

LECTURAS SOBRE MÉTODOS ESTADÍSTICOS EN LA INVESTIGACIÓN



PUBLICACIÓN DEL CENTRO DE APOYO ESTADÍSTICO DE
LA PONTIFICIA UNIVERSIDAD CATÓLICA DE PUERTO RICO

**LECTURAS SOBRE
MÉTODOS ESTADÍSTICOS
EN LA INVESTIGACIÓN**

Publicación del Centro de Apoyo Estadístico

PRIMERA EDICIÓN

JOSÉ NOEL CARABALLO RÍOS, PH.D.
JULIO CÉSAR QUINTANA DÍAZ, PH.D.

LECTURAS SOBRE MÉTODOS ESTADÍSTICOS EN LA INVESTIGACIÓN

Publicación del Centro de Apoyo Estadístico

PRIMERA EDICIÓN



PONTIFICIA UNIVERSIDAD CATÓLICA
DE PUERTO RICO
2021

El material publicado en este libro no puede ser reproducido o transmitido en ninguna forma, medio o formato, total ni parcialmente, sin la autorización escrita del editor y autor.

Lecturas sobre métodos estadísticos en la investigación. Publicación del Centro de Apoyo Estadístico de la Pontificia Universidad Católica de Puerto Rico

Primera edición 2021

© José N. Caraballo Ríos
Julio C. Quintana Díaz

© Pontificia Universidad Católica de Puerto Rico
2250 Boulevard Luis A. Ferré Aguayo, Suite 557
Ponce, Puerto Rico 00717-0655
Teléfono: 787.841.2000
publicaciones.pucpr.edu

Edición: Profa. Olga Bizoso de Montilla
Prensa y Comunicaciones, PUCPR

Diagramación: Carlos Javier Santos Velázquez
coordinador de Publicaciones, PUCPR

Diseño de portada: José González Sotomayor
diseñador gráfico institucional, PUCPR

Impreso en Puerto Rico / Printed in Puerto Rico

Prefacio

*Dra. Iliá C. Rosario Nieves
enlace del Centro de Apoyo Estadístico
Pontificia Universidad Católica de Puerto Rico*

Desde 2012, la Pontificia Universidad Católica de Puerto Rico (PUCPR) comenzó un proyecto con el propósito de proveer apoyo a los estudiantes y a la facultad en sus investigaciones cuantitativas. Esta asesoría comenzó de manera individual. No fue hasta el 2017 que se estableció el Centro de Apoyo Estadístico (CAE) con el fin de ofrecer un programa formal para mejorar la calidad de las investigaciones que se generan en la Institución. Desde el CAE, se provee una serie de seminarios y talleres para que la facultad, los estudiantes y los investigadores se mantengan al día en temas específicos sobre estadísticas, investigación y paquetes estadísticos novedosos.

Con el fin de proveer ayuda adicional a la que ya estaban proveyendo los asesores estadísticos, doctores José Noel Caraballo Ríos y Julio Quintana Díaz, se dieron a la tarea de escribir una serie de artículos, de forma y compilarlos en un cuaderno. Ese cuaderno incluye siete artículos: seis de la autoría del Dr. Caraballo y dos del Dr. Quintana.

Entre los temas se encuentran los siguientes: preguntas de investigación, hipótesis de investigación e hipótesis estadísticas; metanálisis; estudios tipo encuestas; poblaciones escasas u ocultas; ecuaciones estructurales; árboles de clasificación y regresión; y una introducción a la estadística bayesiana. Estos temas resultan medulares para promover la cultura de investigación. La rigurosidad estadística promueve publicaciones de alto impacto, escritos de calidad que son fundamentales para el reconocimiento de la PUCPR en el mundo científico.

Con la publicación de estos artículos, el CAE pretende proveer referencias adicionales en los temas antes mencionados. De igual forma, ofrece información científica adicional para que los investigadores puedan clarificar dudas y utilizar los mismos para la metodología de sus investigaciones. Este cuaderno constituye el primero de una serie de publicaciones relacionadas con la investigación y las estadísticas que publicará la PUCPR.

Tabla de Contenido

Preguntas de investigación, hipótesis de investigación e hipótesis estadísticas	
<i>Dr. José N. Caraballo Ríos</i>	11
La síntesis de investigaciones a través del metanálisis	
<i>Dr. José N. Caraballo Ríos</i>	29
Consideraciones en el diseño de estudios tipo encuesta (muestreo)	
<i>Dr. José N. Caraballo Ríos</i>	47
Muestras y estimación de parámetros en poblaciones escasas u ocultas	
<i>Dr. Julio C. Quintana Díaz</i>	67
Estimación de parámetros en poblaciones ocultas (ensayo II)	
<i>Dr. Julio C. Quintana Díaz</i>	85
Introducción al Análisis mediante Ecuaciones Estructurales	
<i>Dr. José N. Caraballo Ríos</i>	93
Árboles de clasificación y regresión	
<i>Dr. José N. Caraballo Ríos</i>	111
Introducción a la estadística bayesiana	
<i>Dr. José N. Caraballo Ríos</i>	127
Datos sobre el Centro de Apoyo Estadístico	143

Preguntas de investigación, hipótesis de investigación e hipótesis estadísticas

*José Noel Caraballo Ríos, Ph.D.
asesor del Centro de Apoyo Estadístico
Pontificia Universidad Católica de Puerto Rico*

INTRODUCCIÓN

Un área en la que los estudiantes que realizan su primera investigación con frecuencia tienen dudas es en la relación entre las preguntas de investigación, las hipótesis de investigación y las hipótesis estadísticas. La raíz del problema es que, en los cursos introductorios de estadísticas, tradicionalmente, no se discuten aspectos fundamentales de la filosofía de la ciencia que son, en última instancia, la base de la investigación científica en cualquier disciplina de estudios. Este breve ensayo pretende presentar estas ideas fundamentales y su relación con el análisis estadístico. Se discutirán algunos conceptos básicos de la inferencia estadística tales como los valores p , niveles de significación, y pruebas de una y dos colas. Finalmente se discutirá el problema de multiplicidad de pruebas estadísticas y cómo atenderlo.

LA EXPLICACIÓN CIENTÍFICA

De acuerdo con Karl Popper (1956), el objetivo de la ciencia es “... encontrar explicaciones satisfactorias para aquello que nos parezca que necesita explicación.” La explicación, entonces, es un conjunto de aseveraciones lógicamente consistentes que permiten deducir el fenómeno. Al conjunto de aseveraciones que conforman la explicación se le llama una explicación científica, o más generalmente, una **teoría**

científica. De esta manera, la investigación científica consta en elaborar las explicaciones para el fenómeno de interés y en someter a prueba, mediante la experimentación, dichas explicaciones. Cuando los experimentos confirman las teorías, estas se retienen; cuando no, se descartan totalmente o se modifican. Si la teoría se modifica, entonces pasa por un nuevo proceso de verificación y así sucesivamente.

Aunque parezca paradójico, la veracidad de una teoría científica no puede establecerse nunca, razón por la cual Popper desarrolló el concepto de “falseabilidad” de una teoría. El problema al que apunta Popper es que, el hecho de que, aun cuando el resultado de uno o más experimentos esté de acuerdo con la teoría, no hay ninguna garantía de que experimentos futuros también lo estén. Sin embargo, aunque una teoría no se pueda demostrar que sea cierta, sí es posible demostrar su falsedad: bastará con un experimento cuyo resultado no sea el predicho por la teoría. De esta forma, un requisito de una teoría científica es que resulte, aunque sea en principio, “falseable”. Es decir, debe ser posible pensar un experimento que permita poner a prueba la teoría para determinar si esta es falsa.

LAS HIPÓTESIS DE INVESTIGACIÓN

El proceso de someter a prueba una teoría consiste en elaborar y someter a prueba las hipótesis de investigación (véase la **Figura 1**). Una hipótesis de investigación es una aseveración acerca del resultado esperado de un experimento: es lo que el investigador cree que ocurrirá como resultado de su estudio. Considere algunos ejemplos:

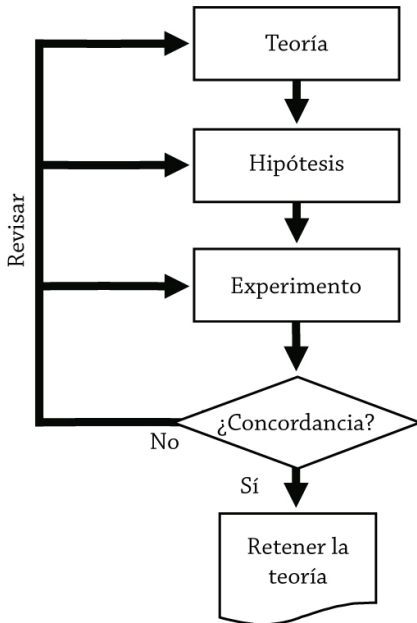
- La droga X reducirá los niveles de colesterol en los pacientes.
- La terapia X reducirá la depresión en los adolescentes.
- El método X producirá mayor aprendizaje de matemáticas que el método Y.

En cada uno de estos ejemplos se presume que hay una teoría que permite hacer dichas predicciones. Note que en todos los casos hay algún efecto que se espera observar. Note, además, la ausencia de mención alguna de parámetros estadísticos (medias, proporciones, correlaciones,

etc.) u otros términos estadísticos tales como “diferencia significativa” y “muestra representativa”, entre otros.

Figura 1

Proceso de someter a prueba una teoría científica.



CRITERIOS QUE DEBE CUMPLIR UNA TEORÍA CIENTÍFICA

Para que una teoría científica sea tomada seriamente tiene que cumplir con dos criterios. En primera instancia, tiene que producir resultados teóricos que concuerden con las observaciones realizadas. Es decir, tiene que explicar el fenómeno que se pretendía explicar. Sin embargo, esto no es suficiente: tiene además que predecir resultados de observaciones que no se hayan realizado aún. La razón para ello es que siempre es posible construir, *a posteriori*, una teoría que prediga observaciones realizadas con anterioridad. Sin embargo, no es posible construir *a priori* una teoría para explicar observaciones que aún no se hayan realizado. La siguiente sección presenta un ejemplo que permite aclarar esta idea.

UN EJEMPLO: LA TEORÍA GENERAL DE LA RELATIVIDAD

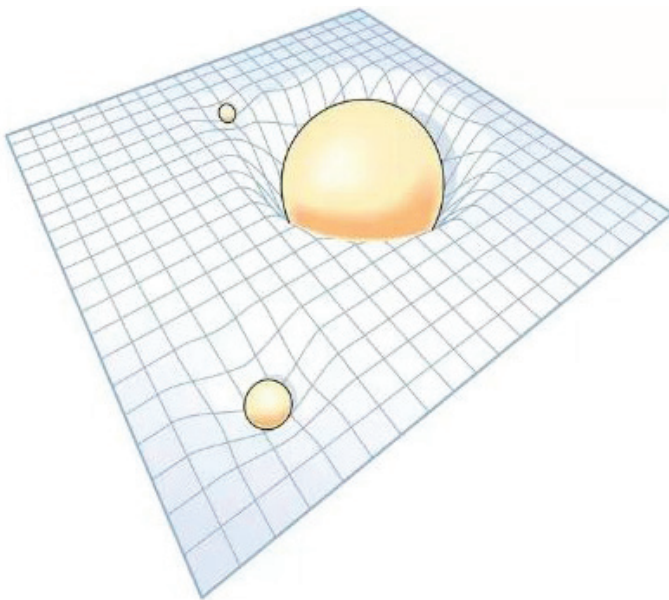
La Teoría General de la Relatividad, formulada por Albert Einstein en 1915, extiende la Teoría Especial para los casos en que los objetos aceleran. Esta teoría plantea la idea de que la gravedad se debe a la flexibilidad del espacio-tiempo. En forma matemática, la teoría se expresa mediante la siguiente ecuación:

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu}$$

En esencia, esta ecuación muestra la equivalencia entre la gravedad (lado derecho) con la geometría del espacio-tiempo (lado izquierdo). Un objeto masivo distorsiona el espacio a su alrededor, como se ilustra en la **Figura 2**. Entonces, un segundo objeto en sus cercanías será “atraído” debido a la geometría (depresión) del espacio.

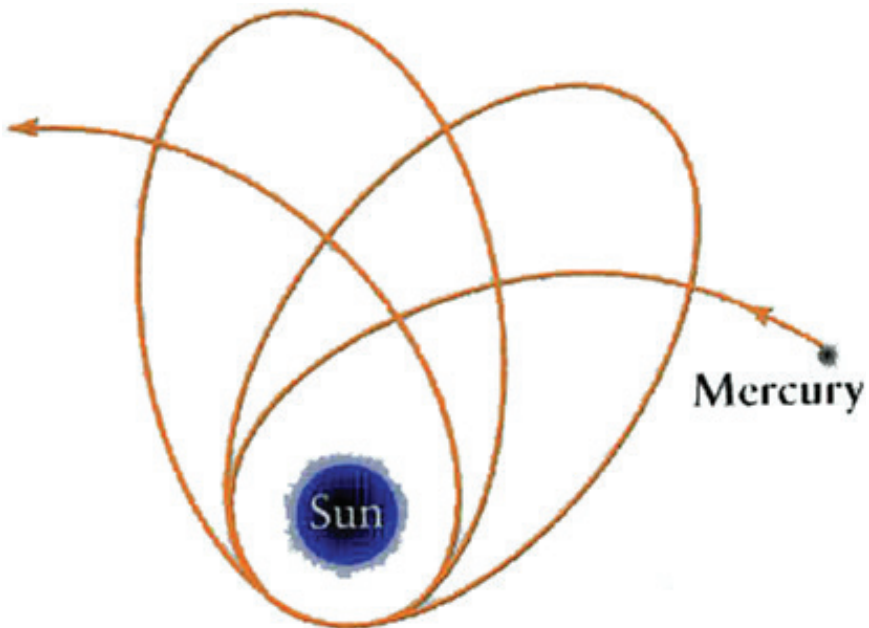
Figura 2

Representación de dos dimensiones de la distorsión del espacio-tiempo.



La Teoría de la Relatividad General puede explicar los fenómenos gravitatorios al igual que la Teoría de Gravitación de Isaac Newton. De hecho, es capaz de explicar el fenómeno de precesión de la órbita del planeta Mercurio, cosa que la teoría de Newton no puede hacer. La precesión se refiere a la rotación en el espacio del eje mayor de la elipse que forma la órbita del planeta (véase la **Figura 3**). Note que esta evidencia no es suficiente, pues se conocía con anterioridad a la teoría de Einstein. Einstein, sin embargo, hizo una nueva predicción: si el espacio alrededor de un objeto masivo se distorsiona, entonces un haz de luz que pase cerca de dicho objeto se desviará de su trayectoria rectilínea (véase la **Figura 4**).

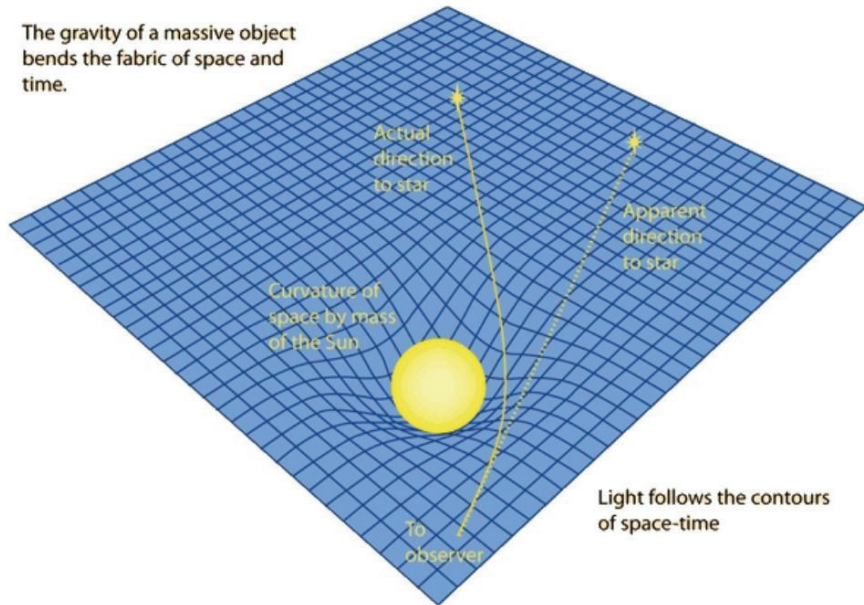
Figura 3
Órbita de Mercurio.



Nota. La órbita de Mercurio, cuya forma es elíptica, no está fija en el espacio sino que tiende a girar, fenómeno conocido como precesión.

Figura 4

Desviación de un haz de luz que pasa cerca de un objeto con un espacio alrededor distorsionado.



Nota. La depresión que genera el Sol en el espacio-tiempo hace que un haz de luz que pase cerca de este se desvíe de su trayectoria rectilínea, dando lugar a que se observe una posición “aparente” de la estrella y no su posición real.

Einstein ideó un experimento para probar esta hipótesis. Las estrellas principales en el firmamento están catalogadas y su posición en la esfera celeste es fija y conocida. Por otro lado, el Sol se mueve por la esfera celeste en la región en la que se encuentran las constelaciones del Zodiaco. Bastaría entonces con ver si, cuando el Sol pasa cerca de una estrella, la posición de esta cambia. Claro está, la luz del Sol es demasiado intensa en comparación con la de las estrellas, por lo que no es factible hacer esto cualquier día. Einstein se percató que esto sería factible hacerlo durante un eclipse solar. Él calculó cuánto sería la magnitud y dirección de la desviación que sufrirían las estrellas durante el eclipse de Sol de 1919. Sir Arthur Eddington fue quien hizo las medidas astronómicas durante este eclipse, las cuales concordaron con las predicciones hechas por Einstein. La **Figura 5(a)** muestra una de las

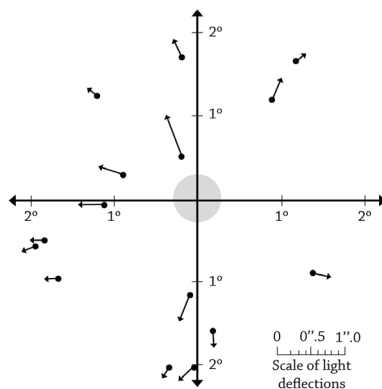
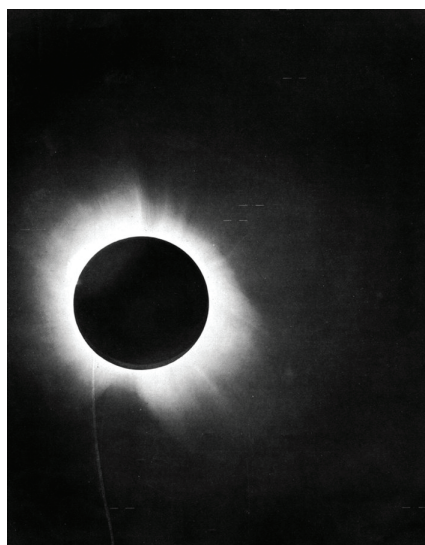
fotos tomadas por Eddington durante el eclipse de 1919. La brillantez de las estrellas es muy débil como para apreciar su posición sin la ayuda de una lente de aumento. La **Figura 5(b)** muestra la posición verdadera (punto redondo) y la posición aparente (punta de la flecha) de las estrellas cercanas al Sol (círculo gris en el centro del eje de coordenadas).

La Teoría General de la Relatividad ha generado muchas otras predicciones de fenómenos que no se habían observado previamente y que han sido corroborados, por lo que es una teoría muy firme. Esto no significa que en el futuro alguna nueva predicción no se cumpla o que se realice alguna observación nueva que no pueda ser explicada por esta teoría.

Figura 5

a. Eclipse solar de 1919.

b. Posiciones reales vs. posiciones aparentes.



Nota. (a) Foto tomada por Sir Arthur Eddington durante el eclipse solar de 1919. (b) Posiciones reales (círculos negros) y posiciones aparentes (puntas de flechas) de las estrellas cercanas al Sol durante el eclipse de 1919. El círculo gris en el centro del eje de coordenadas representa al Sol.

LA INVESTIGACIÓN HIPOTÉTICO-DEDUCTIVA Y LA EXPLORATORIA

El tipo de investigación descrita en el ejemplo anterior se denomina “hipotético-deductiva” pues consiste en someter a prueba las predicciones o hipótesis que se deriven de una teoría científica. Sin embargo, no toda investigación es de este tipo. Por ejemplo, las encuestas de opinión, una metodología utilizada a menudo en las ciencias sociales, usualmente no se realiza para someter a prueba una teoría. A los tipos de investigaciones en las que no se somete a prueba una teoría se les llama “exploratorias”. Esta diferencia, que a menudo los estudiantes pasan por alto, tiene implicaciones importantes para la redacción de los informes de investigación y también al momento de analizar los datos.

En cuando a la redacción del informe de investigación, en las investigaciones de tipo exploratorio no se redactan hipótesis de investigación. Lo adecuado en tal caso es presentar solo las preguntas de investigación en la sección o capítulo introductorio del informe. A menudo los estudiantes presentan hipótesis estadísticas como si fueran hipótesis de investigación, lo cual tampoco es correcto. De hecho, en las investigaciones exploratorias, al igual que en las hipotético-deductivas, corresponde presentar las hipótesis estadísticas en la sección de Metodología. En el caso de las investigaciones hipotético-deductivas corresponde presentar las hipótesis de investigación al final de la sección con el Marco Teórico. Esto es así, pues como se observó anteriormente, las hipótesis son predicciones que se derivan de la teoría.

EL SIGNIFICADO DEL TÉRMINO “EXPERIMENTO”

En las secciones anteriores se ha utilizado el término “experimento” casualmente, como si no tuviese más que una acepción. Sin embargo, esto no es así, por lo que vale la pena comentar sobre sus significados, antes de seguir adelante.

En las ciencias naturales, de donde se deriva inicialmente este término, un experimento se refiere a cualquier investigación cuantitativa. Así, por ejemplo, uno aplica este término para describir una investigación de laboratorio en la que se midan las propiedades de alguna sustancia u objeto, o en la astronomía, cuando se haga alguna observación o medida cuantitativa. En las ciencias sociales, incluyendo

la educación y la psicología, y en la medicina, el término “experimento” tiene una connotación más limitada. Un experimento es un tipo de investigación en el cual los objetos de estudio (personas, animales, etc.) se asignan aleatoriamente a dos o más tratamientos experimentales. En la medicina, los estudios clínicos para probar la efectividad de un medicamento son de este tipo.

Finalmente, en estadística, un experimento se refiere a cualquier procedimiento que pueda ser repetido indefinidamente y tenga un conjunto bien definido de posibles resultados, conocido como el espacio muestral (*sample space*). Se dice que un experimento es *aleatorio* si tiene más de un posible resultado (de lo contrario sería determinista). Así que lanzar un dado para anotar el número que queda en la cara de arriba es un experimento. También lo es realizar una encuesta. Por ejemplo, en una encuesta para determinar quién ganará una elección, se presume que hay un número definido de posibles opciones, así como alguna distribución probabilística conocida. Si la muestra de participantes es aleatoria, entonces se cumple también el requisito de que las respuestas observadas sean aleatorias.

LAS HIPÓTESIS ESTADÍSTICAS

Una hipótesis estadística es una aseveración acerca de un *parámetro* (es decir, sobre el valor de un índice, tal como una media, diferencia entre medias, proporción, correlación o un coeficiente de regresión). Hay dos tipos de hipótesis estadísticas conocidas como “hipótesis nula”, representada simbólicamente por H_0 , y la “hipótesis alterna”, presentada por H_a . La hipótesis nula es la que se somete a prueba estadísticamente, mientras que la hipótesis alterna es la que se retiene si la primera se rechaza.

La hipótesis nula tiene ese nombre pues, contrario a una hipótesis de investigación, asevera que un tratamiento experimental no será efectivo o, en términos más generales, que no existe una relación entre las variables bajo estudio. Las hipótesis estadísticas se diferencian también de las de investigación en cuanto a que las primeras son aseveraciones acerca de parámetros estadísticos, mientras que las segundas no. Es común, además, presentar las hipótesis estadísticas en forma simbólica, aunque también se pueden presentar textualmente.

Considere el ejemplo mencionado anteriormente:

Hipótesis de investigación: La droga X reducirá los niveles de colesterol en los pacientes.

Hipótesis nula (simbólica): $H_0: \mu_E = \mu_C$

Hipótesis nula (textual): No existirá una diferencia estadísticamente significativa en el colesterol medio de los pacientes que reciban la droga X y los que no la reciban.

Hipótesis alterna (simbólica): $H_a: \mu_E < \mu_C$

Hipótesis alterna (textual): El nivel de colesterol medio de los pacientes que reciban la droga X será significativamente menor que el de los pacientes que no reciban la droga.

Note que en la redacción de las hipótesis estadísticas se menciona un parámetro: en el ejemplo, la media de cada grupo (o más bien, la diferencia entre las medias de los grupos). Además, se utiliza el término estadístico “significativa” o “significativamente”. En una próxima sección se tratará en más detalle lo que quiere decir “estadísticamente significativo”.

Una pregunta que a menudo hacen los estudiantes que toman su primer curso de inferencia estadística es: ¿Por qué se somete a prueba la hipótesis nula y no la alterna? La respuesta es que, al igual que una teoría, las hipótesis son aseveraciones con carácter de universalidad. Es decir, para probar la veracidad de una hipótesis, habría que conducir un número infinito de experimentos y en todos ellos la hipótesis podría predecir correctamente sus resultados. Claramente, esto no sería viable. Sin embargo, es sencillo determinar si la hipótesis es incorrecta: solo bastaría con hacer un estudio en el que los resultados no fueran los esperados bajo la hipótesis. De nuevo, si los resultados fueran los esperados bajo la hipótesis, se retendría la hipótesis hasta tanto se demostrara su falsedad. Así, cuando se somete a prueba la hipótesis nula y se rechaza, se hace porque la evidencia favorece a la hipótesis de investigación, pero no porque se haya probado “cierta” esta última. Retener la hipótesis nula, por el contrario, indica que los datos observados no son consistentes con la hipótesis de investigación.

LAS PRUEBAS DE UNA Y DOS COLAS

Usualmente, para probar si existe una diferencia significativa entre los promedios de dos grupos en un experimento, se hace uso de prueba estadística t de Student. Cuando se hace esta prueba, la hipótesis alterna puede ser direccional ($H_a: \mu_E > \mu_C$ ó $H_a: \mu_E < \mu_C$), o no direccional ($H_a: \mu_E \neq \mu_C$). Para las hipótesis direccionales, la región de rechazo se encuentra a uno de los dos extremos de la distribución de la estadística t , por lo cual se le llama “prueba de una cola”. Para la hipótesis no direccional, la región de rechazo se distribuye equitativamente en ambos extremos de la distribución de t , por lo cual se le denomina “prueba de dos colas” (véase la **Figura 6**).

Figura 6

a. Prueba de hipótesis de una cola.

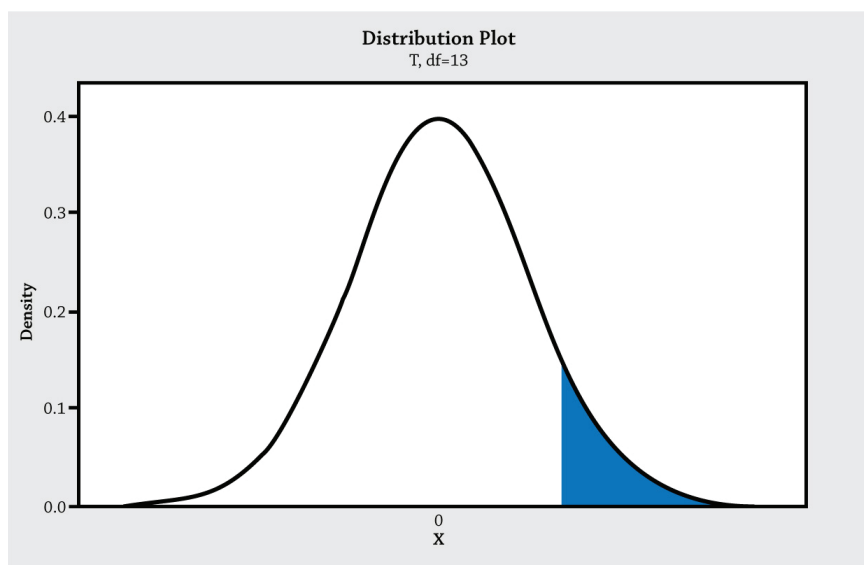
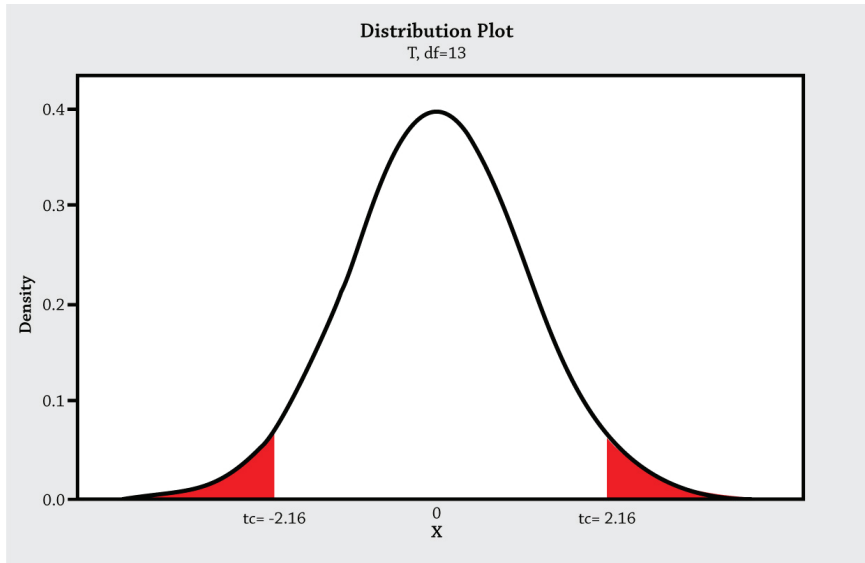


Figura 6 (continuación)

b. Prueba de hipótesis de dos colas.



Nota. En una prueba de una cola la región de rechazo se encuentra en un uno de los dos extremos de la distribución (a), mientras que en una prueba de dos colas la región de rechazo se distribuye equitativamente en los dos extremos (b). Note que el área asociado a la región de rechazo es numéricamente igual en (a) que en (b).

De la discusión en la sección anterior debe ser evidente que, ya que en los estudios hipotético-deductivos las hipótesis de investigación —y por lo tanto, las hipótesis alternas— son direccionales, la prueba estadística debe ser de una cola. En los estudios exploratorios, en los cuales no haya hipótesis científica que indique direccionalidad, las hipótesis alternas son no direccionales, por lo que la prueba estadística será de dos colas. Si en un estudio hipotético-deductivo se utiliza una prueba de dos colas, entonces el nivel de significación se reduciría a la mitad. Por ejemplo, para un nivel de significación $\alpha = 0.05$, el área asociada a la región de rechazo es $0.05/2 = 0.025$. Como en los estudios hipotético-deductivos las hipótesis alternas son direccionales, entonces uno de los dos extremos quedaría descartado. En tal caso, habría que quedarse con un solo extremo de la distribución y, por lo tanto, con la mitad del área en la región de rechazo. Así que, si se realiza una prueba de hipótesis al nivel $\alpha = 0.05$, pero se hace una prueba de dos

colas, el verdadero nivel de significación será $\alpha = 0.025$. Si el valor de probabilidad obtenido fuese de, quizá 0.030, se retendría la hipótesis nula. Sin embargo, como este valor es menor a 0.05 lo que se debería hacer es rechazarla.

LA INTERPRETACIÓN DEL VALOR P Y α

La **Figura 7** presenta un ejemplo del resultado que despliega el programa SPSS al realizar una prueba t para grupos independientes. En esta tabla la estadística F y su correspondiente valor de p se utilizan para someter a prueba la hipótesis de homogeneidad de varianzas. Las siguientes dos columnas se utilizan para someter a prueba la hipótesis de igualdad entre las medias de dos grupos. Como se indica en la sexta columna, los valores de probabilidad que se presentan para esta prueba corresponden a una prueba no direccional (dos colas). ¿Qué representa esta probabilidad?

Gráficamente este resultado se presenta en la **Figura 6(b)**. La suma de las áreas en rojo es el valor de probabilidad que presenta la tabla, 0.014. Lo que este valor indica es lo siguiente: si una variable (en este caso la diferencia entre las medias) se distribuye como una distribución t con 13 grados de libertad, entonces, la probabilidad de obtener un valor de t que sea mayor a +2.85 o menor a -2.85 es de 0.014. En ese caso particular, asumiendo un nivel de significación $\alpha = 0.05$, se rechazaría la hipótesis nula ya que el valor obtenido de p es menor que α ($0.014 < 0.05$). Vale la pena señalar que el valor de p que se obtenga como resultado de una prueba estadística no es indicativo de la magnitud del efecto de un tratamiento. Es decir, un valor $p = 0.001$ no indica que el tratamiento sea más efectivo que en un estudio similar con $p = 0.015$, aunque el segundo valor de p sea quince veces más grande que el primero. Así que no es correcto afirmar que “la diferencia observada es altamente significativa” o decir que, porque p es muy pequeño, el tratamiento es “muy efectivo”.

Note que los valores de t en la región de rechazo son posibles, pero poco probables. Así que cuando se rechaza la hipótesis nula, lo que se quiere decir es que la diferencia entre las medias es tan grande que es poco probable que los dos grupos provengan de una misma población. Por supuesto, el que sea poco probable no quiere decir que la hipótesis nula esté incorrecta, por lo que se cometería un error al rechazarla. Como

al fijar el nivel de significación se fija la probabilidad de cometer un error, a este también se le llama “error Tipo I”. Así que el error Tipo I no es otra cosa que la probabilidad de rechazar la hipótesis incorrectamente. Por supuesto, es posible cometer el error contrario: retener la hipótesis nula cuando debería ser rechazada. A este último error se le conoce como el “error Tipo II”.

Figura 7

Ejemplo de tabla de resultados de SPSS para una prueba t de grupos independientes.

		Independent Samples Test									
		Levene's Test for Equality of Variances		t-test for Equality of Means						95% Confidence Interval of the Mean	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper	
TAPSPEED	Equal variances assumed	.319	.582	-2.845	13	.014	-14.00	4.92	-24.63	-3.37	
	Equal variances not assumed			-2.927	11.865	0.013	-14.00	4.78	-24.44	-3.56	

En el párrafo anterior hay una idea fundamental que merece elaboración. Si se realiza una investigación y la prueba estadística, podría decirse una prueba *t*, resulta significativa ¿qué haría pensar que la diferencia encontrada es evidencia de un efecto real y no es un resultado espurio? La respuesta depende de si el estudio es hipotético-deductivo o si es exploratorio. En el estudio hipotético-deductivo hay una teoría que predice el resultado. Si el resultado es el esperado, entonces habría razones para pensar que sería correcto rechazar la hipótesis nula: la explicación que provee la teoría. En un estudio exploratorio, en ausencia de una teoría, no se puede asegurar que el resultado no sea espurio. Por este motivo, un estudio exploratorio no es válido para probar una hipótesis científica. Lo más que podría hacerse cuando se obtiene una diferencia significativa es postular la posibilidad de que exista un efecto real. Para confirmar si lo es o no lo es, habría que elaborar una teoría que explicara este resultado y realizar un experimento posterior del tipo hipotético-deductivo.

EL PROBLEMA DE LA MULTIPLICIDAD DE PRUEBAS

La multiplicidad de pruebas se refiere a la realización de más de una prueba estadística con un mismo conjunto de datos. Existen varios posibles escenarios en donde esto podría suceder:

- En estudios hipotético-deductivos: a) cuando haya más de una hipótesis principal, o cuando haya una (o más) hipótesis principal, pero también hipótesis secundarias. b) Cuando en el diseño de la investigación haya más de dos tratamientos.
- En los estudios exploratorios, cuando se desee realizar más de una prueba para determinar si existen relaciones entre las variables de interés.

El problema es el siguiente: si se llevan a cabo varias pruebas de hipótesis con un mismo conjunto de datos, la probabilidad de que alguna de las pruebas resulte significativa es mayor que el valor nominal de α . Por ejemplo, si se llevan a cabo cinco pruebas de hipótesis al nivel $\alpha = 0.05$, entonces la probabilidad de que alguna de las pruebas resulte significativa solo por casualidad es igual a $\alpha_T = 1 - (1 - 0.05)^5 = 0.23$. Es decir, que hay casi una probabilidad de 1 en 4 de que alguna de las pruebas resulte significativa de forma espuria.

Hay diferentes alternativas para tratar este problema, dependiendo de si el estudio es hipotético-deductivo o exploratorio, si hay más de una variable dependiente de interés, y si hay más de dos grupos en un experimento. En los estudios hipotético-deductivos, como se sabe de antemano la cantidad de hipótesis a probar, una manera de controlar el error Tipo I global es reducir el nivel de significación de las pruebas individuales a α/n , donde n es el número de pruebas de hipótesis a realizarse. Por otro lado, si hay más de una hipótesis primaria, es posible realizar una prueba multivariada, tal como un MANOVA, en vez de pruebas t individuales para cada hipótesis. También es posible construir un índice compuesto por las variables dependientes de interés y hacer una sola prueba usando el índice construido como la variable dependiente. En el caso de un experimento en el cual haya más de dos grupos, cuando se encuentre una estadística F significativa, entonces se deben utilizar las pruebas de comparaciones múltiples de Dunnett, si

solo se desea comparar cada grupo experimental con el grupo control, o la prueba de Bonferroni, si las comparaciones que se derivan de las hipótesis de investigación son más generales.

En el caso de estudios exploratorios, si se realiza más de una prueba, es importante recordar que la probabilidad de obtener resultados espurios aumenta con el número de pruebas que se realice. Un cálculo rápido con la fórmula presentada arriba le servirá al investigador para tener una idea de cuál es esta probabilidad, y así no entusiasmarse demasiado si encuentra algún resultado significativo. En los estudios exploratorios en los cuales haya más de dos grupos, si se obtiene una F significativa se debe usar las pruebas de comparaciones múltiples LSD, HSD, Tukey o Scheffé, o alguna otra que sea apropiada para explorar diferencias entre las medias de los grupos.

RESUMEN

La idea principal esbozada en este escrito es que el tipo de investigación (hipotético-deductiva o exploratoria) tiene implicaciones sobre sí corresponde redactar hipótesis de investigación, la redacción de la hipótesis estadística alterna y los análisis que se deben realizar cuando haya multiplicidad de pruebas estadísticas.

Se observó que las hipótesis de investigación son deducciones lógicas de la teoría que predicen los resultados de un experimento. Por lo tanto, solo es apropiado redactarlas en los estudios hipotético-deductivos. Como las hipótesis de investigación indican la direccionalidad de un efecto, las hipótesis alternas correspondientes deben ser del tipo direccional también. En un estudio exploratorio corresponde redactar hipótesis alternas no direccionales. Finalmente, el problema de multiplicidad de pruebas se atiende de manera distinta en estos tipos de estudios. En particular, cuando hay más de dos grupos en los estudios hipotético-deductivos se deben utilizar pruebas de comparaciones múltiples que permitan controlar el error Tipo I para un número limitado de comparaciones o contrastes entre medias, tales como las pruebas de Dunnet y Bonferroni, mientras que en los descriptivos se deben utilizar pruebas que permitan un número ilimitado de comparaciones o contrastes tales como las de Tukey y Scheffé.

El tipo de estudio también determina la manera en que se redacta un informe de investigación. En ambos casos, las preguntas de investigación se presentan en la sección o capítulo introductorio. En los estudios hipotético-deductivos, las hipótesis de investigación deben presentarse al final de la sección de teoría del escrito. Además, las hipótesis estadísticas (nula y alterna) deben presentarse en la sección sobre metodología del documento. Un detalle adicional que debe tenerse en cuenta en la redacción del informe de investigación es no redactar las hipótesis de investigación como si fuesen hipótesis estadísticas.

BIBLIOGRAFÍA

- Copi, I. M., Cohen, C. & Rodych, V. (2019). *Introduction to Logic*, 15th Ed. New York, NY: Routledge.
- Einstein, A. (1961). *Relativity: The Special and the General Theory*. New York, NY: Crown Publishers, Inc.
- Kuehl, R. O. (2000). *Design of Experiments: Statistical Principles of Research Design and Analysis*. Pacific Grove, CA: Duxbury.
- Popper, K. (1956). *The Aim of Science. En: Realism and the Aim of Science*, 2000. New York, NY: Routledge.

La síntesis de investigaciones a través del metanálisis

*José Noel Caraballo, Ph.D.
asesor del Centro de Apoyo Estadístico
Pontificia Universidad Católica de Puerto Rico*

INTRODUCCIÓN

El metanálisis es un tipo de análisis secundario, en el cual se analizan los resultados de un conjunto de estudios cuantitativos para obtener un estimado global de un parámetro (diferencia estandarizada entre medias o proporciones, correlación y *odds ratio*, entre otros). Aunque la técnica tiene sus raíces en el siglo XVII, no fue hasta principios del siglo 20 que se llevaron a cabo los primeros estudios metanalíticos formales. La técnica ganó popularidad en 1981 con la publicación del primer texto sobre el tema por Glass, McGaw y Smith (1981). El formalismo estadístico fue elaborado unos años más tarde por Hedges y Olkin (1985). Desde entonces se han publicado cientos de metanálisis que cubren la gran mayoría de las disciplinas en las que se realizan investigaciones cuantitativas.

La idea general del metanálisis es calcular una estadística para cada estudio, llamada la **magnitud del efecto**, y luego combinarlas para obtener un estimado global del efecto producido por el tratamiento experimental o un estimado de la magnitud de la relación entre dos variables. Además, el metanálisis puede permitir explorar relaciones entre variables, incluso en casos en donde no sea posible o sea muy difícil de llevar a cabo en un estudio primario. Realizar un metanálisis puede ser el objetivo final de un estudio, pero también puede ser utilizado como parte de la revisión de la literatura en un estudio primario.

TIPOS DE METANÁLISIS

Existen dos tipos de metanálisis: de efectos fijos y de efectos aleatorios. El modelo de efectos fijos asume que hay un valor poblacional verdadero y que las fluctuaciones de ese valor observadas en los distintos estudios son debido al error de muestreo. El modelo de efectos aleatorios, por el contrario, asume que las magnitudes de los efectos dependen de características de los estudios, tales como edad, etnia, estado de salud de los participantes, o alguna otra característica del estudio. El metanálisis de efectos fijos es adecuado en situaciones en las que los estudios que se sintetizarán utilicen la misma población de sujetos (u objetos), misma metodología, idénticos tratamientos de control y experimental, y los mismos instrumentos de medición. Esta restricción es típica de los estudios médicos clínicos. El metanálisis de efectos aleatorios es adecuado en casos en que no se puedan cumplir una o más de estas condiciones. Esto último es más típico en los estudios educativos y psicológicos.

En este ensayo se utilizará, como ejemplo, la síntesis de estudios sobre la efectividad de los mapas de conceptos como estrategia de aprendizaje en ciencias. Estos estudios pueden diferir en metodología (cuasiexperimentos vs. experimentos); contenido utilizado en el estudio (biología, física, química, u otras ciencias); instrumentos de medición (pruebas desarrolladas por los investigadores no estandarizadas); y nivel educativo en el que se realice el estudio (escuela elemental, escuela intermedia, escuela superior, nivel universitario), entre otros. Por este motivo, este ejemplo ilustrará el análisis bajo el modelo de efectos aleatorios. Aunque en la actualidad haya paquetes estadísticos para llevar a cabo metanálisis, las fórmulas para llevar a cabo los cálculos son lo suficientemente simples como para hacerlos usando el programa Excel.

PROCEDIMIENTO GENERAL

El metanálisis requiere de un acervo de artículos de investigaciones primarias. El investigador debe, antes que nada, definir el problema que desee estudiar, las variables dependientes e independientes que desee incluir y delimitar la población de estudios que desee incluir en su síntesis. Además, deberá establecer cualquier otro criterio de inclusión en la muestra de artículos que sintetizará.

En general, los estudios primarios se obtienen haciendo búsquedas exhaustivas en bases de datos de artículos de la disciplina bajo estudio. Si el investigador decide incluir investigaciones no publicadas en revistas arbitradas, se puede ampliar la búsqueda incluyendo disertaciones doctorales (disponibles en el *Dissertation Index*) e incluso los resúmenes o *Proceedings* de conferencias profesionales.

Una vez se obtengan los artículos, hay que revisarlos individualmente para confirmar que cumplan con los criterios de inclusión y entonces codificarlos con una guía preparada con anterioridad. La codificación recoge los resultados del estudio y las características más relevantes del mismo (tamaño de muestra, población considerada, instrumentos usados en la recopilación de datos, entre otras). Luego de esto, para cada estudio se calcula una estadística que mida cuán efectivo fue el tratamiento experimental o cuán fuerte fue la relación entre las variables, llamado la magnitud del efecto (*effect size*).

Una vez calculada la magnitud del efecto de cada estudio en la muestra, se procede a agregar dichos resultados para obtener la magnitud de efecto global y hacer un intervalo de confianza alrededor de su promedio. Adicional a esto, se pueden llevar a cabo análisis para establecer la relación entre la magnitud del efecto y las características de los estudios.

LA MAGNITUD DEL EFECTO

La magnitud del efecto es una cantidad calculada en un estudio cuantitativo que indica cuán grande es el efecto de un tratamiento experimental o cuán grande es la relación entre dos variables. En este escrito se considerarán solo estudios experimentales para los cuales se tomará la diferencia entre medias estandarizada como el índice de magnitud del efecto. Para otros índices el procedimiento de agregación de resultados es similar al que se presenta aquí.

La diferencia entre medias estandarizadas, o estimador de Glass, no es otra cosa que la diferencia entre las medias de los grupos control y experimental, dividido entre una desviación estándar, que puede ser la del grupo control, o, más generalmente, la desviación estándar ponderada de los dos grupos. La mayoría de los estudios que se citan en el ejemplo que se dará a continuación son experimentos o cuasiexperimentos de dos grupos, en los cuales, la significancia se determinó mediante la estadística *t*, por lo que tienen toda la información necesaria para calcular

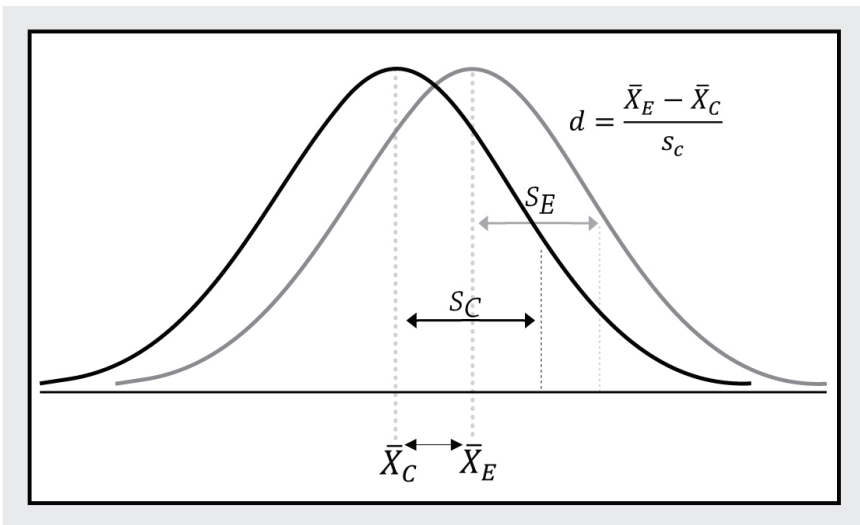
la magnitud del efecto. En estudios con más de dos grupos, estudios con diseños de más de un factor o en los que se haga uso de covariados, la estimación de la desviación estándar ponderada es más elaborada. En este escrito no se presentarán estos casos, pero el lector interesado puede consultar el libro de Hedges, Woodworth y Shimansky (1989) en el que se discute cómo obtener la desviación estándar ponderada en esos casos.

El valor de g dice cuánto más alto es el promedio del grupo experimental que el del control, en unidades de la desviación estándar del grupo control. Usando una tabla de la distribución z , el lector puede verificar que media desviación estándar sobre el promedio ($z = 0.5$) equivale a una percentila de 69.15. Por ejemplo, en un estudio sobre aprendizaje, esta magnitud del efecto se puede interpretar entonces como: “El efecto del tratamiento experimental es aumentar la percentila del estudiante promedio de 50 a 69.15”.

Cohen (1988) indica que una magnitud del efecto de alrededor de 0.20 es considerada baja, de alrededor de 0.50 es considerada moderada, mientras que alrededor de 0.8 es considerada alta.

Figura 1

Magnitud del efecto (Glass).



Nota. La magnitud del efecto (Glass) es la diferencia estandarizada entre los promedios de los grupos control y experimental.

En su libro, Hedges y Olkin demuestran que el estimador de Glass está viciado, por lo que requiere de un factor multiplicativo para que no lo sea. El estimador de Hedges y Olkin está dado por:

$$g = j \cdot d$$

donde d es el estimador de Glass, dado por:

$$d = \frac{\bar{X}_E - \bar{X}_C}{s_c}$$

y J es el factor desarrollado por Hedges y Olkin para eliminar el sesgo en g :

$$J = 1 - \frac{3}{4m - 1}$$

En esta última ecuación m representa los grados de libertad, es decir, la suma de los tamaños de muestra de los dos grupos menos 2.0. La siguiente ecuación se utiliza para calcular el error estándar de g , para así construir el intervalo de confianza alrededor de cada magnitud de efecto individual:

$$SE_g = J \cdot SE_d$$

donde

$$SE_d = \sqrt{\frac{n_E + n_C}{n_E n_C} + \frac{d^2}{2(n_E + n_C)}}$$

El intervalo de confianza estará dado entonces por:

$$g \pm z \cdot SE_g$$

En esta última ecuación, el valor de z dependerá del nivel de confianza deseado. Para un nivel de confianza de 95%, $z = 1.96$, mientras que para 99% de confianza, $z = 2.33$.

EJEMPLO: CÁLCULO DE LA MAGNITUD DEL EFECTO

Los datos de este ejemplo provienen de un estudio publicado en Internet por Aziz y Rahman (véase la **Tabla 1**). El estudio compara el uso de mapas de conceptos sobre el aprendizaje de ciencias en estudiantes de escuela superior.

Tabla 1

Resultados obtenidos por Aziz y Rhaman en su estudio

	E	C
\bar{X}	18.43	11.60
s	5.09	4.89
n	60	60

Cálculo de d:

$$d = \frac{\bar{X}_E - \bar{X}_C}{s_c} = \frac{18.43 - 11.60}{4.89} = 1.3967$$

Cálculo de J:

$$J = 1 - \frac{3}{4m - 1} = 1 - \frac{3}{4(60 + 60 - 2) - 1} = 0.9936$$

Cálculo de la magnitud del efecto:

$$g = J \cdot d = (0.9936)(1.3967) = 1.39$$

Cálculo del error estándar de d:

$$SE_d = \sqrt{\frac{n_1 + n_2}{n_1 n_2} + \frac{d^2}{2(n_1 + n_2)}} = \sqrt{\frac{60 + 60}{(60)(60)} + \frac{1.39^2}{2(60 + 60)}} = 0.2034$$

Cálculo del error estándar de g :

$$SE_g = J \cdot SE_d = (0.9936)(0.2034) = 0.20$$

Por lo tanto, el intervalo de confianza al 95% es:

$$1.39 \pm 1.96(0.20) = 1.39 \pm 0.39$$

Así que se tiene un 95% de confianza de que el valor verdadero de g de este estudio se encuentre entre 1.0 y 1.78. Note que, bajo el criterio de Cohen, esta es una magnitud de efecto grande ($g > 0.8$). En una distribución normal, un valor de 1.39 corresponde a una percentila de 91.7%. Es decir, que en este estudio el efecto del tratamiento es tal que aumenta la percentila del estudiante promedio de 50% a casi 92%.

AGREGACIÓN DE RESULTADOS

Una vez se calcule la magnitud del efecto y su correspondiente error estándar para cada estudio, entonces se procede a agregarlos. Esto implica calcular una media y un error estándar global, que son los estimados para los parámetros de la población de estudios. Ambos parámetros se calculan usando factores de ponderación. En esencia, el factor de ponderación lo que hace es darles mayor peso a los estudios con mayor cantidad de observaciones.

El estimado global de la magnitud del efecto y su desviación estándar están dados por:

$$\bar{g} = \frac{\sum g_i \cdot w_i^*}{\sum w_i^*}$$

donde

$$w_i^* = \frac{1}{(SE_g)_i^2 + T^2}$$

y

$$T^2 = \frac{Q - df}{C}$$

$$df = k - 1 \text{ (k = número de estudios)}$$

$$Q = \sum w_i g_i^2 - \frac{(\sum w_i g_i)^2}{\sum w_i}$$

$$C = \sum w_i - \frac{\sum w_i^2}{\sum w_i}$$

$$w_i = \frac{1}{(SE_g)_i^2}$$

El error estándar en el estimado global está dado por:

$$SE_{\bar{g}} = \sqrt{\frac{1}{\sum w_i^*}}$$

y el intervalo de confianza al 95% alrededor del estimado global por:

$$\bar{g} \pm (1.96)SE_{\bar{g}}$$

EJEMPLO: AGREGACIÓN DE RESULTADOS

La **Tabla 2** presenta datos de diez investigaciones primarias sobre la efectividad de los mapas de conceptos como estrategia de aprendizaje en ciencias. La tabla incluye la fuente (D = disertación, A = artículo publicado en una revista arbitrada); si fue (S) o no (NS) significativo el resultado; los tamaños de las muestras de los grupos experimentos y control; y el valor de la magnitud del efecto y los errores estándar SE_d y SE_g , calculados usando el procedimiento presentado en la sección anterior. Note que, de los diez estudios, ocho de ellos no encontraron diferencias significativas entre el tratamiento experimental y el control, por lo que en una revisión de literatura tradicional se inclinaría a pensar que los mapas de conceptos no son una estrategia de aprendizaje efectiva.

Tabla 2

Muestra de estudios sobre el uso de mapas de conceptos como estrategia de aprendizaje en ciencias.

autor	fuentes	sig.	n _b	n _c	g	SE _d	SE _g
Abayomi (1988)	D	NS	97	59	0.25	0.166	0.161
Bodolus (1986)	D	NS	113	185	0.06	0.119	0.116
Franklin (1991)	D	NS	78	67	0.12	0.167	0.160
Gurley (1982)	D	NS	40	35	0.01	0.231	0.214
Heinze-Fry (1987)	D	NS	20	20	0.30	0.318	0.268
Henze-Fry & Novak (1990)	A	Ns	17	20	0.61	0.337	0.274
Huang (1991)	D	NS	76	62	-0.63	0.175	0.168
Jegede, et. al (1990)	A	S	22	29	2.02	0.346	0.297
Lehman, et. al (1985)	A	NS	125	125	0.06	0.127	0.123
Okebukola (1990)	A	S	63	75	1.73	0.200	0.190

Para calcular los valores de Q, C, T2 y w*, se necesita calcular primero los valores de w. Para el estudio de Abayomi tenemos que:

$$w_1 = \frac{1}{(SE_g)_1^2} = \frac{1}{0.161^2} = 38.81$$

La **Tabla 3** presenta los valores de w, así como las cantidades wg, wg² y w², necesarias para completar los cálculos. Usando los resultados de la **Tabla 3**, resulta que:

$$Q = \sum w_i g_i^2 - \frac{(\sum w_i g_i)^2}{\sum w_i} = 152.54 - \frac{(83.65)^2}{340.79} = 132.01$$

$$C = \sum w_i - \frac{\sum w_i^2}{\sum w_i} = 340.79 - \frac{15796.34}{340.79} = 320.26$$

$$T^2 = \frac{Q - df}{C} = \frac{132.01 - 9}{320.26} = 0.3841$$

La **Tabla 3** presenta los valores de w , así como las cantidades wg , wg^2 y w^2 , necesarias para completar los cálculos. Usando los resultados de la **Tabla 3**, resulta que:

Tabla 3

Resultados parciales necesarios para el cálculo de Q.

autor	d	SE _d	SE _g	w	(w)(g)	(w)(g ²)	w ²	
Abayomi (1988)	0.25	0.166	0.161	38.81	9.70	2.425	1505.884	
Bodulus (1986)	0.06	0.119	0.116	74.03	4.44	0.267	5481.078	
Franklin (1991)	0.12	0.167	0.160	38.95	4.67	0.561	1517.350	
Gurley (1982)	0.01	0.231	0.214	21.91	0.22	0.002	479.923	
Heinze-Fry (1987)	0.30	0.318	0.268	13.94	4.18	1.255	194.454	
Henze-Fry & Novak (1990)	0.61	0.337	0.274	13.30	8.12	4.951	177.021	
Huang (1991)	-0.63	0.175	0.168	35.32	-22.25	14.017	1247.189	
Jegade, et. al (1990)	2.02	0.346	0.297	11.35	22.92	46.305	128.781	
Lehman, et. al (1985)	0.06	0.127	0.123	65.61	3.94	0.236	4304.419	
Okebukola (1990)	1.73	0.200	0.190	27.57	47.70	82.521	760.238	
				Σ =	340.79	83.65	152.54	15796.34

Con el valor calculado para T^2 se puede entonces calcular w_i^* para cada estudio. Por ejemplo, para el estudio de Abayomi:

$$w_1^* = \frac{1}{(SE_g)_1^2 + T^2} = \frac{1}{0.161^2 + 0.3841} = 2.44$$

La **Tabla 4** presenta los resultados de g , w^* y gw^* para los 10 estudios de este ejemplo.

Tabla 4

Valores de g , w^ y gw^* para los diez estudios sobre mapas de conceptos.*

autor	g	w^*	gw^*
Abayomi (1988)	0.25	2.439876	0.609969
Bodulus (1986)	0.06	2.515124	0.150907
Franklin (1991)	0.12	2.440457	0.292855
Gurley (1982)	0.01	2.327017	0.02327
Heinze-Fry (1987)	0.3	2.193948	0.658184
Henze-Fry & Novak (1990)	0.61	2.177475	1.32826
Huang (1991)	-0.63	2.424809	-1.52763
Jegede, et. al (1990)	2.02	2.117713	4.27778
Lehman, et. al (1985)	0.06	2.504198	0.150252
Okebukola (1990)	1.73	2.378938	4.115563
$\Sigma =$		23.51956	10.07941

Usando los resultados de la **Tabla 4**, se puede proceder a calcular el estimado poblacional de la magnitud del efecto:

$$\bar{g} = \frac{\sum g_i \cdot w_i^*}{\sum w_i^*} = \frac{10.08}{23.52} = 0.43$$

De acuerdo con el criterio de Cohen, esta magnitud de efecto es mediana. El error estándar de este estimado está dado por:

$$s_{\bar{g}} = \sqrt{\frac{1}{\sum w_i^*}} = \sqrt{\frac{1}{23.52}} = 0.21$$

Así que el intervalo de confianza al 95% para la magnitud del efecto será:

$$\mu_g = \bar{g} \pm 1.96s_{\bar{g}} = 0.43 \pm (1.96)(0.21) = 0.43 \pm 0.41 = (0.02, 0.83)$$

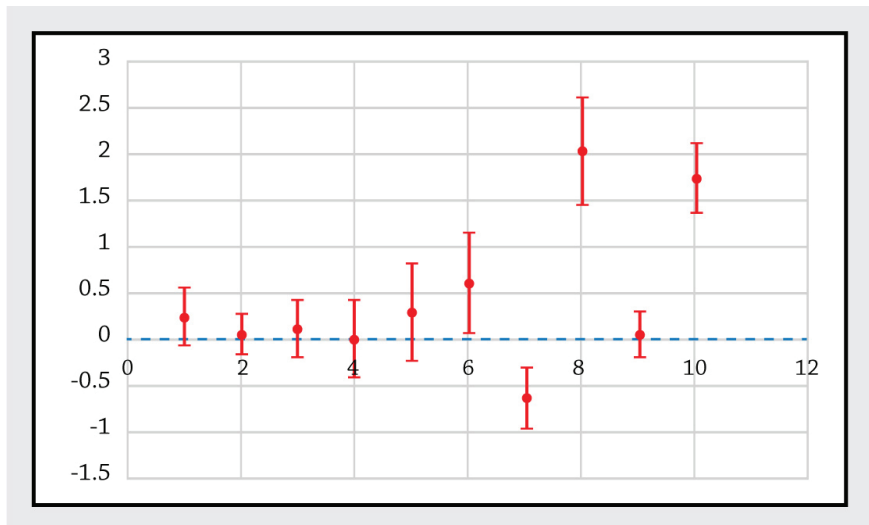
Note que el intervalo de confianza no incluye el valor de 0, por lo que podría decirse que existe un efecto significativo a favor del uso de mapas de conceptos como estrategia de aprendizaje. Esta conclusión es la opuesta a la que se llegó cuando simplemente se determinó la proporción de estudios con resultados no significativos (8 de 10).

ANÁLISIS DE HETEROGENEIDAD

Aunque en este ejemplo se encuentra una magnitud de efecto grande, cabe preguntarse si los resultados de los estudios son o no son homogéneos. Se podría tener una idea haciendo una gráfica de g para cada estudio, con su intervalo de confianza (véase la **Figura 1**). Además, se podría utilizar el valor de Q calculado anteriormente para realizar una prueba estadística de heterogeneidad. Como se puede apreciar en la **Figura 2**, los resultados parecen ser bastante heterogéneos: en seis estudios el intervalo de confianza al 95% incluye el valor de cero; en tres, la magnitud del efecto es positiva y grande ($g > 0.5$); finalmente, hay un estudio en el cual la magnitud del efecto es negativa, es decir, la media del tratamiento de control es mayor que la del tratamiento experimental.

Figura 2

Gráfica de heterogeneidad para los estudios sobre uso de mapas de conceptos.



Para determinar si la heterogeneidad es estadísticamente significativa, se hace uso de la estadística Q . Aunque la fórmula presentada anteriormente es una forma computacional, la fórmula básica se basa en la diferencia entre los valores de g individuales y el promedio del grupo. Así que mientras mayor sea la dispersión de los valores de g , mayor será el valor de Q . La estadística Q se distribuye como una estadística χ^2 con $k - 1$ grados de libertad, donde k es el número de estudios en la síntesis. En este caso, $k - 1 = 9$, por lo que el valor crítico para $\alpha = 0.05$ es de χ^2 es 3.325. Como el valor encontrado para Q es mayor que este valor crítico ($132.01 > 3.325$), hay que rechazar la hipótesis nula de no diferencia entre los valores de g y concluir que los resultados son heterogéneos, lo que confirma el análisis gráfico realizado anteriormente.

RELACIONES ENTRE VARIABLES

Cuando hay heterogeneidad en los resultados cabe preguntarse a qué se debe la misma. Es en este punto en el que las características de los estudios que han sido codificadas pueden ser utilizadas en el análisis. En la **Tabla 2** se puede ver que los estudios se dividen en dos grupos: seis disertaciones y cuatro estudios publicados en revistas arbitradas. En las seis disertaciones los resultados fueron no significativos, mientras que en el 50% de los estudios publicados en revistas arbitradas los resultados fueron estadísticamente significativos. Se podría entonces explorar si existe relación entre la magnitud del efecto y el tipo de publicación.

La estrategia consiste en realizar el equivalente a un ANOVA, en donde las magnitudes de los efectos se dividen en dos grupos (publicados en revistas arbitradas o disertaciones) y determinar si la diferencia en los promedios de estos dos grupos es significativamente distinta de cero. El ANOVA que se realizará hace uso nuevamente de la estadística Q , la cual se determina para cada uno de los grupos en los que se dividió la muestra de estudios. Esto último requiere el cálculo de g promedio para cada grupo, así como las cantidades w^* y gw^* . Una vez obtenidos los valores de Q para cada grupo, se puede calcular la estadística Q_B , la cual es un estimado de la varianza entre los grupos:

$$Q_B = \sum_{i=1}^p Q_i$$

donde p es el número de grupos. El valor de Q_b se compara con el valor crítico de χ^2 para $p - 1$ grados de libertad. Si esta estadística resulta significativa, se podrá concluir que la variable utilizada para hacer los grupos explica (alguna porción) de la heterogeneidad.

El análisis de heterogeneidad puede ser repetido nuevamente para determinar si los resultados en cada grupo son o no son homogéneos. De resultar heterogéneo algún grupo, entonces se puede repetir el procedimiento anterior para explorar la relación entre la magnitud del efecto y alguna otra característica relevante del estudio.

EJEMPLO: ANÁLISIS CUANTITATIVO DE LA HETEROGENEIDAD

La **Tabla 5** muestra los resultados de w^* y gw^* para los dos grupos de estudios construidos utilizando el tipo de publicación. En el grupo de estudios publicados en revistas arbitradas el valor promedio de g es de 1.09, mientras que en el de las disertaciones es de 0.0018.

Tabla 5

Resultados parciales para el cálculo de la estadística Q para cada grupo de estudio.

autor	fuelle	g	SE g	w	wg	wg^2
Henze-Fry & Novak (1990)	A	0.61	1.084	8.7833	5.3578	3.2683
Jegede, et. al (1990)	A	2.02	1.087	8.3374	16.8416	34.0200
Lehman, et. al (1985)	A	0.06	1.038	62.4719	3.7483	0.2249
Okebukola (1990)	A	1.73	1.049	24.9687	43.1958	74.7287
		$\Sigma=$	1.443	104.5613	69.1435	112.2419
autor	fuelle	g	SE g	w	wg	wg^2
Abayomi (1988)	D	0.25	0.3193	36.4183	9.1046	2.2761
Bodulus (1986)	D	0.06	0.2979	70.1213	4.2073	0.2524
Franklin (1991)	D	0.12	0.3198	35.9770	4.3172	0.5181
Gurley (1982)	D	0.01	0.3578	18.6664	0.1867	0.0019
Huang (1991)	D	-0.63	0.3243	32.5468	-20.5045	12.9178
Heinze-Fry (1987)	D	0.30	0.4190	9.8888	2.9666	0.8900
		$\Sigma=$	2.0381	203.6186	0.2779	16.8563

Utilizando los datos de la **Tabla 5** se encuentra que:

$$Q_1 = \sum w_i g_i^2 - \frac{(\sum w_i g_i)^2}{\sum w_i} = 112.24 - \frac{69.14^2}{104.56} = 66.52$$

$$Q_2 = \sum w_i g_i^2 - \frac{(\sum w_i g_i)^2}{\sum w_i} = 16.86 - \frac{0.28^2}{203.62} = 16.86$$

Por lo que:

$$Q_B = \sum_{i=1}^p Q_i = Q_1 + Q_2 = 66.52 + 16.86 = 83.38$$

El valor crítico, para $\alpha = 0.05$ y con $k - 1 = 2 - 1 = 1$ grados de libertad es de 0.0039. Como el valor calculado es mayor que el crítico, se concluye que existe una diferencia significativa entre el promedio de la magnitud del efecto en el grupo de artículos publicados en revistas arbitradas y promedio de las disertaciones. Como se observa, los estudios publicados en revistas arbitradas muestran una magnitud de efecto alta, mientras que en las disertaciones la magnitud del efecto es prácticamente nula.

RESUMEN

El metanálisis es una técnica muy útil para la síntesis de investigaciones. Una síntesis rigurosa a través del metanálisis puede ser utilizada como tema de una disertación doctoral, o una publicación en una revista arbitrada. También puede ser útil para complementar una revisión de literatura tradicional, siempre y cuando los estudios sean cuantitativos. En esta metodología, es de suma importancia definir la población bajo estudio y crear una guía de codificación de los estudios que sea comprensiva. Una vez identificados los estudios, se procede a calcular las magnitudes de efectos para luego agregarlas y obtener un estimado global. El análisis de heterogeneidad puede indicar si hay consistencia en los resultados. De no haberlas, se puede utilizar alguna de las variables codificadas para explicar la heterogeneidad.

Como se vio en el ejemplo utilizado, el análisis de literatura tradicional, basado en determinar la proporción de estudios con

resultados significativos, puede dar una visión sesgada sobre la efectividad de un tratamiento. Como el metanálisis considera la precisión (o el tamaño de muestra) de los estudios, este provee un estimado no sesgado de la magnitud del efecto global. De esta manera, el metanálisis resulta ser una mejor técnica para la síntesis de las investigaciones.

Aunque existen paquetes estadísticos para realizar el metanálisis, los cálculos son lo suficientemente sencillos como para llevarlos a cabo utilizando una calculadora de mano o una hoja de cálculo, tal como Excel.

BIBLIOGRAFÍA

- Abayomi, B. J. (1988). *The effects of concept mapping and cognitive style on science achievement* [Doctoral dissertation, Georgia State University]. Dissertation Abstracts International, 4 9 (6), 1420A.
- Aziz, T. y Rahman, A. (2015, Sept.). *Effect of Concept Mapping Strategy on Students' Achievement in Science at Secondary Level*. Obtenido de https://www.researchgate.net/publication/281848894_Effect_of_Concept_Mapping_Strategy_on_Students'_Achievement_in_Science_at_Secondary_Level
- Bodolus, J. E. (1986). *The use of concept mapping strategy to facilitate meaningful learning for ninth grade students in science*. (Doctoral dissertation, Temple University). Dissertation Abstracts International, 4 7 (9), 3387A.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., y Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. John Wiley and Sons, Ltd.
- Cohen, J. (1988). *Statistical Power Analysis* 2nd Ed. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cooper, H. (2017). *Research Synthesis and Meta-Analysis: A Step-by-Step Approach*. Los Angeles, CA: Sage.
- Franklin, C. E. (1991). *An experiment testing the effects of concept mapping on science anxiety and acquisition of scientific knowledge among eighth grade students low in integrative complexity*. (Doctoral dissertation, The Ohio State University). Dissertation Abstracts International, 2 5 (5), 1689A.

- Glass, G. V., McGaw, B. y Smith, M. L. (1987). *Meta-Analysis in Social Research*. Beverly Hills, CA: Sage Publications.
- Gurley, L. A. (1982). *Use of Gowin's vee and concept mapping strategies to teach students' responsibility for learning in high school biological sciences* (Doctoral dissertation, Cornell University, 1982).
- Hedges, L. V., Shymansky, J. A., y Woodworth, G. (1989). *Modern Methods of Meta-Analysis: A Practival Guide*. Washington, D.C.: National Science Teachers Association.
- Hedges, L. V., y Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. New York, NY: Academic Press, Inc.
- Heinze-Fry, J. A. & Novak, J. D. (1990). *Concept mapping brings long-term movement toward meaningful learning*. *Science Education*, 74 (4), 461-472.
- Heinze-Fry, J.A. (1987). *Evaluation of concept mapping as a tool for meaningful education of college biology students*. (Doctoral dissertation, Cornell University, 1987). *Dissertation Abstracts International*, 48, A D 87-0889 8.
- Huang, W. (1991). *Concept mapping and chemistry achievement, integrated process skills, logical thinking abilities, and gender at teachers' college in Taiwan*. (Doctoral dissertation, Florida Institute of Technology, 1991, Melbourne, Fl.)
- Jegede, O. J., Alaiyemola, F. T. & Okebukola, P. A. (1990). "The effect of concept mapping on students' anxiety and achievement in biology". *Journal of Research in Science Teaching*, 27 (10), 951-960.
- Lehman, J. D., Carter, C. & Kahle, J. B. (1985). "Concept mapping, vee mapping, and achievement: Results of a field study with black high school students". *Journal of Research in Science Teaching*, 22 (7), 663-673.
- Okebukola, P. A. (1990). "Attaining meaningful learning of concepts of genetics and ecology: An examination of the potency of concept mapping technique". *Journal of Research in Science Teaching*, 27 (5), 493-504.

Consideraciones en el diseño de estudios tipo encuesta (muestreo)

*José Noel Caraballo, Ph.D.
asesor del Centro de Apoyo Estadístico
Pontificia Universidad Católica de Puerto Rico*

INTRODUCCIÓN

En disciplinas como Educación, Psicología, Salud pública, Epidemiología, Sociología y Mercadeo, entre otras, a menudo se realizan encuestas para determinar algún parámetro de interés en una población, tal como la intención de votar en una elección o la satisfacción con algún producto comercial. Para que los estimados obtenidos de estos parámetros sean fiables, se parte del supuesto de que la muestra seleccionada es representativa de la población de interés. Es decir, si en una encuesta se encuentra que cierta proporción de la población, por ejemplo 0.55 (55%) tiene intención de votar por el partido político X en la próxima elección, queremos creer que ese será el comportamiento de la población general. Es necesario entonces asegurar que, en efecto, la muestra seleccionada para el estudio represente a la población.

En este escrito se discutirán varios de los métodos que tiene disponibles el investigador para obtener una muestra representativa, incluyendo sus ventajas y limitaciones. Además, se discutirá cómo determinar la cantidad de sujetos (u objetos) que se requiere para obtener un estimado preciso del parámetro de interés. Se discutirá también cómo calcular los pesos de muestreo en los diseños complejos y se ilustrará con un ejemplo obtenido de la literatura.

¿POR QUÉ SE MUESTREA?

En muchas ocasiones no es factible, por razones de logística o de costos, obtener información de cada persona de la población. Por ejemplo, si se quisiera determinar cómo votará la población en Puerto Rico en las próximas elecciones, ¿se contaría con los recursos económicos, tecnológicos y humanos para entrevistar a cerca de dos millones de votantes? Y si estuvieran disponibles estos recursos, ¿sería posible localizar a cada uno de los votantes y lograr entrevistarlos? La respuesta es que, en la inmensa mayoría de las situaciones, una o ambas de estas razones evita que se pueda entrevistar a toda una población. Incluso el Censo General de los Estados Unidos, que, por definición, trata de obtener información sobre toda la población estadounidense, nunca puede lograr que el 100% de las familias contesten su cuestionario.

El muestreo, sin embargo, permite obtener un estimado, con un margen de error razonable, de los parámetros de una población. Bien utilizado, el método de muestreo permitirá encontrar estimados no sesgados a un costo muchísimo menor de lo que costaría auscultar a toda una población.

¿QUÉ SE BUSCA ESTIMAR?

El muestreo se lleva a cabo con el propósito de estimar algún parámetro poblacional. Los parámetros más comunes son medias (o promedios aritméticos), proporciones (o prevalencias) y totales. Por ejemplo, se puede llevar a cabo una encuesta para determinar la media y desviación estándar del ingreso anual de una población, calcular la prevalencia de personas que sufren de asma, o calcular el total de personas deambulantes en una ciudad. Sin embargo, es posible estimar otros parámetros tales como correlaciones, *odds ratios*, coeficientes de regresión, o incluso, pesos factoriales. Para cada uno de estos parámetros existe un estimador.

¿CÓMO SE REPORTA UN RESULTADO DE UNA ENCUESTA?

Antes de entrar en los detalles de cómo seleccionar una muestra, es importante tener claro cómo se reporta el resultado de una encuesta.

Típicamente esto se hace mediante un intervalo de confianza, tal como:

$$\hat{\theta} \pm \lambda_{1-\alpha} \hat{\sigma}_{\theta}$$

En esta ecuación $\hat{\theta}$ es el estimado del parámetro, $\hat{\sigma}_{\theta}$ es el error estándar del estimado del parámetro y $\lambda_{1-\alpha}$ es el valor de la estadística de prueba el cual depende de los grados de libertad asociados al error estándar y al nivel de confianza deseado. Los niveles de confianza más usados corresponden a niveles de significación α de 0.05 y 0.01: $1 - 0.05 = 0.95$ o 95% y $1 - 0.01 = 0.99$ o 99%.

Por ejemplo, suponga que la prevalencia de Hepatitis C en las prisiones de Puerto Rico fuese de 0.30, con un error estándar de 0.015. Entonces, asumiendo una distribución normal, el intervalo de confianza al 95% estaría dado por:

$$0.30 \pm (1.96)(0.015) = 0.30 \pm 0.029$$

Es decir, que hay 95% de confianza de que la prevalencia de Hepatitis C en las prisiones de Puerto Rico se encuentre entre 0.27 y 0.33. Note la importancia de informar no solo el promedio, sino también el margen de error. Sin el último, el estimado de la prevalencia no nos da información útil alguna. Suponga, que el error fuese de 0.30. Entonces el intervalo de confianza sería de 0 a 0.60, es decir, entre ningún caso de Hepatitis C y 60% de la población penal con Hepatitis C. Este segundo ejemplo tiene el problema de ser poco preciso como para ser útil, posiblemente debido a que el tamaño de la muestra seleccionado no era adecuado para tener una precisión aceptable.

DISEÑO DE ESTUDIOS TIPO ENCUESTA

Entre las consideraciones al momento de diseñar un estudio tipo encuesta se encuentran las siguientes: el método de muestreo, el nivel de confianza al que se desean reportar los resultados, el margen de error deseado en los estimados y el tamaño de la muestra. Existen varias opciones en cuanto al método de muestreo a utilizar. En este escrito se considerarán solo aquellas que se basen en un muestreo

probabilístico, es decir instancias en las que las probabilidades de selección de los individuos que componen la población sean conocidas. Estas opciones incluyen el muestreo aleatorio simple, el muestreo sistemático, el muestreo estratificado, el muestreo por conglomerados, o, combinaciones de estos. En las próximas secciones se discutirán estos métodos de muestreo y sus ventajas y desventajas.

El nivel de confianza fue discutido en la sección anterior. Típicamente se hace uso de dos niveles, 95% y 99%, aunque es posible utilizar cualquier otro. Recuerde que el nivel de confianza dice cuán confiados se está de que el intervalo que se ha construido captura el valor del parámetro de interés.

El margen de error (el término que se suma o resta del estimado para construir el intervalo de confianza) dice cuán preciso se desea que sea el estimado. De esta manera, la precisión deseada determina cuán ancho será el intervalo de confianza. Piense en el margen de error como las divisiones más pequeñas de una regla. Si la separación entre estas es grande, habrá menos precisión en una medida que si la separación es pequeña. La precisión es una cantidad que establece el investigador de acuerdo con su criterio, sin embargo, cabe señalar que mientras mayor es la precisión deseada mayor será el tamaño de la muestra que hará falta en el estudio.

Finalmente, el tamaño de la muestra depende del margen de error, el nivel de confianza y la desviación estándar de la variable que se desea estimar. Como se verá más adelante, en diseños complejos, el tamaño de la muestra está influenciado también por su nivel de complejidad. De los primeros tres factores, solo el último no es seleccionado *a priori* por el investigador. La desviación estándar de la variable de interés típicamente hay que obtenerla de una investigación previa o de un estudio piloto. La excepción es cuando el parámetro que se desea estimar es una proporción. En tales casos se utiliza la desviación estándar máxima, la que es sencillamente igual a 0.25.

MÉTODOS DE MUESTREO

MUESTREO ALEATORIO SIMPLE

El muestreo aleatorio simple es, conceptualmente, la forma más básica para obtener una muestra probabilística. Este se basa en el principio de que todos los elementos de la población tienen igual probabilidad de ser seleccionados en la muestra. Para ello, todo lo que hay que hacer es preparar una lista de todas las unidades en la población, enumerarlas y seleccionar una muestra de tamaño n mediante un proceso aleatorio. Por ejemplo, suponga que hay N unidades en la población. Entonces se puede colocar N fichas (numeradas de 1 a N) en una tómbola, agitarlas, y, sin mirar en la tómbola, sacar una a la vez hasta obtener el tamaño de muestra deseado. Otra forma de obtener una muestra aleatoria es usando un programa de computadoras. El programa Excel, por ejemplo, tiene una función, `RAND()`, que genera un número aleatorio (pseudoaleatorio) entre 0 y 1, que puede ser utilizado para obtener una muestra aleatoria.

Observe que este método es sumamente sencillo, pero parte del supuesto de que se tiene, o se puede obtener, la lista de todas las unidades en la población. Este es un gran problema, pues no siempre se tiene a la mano la lista de todas las unidades de la población. Por ejemplo, en el sistema de educación pública de Puerto Rico no existe una lista actualizada de todos sus estudiantes, así que es imposible hacer un muestreo aleatorio simple de ellos. Aun cuando se tuviese dicha lista, el procedimiento sería poco eficiente, pues si se seleccionara una muestra de 100 estudiantes, habría que recorrer todo el país para entrevistarlos.

El muestreo aleatorio simple tiene otra desventaja: no hay forma de asegurar que la muestra seleccionada sea representativa de la población con respecto de alguna característica en particular. Por ejemplo, suponga que en una escuela hay 400 alumnos, 200 hembras y 200 varones y que se desea seleccionar una muestra de tamaño $n = 20$. Mediante un proceso aleatorio simple es posible que los 20 alumnos seleccionados sean de un mismo sexo, o mayoritariamente de un sexo (por ejemplo, 15 hembras y 5 varones). Si la variable que se mide está relacionada con el sexo del encuestado, entonces esta muestra podría considerarse como

no representativa de la población y los estimados obtenidos estarían sesgados (o viciados). Los métodos que se discutirán a continuación permiten controlar mejor estas variables, de manera que sea menos probable que la muestra no sea representativa.

MUESTREO SISTEMÁTICO

El muestreo sistemático es otra forma sencilla de obtener una muestra. Suponga que usted tiene una lista de los N elementos de la población y desea obtener una muestra de tamaño n . Divida la lista en segmentos de tamaño N/n . Luego seleccione al azar un elemento del primer segmento usando muestreo aleatorio simple. Una vez obtenga ese elemento, seleccione en todos los demás segmentos el elemento que está en la misma posición que estuvo el primero. Por ejemplo, si $N = 1000$ y $n = 10$, $N/n = 100$. Es decir, dividiría la lista en 100 segmentos con 10 elementos. Suponga que en el primer segmento resultó seleccionado al azar el 4to elemento. Entonces, seleccione el cuarto elemento en los restantes 99 segmentos. Con estos completará el tamaño de muestra deseado, 100.

El muestreo sistemático tiene la ventaja de que permite controlar alguna variable de interés. Por ejemplo, suponga que los elementos en cuestión son ciudades. Si la lista se ordena por tamaño de ciudad, que podría ser de mayor a menor, entonces al realizar el muestreo sistemático se asegurará de que los tamaños de las ciudades en la muestra estén en proporción con los tamaños en la población. El muestreo sistemático utilizado de esta forma se le conoce como “estratificación implícita”.

MUESTREO ESTRATIFICADO

El muestreo estratificado permite controlar alguna variable de interés, de manera que la muestra obtenida sea representativa de la población en cuanto a ella. La variable de interés se le conoce como “estrato”, de ahí el nombre de este método de muestreo. Una población de individuos se podría estratificar por su sexo, nivel socioeconómico, edad o preferencia política, entre otras variables. Las ciudades se podrían estratificar de acuerdo con su tamaño, sea geográfico o poblacional. Es posible, incluso, estratificar por más de una variable de forma simultánea.

Volviendo al ejemplo de la escuela con 200 niñas y 200 niños, se podrían seleccionar 10 niñas al azar de la lista de 200 niñas y 10 niños de la lista de 200 niños. Esto aseguraría que la proporción de niñas (o niños) en la muestra sea la misma que en la población: 50%. El muestreo estratificado aquí descrito se conoce por muestreo estratificado proporcional. Existen otras variantes de este método que también permiten obtener estimados no viciados del parámetro de interés: igual tamaño, distribución óptima y distribución óptima por costo. Estos métodos utilizan pesos para compensar cuando las proporciones de los elementos no son las mismas en la población que en la muestra.

El muestreo estratificado tiene el mismo problema que el muestreo aleatorio simple y el sistemático: sigue siendo necesario tener una lista de todos los elementos de la población. Sin embargo, tiene otra ventaja adicional a la descrita anteriormente, provee mayor precisión en los estimados del parámetro de interés. Es decir, los estimados de los errores estándar en el muestreo estratificado serán siempre iguales o menores, pero en ningún caso mayores, que los obtenidos por el muestreo aleatorio simple o sistemático. Esto entonces resultará en intervalos de confianza más estrechos, es decir, más precisos. Además, es posible obtener estimados separados para las categorías de la variable de estratificación. Así que, en el ejemplo anterior, se podrían obtener estimados separados, con sus correspondientes intervalos de confianza, para las niñas y para los niños.

MUESTREO POR CONGLOMERADOS

El último método básico de muestreo es el muestreo por conglomerados. Su nombre proviene de que, en muchas situaciones, las unidades bajo estudio están agrupadas naturalmente. Por ejemplo, los alumnos en un sistema escolar están agrupados por región educativa, distrito, pueblo, escuela, grado, y sección. La idea del muestreo por conglomerados es hacer una lista de los conglomerados y seleccionar un número de ellos al azar. Entonces, dentro de cada conglomerado se pueden encuestar todas las unidades, o, usar alguno de los métodos descritos anteriormente para obtener una muestra.

Note que el muestreo por conglomerados no requiere la lista completa de las unidades en la población: solo se requiere la lista de

unidades dentro de cada uno de los conglomerados seleccionados. Estas listas a menudo se pueden conseguir con más facilidad, o construirse en el momento. Por ejemplo, en el estudio en un sistema escolar, si las escuelas son los conglomerados, entonces lo que se requiere son las listas de estudiantes en las escuelas seleccionadas y no la lista de todos los estudiantes en el sistema escolar del país.

Casi todas las encuestas importantes sobre los habitantes de un país se llevan a cabo utilizando conglomerados por ser viable y por ser más económico. Por ejemplo, si se quiere una muestra aleatoria de 10,000 estudiantes en un país, es más económico visitar un número reducido de escuelas, que viajar por todo el país para entrevistar a los 10,000 estudiantes. En general, el número de conglomerados que se requiere en este tipo de muestreo es de alrededor de 20.

La desventaja del muestreo por conglomerados estriba en que, para un tamaño de muestra dado, los estimados tienen menor precisión que en el muestreo aleatorio simple. Así que, para un nivel de precisión preestablecido, será necesario obtener un tamaño de muestra que pueda ser considerablemente mayor que el que se requeriría en el muestreo aleatorio simple. Sin embargo, en general, el costo de aumentar el tamaño de la muestra es todavía mucho menor que el costo en el que se incurriría si se tuviese que visitar individualmente la muestra aleatoria simple.

DISEÑOS COMPLEJOS

Los métodos básicos de muestreo discutidos anteriormente se pueden combinar en diseños más complejos. Por ejemplo, es posible que se establezcan conglomerados en un primer nivel de muestreo y dentro de cada conglomerado se hagan estratificaciones u otros conglomerados. Por supuesto, la selección de la muestra en el último nivel será mediante muestreo aleatorio simple, al igual que cualquier muestreo por conglomerados que se realice. Esta es la estrategia que se sigue en encuestas a nivel de país, donde hay gran cantidad de elementos a muestrear y se encuentran dispersos geográficamente. En una sección posterior se mostrará un ejemplo en el cual se hace uso de un diseño complejo.

TAMAÑO DE LA MUESTRA

MUESTREO ALEATORIO SIMPLE

Cuando hablamos sobre los intervalos de confianza se indicó que la cantidad $\lambda_{1-\alpha} \hat{\sigma}_\theta$ representaba la precisión deseada. Suponga que el parámetro de interés es una media y que la distribución muestral de esta es normal. Entonces la precisión d será igual a:

$$d = z_{1-\alpha} s_{\bar{X}}$$

o

$$d = z \frac{s_x}{\sqrt{n}}$$

En esta última ecuación, s_x es la desviación estándar de los datos obtenidos en la encuesta y n es el tamaño de la muestra. El valor de z dependerá del nivel de confianza: para 95% será 1.96, mientras que para 99% será igual a 2.33. Si se despeja la ecuación anterior para n , se obtendrá la fórmula básica para el tamaño de muestra:

$$n = \frac{z^2 s^2}{d^2}$$

Si en vez de una media el parámetro a estimar es una proporción, entonces $s = \sqrt{p(1-p)}$, donde p es el estimado de la proporción determinado con la muestra. De esta manera la fórmula para el tamaño de la muestra requerido para estimar una proporción es:

$$n = \frac{z^2 (p(1-p))^2}{d^2}$$

Para hacer uso de estas dos fórmulas el investigador tiene que especificar la precisión deseada, d , y el nivel de confianza (que determina el valor de z). Queda entonces especificar los valores de s en la primera, o el valor de p en la segunda. El valor de s usualmente se consigue de estudios anteriores en los cuales se ha medido la variable de interés (con el mismo instrumento) o realizando un estudio piloto. En el caso de p , aunque tal vez sería posible hacer lo mismo, usualmente se utiliza un valor de 0.5. Esto es así, porque el producto $p(1-p)$ es máximo cuando $p = 0.5$. De esta manera, el valor de n que se obtiene asegura que se obtendrá la precisión deseada. Así que la segunda ecuación se puede reescribir como:

$$n = \frac{z^2(0.25)}{d^2}$$

Estas ecuaciones asumen que la población es grande. En casos en que la población tenga un tamaño pequeño (menor a 1,000) es necesario hacer una corrección. Las fórmulas corregidas para poblaciones finitas son:

$$n = \frac{1}{\left(\frac{d}{zs}\right)^2 + \frac{1}{N}}$$

y

$$n = \frac{1}{\frac{d^2}{0.25z^2} + \frac{1}{N}}$$

donde N es el tamaño de la población.

Por ejemplo, suponga que se desea determinar la prevalencia de caries en la población de niños de escuela elemental con una precisión de 0.03 y un nivel de confianza de 99%. Entonces el tamaño de la muestra requerido sería:

$$n = \frac{(2.33)^2(0.25)}{(0.03)^2} = 1,508$$

Si la población fuese un distrito escolar, en vez de un país, y el tamaño de esta población fuese de 1,000 niños, entonces el tamaño de la muestra sería:

$$n = \frac{1}{\frac{0.03^2}{0.25(2.33)^2} + \frac{1}{1000}} = 601$$

Para estimar un total, las fórmulas son las siguientes:

$$n = \frac{N^2 z^2 s^2}{d^2}$$
$$n = \frac{1}{\frac{d^2}{N^2 z^2 s^2} + \frac{1}{N}}$$

MUESTREO ESTATIFICADO

Para determinar el tamaño de la muestra en el muestreo estratificado es necesario calcular el tamaño de muestra para el muestreo aleatorio simple, como se mostró en la sección anterior, pero además hay que saber el tipo de estratificación que se desea hacer. Cuando se desea un muestreo estratificado proporcional, el número de elementos en cada estrato se calcula de la siguiente forma:

$$n_h = \frac{nN_h}{N}$$

Donde n es el tamaño de la muestra para un muestreo aleatorio simple, N es el tamaño de la población y N_h es el tamaño de la población en el estrato h . En el diseño de muestreo estratificado con distribución óptima, es decir con el tamaño de muestra menor posible, la fórmula es:

$$n_h = \frac{nN_h\sigma_h}{\sum N_h\sigma_h}$$

Note que esta fórmula requiere estimados de las desviaciones estándares dentro de cada estrato. Finalmente, en el muestreo estratificado con costo mínimo la fórmula es:

$$n_h = \frac{(c - c_0)nN_h\sigma_h/\sqrt{c_h}}{\sum N_h\sigma_h/\sqrt{c_h}}$$

donde c_0 es el costo fijo para realizar la encuesta, c_h es el costo para el estrato h y c es el costo total (fijo mas costos de los estratos).

MUESTREO POR CONGLOMERADOS

En el caso del muestreo por conglomerados, las fórmulas para calcular el tamaño de la muestra son más complejas y no se presentarán en este escrito. En todo caso, se puede utilizar el método que se presentará para diseños complejos. El lector interesado podrá consultar algún texto básico como el de Cochran (1977) o el de Lohr (1999).

TAMAÑO DE MUESTRA EN UN DISEÑO COMPLEJO

La estrategia para determinar el tamaño de muestra requerido en un diseño complejo consiste en determinar el tamaño de la muestra para el diseño aleatorio simple y multiplicar este por un factor definido por la complejidad del diseño, llamado *Deff* (*design effect factor*). Para obtener el tamaño de la muestra en un diseño complejo, multiplique el tamaño de muestra para un muestreo aleatorio por el efecto del diseño:

$$n_{dc} = (Deff) \cdot n$$

donde n_{cd} es el tamaño de muestra para el diseño complejo y n es el obtenido usando la fórmula para muestreo aleatorio simple. Para calcular el *Deff* es necesario contar cuántos conglomerados hay en el diseño y estimar su tamaño promedio. Entonces el *Deff* se determina mediante la ecuación:

$$Deff = 1 + (\bar{b} - 1)\rho$$

donde b es el tamaño promedio de los conglomerados y ρ es la correlación intraclase. La correlación intraclase es una medida de cuán correlacionados están los datos dentro de un conglomerado. Esta correlación puede ser obtenida de estudios previos o de estudios pilotos. Otra posibilidad es utilizar un valor de *Deff* basado en la experiencia. Estos valores típicamente son de entre 1.0 a 3.5. Note que un *Deff* igual a 2.5 lo que dice es que se necesitarán dos veces y media el tamaño de muestra calculado para muestreo aleatorio simple. Por ejemplo, si $n = 400$, entonces $n_{dc} = (2.5)(400) = 1,000$.

EL ANÁLISIS DE LOS DATOS DE UNA ENCUESTA

Como se mencionó en la introducción, una encuesta se lleva a cabo para estimar algún parámetro de interés. Aunque en este escrito se habla sobre medias, proporciones y totales, es posible estimar los coeficientes de un modelo de regresión, los efectos en un diseño experimental, o los pesos factoriales en un análisis factorial, entre otros. Las fórmulas para obtener estos estimados pueden ser bastante complejas, sin embargo, en la era de las computadoras hay numerosos programas para llevarlos a cabo. Entre estos se destacan Stata, SAS, M+, y el módulo de SPSS para análisis de encuestas. Es importante enfatizar que analizar los datos de una encuesta con programas o algoritmos que no tomen en consideración el diseño de muestreo es un error mayor y puede resultar, en la mayoría de los casos, en intervalos de confianza menores que los correctos. Es decir, se puede llegar a concluir que algún efecto es significativo, cuando en realidad la evidencia no lo justifica.

Los programas (o algoritmos) especializados para el análisis de datos de encuestas se caracterizan porque requieren una estructura de datos que incluya, no solo la (o las) variable(s) de interés de la encuesta, sino también variables que especifiquen, para cada elemento encuestado, a qué estrato pertenece, a que conglomerado pertenece y cuál es su peso de muestreo correspondiente. Es decir, para cada récord, línea de datos, o individuo encuestado, tienen que registrarse estas variables y luego ser usadas en las especificaciones del programa para llevar a cabo los análisis.

DETERMINACIÓN DE LOS PESOS DE MUESTREO

Como se indicó en la sección anterior, para el análisis correcto de los datos de una encuesta es necesario, en la mayoría de los casos, determinar cuánto peso se le dará a cada observación realizada. El peso es necesario en los casos en los que las probabilidades de selección de las unidades no sean constantes y sean iguales al recíproco de esta probabilidad. La fórmula general para calcular los pesos es:

$$w_{hk} = \frac{\sum Mos_h}{ab}$$

donde w_{hk} es el peso que se le asigna a cada observación en el estrato **h** y el conglomerado **k**, **a** es el número de conglomerados que fueron seleccionados en el estrato **h**, y **b** es el tamaño de la muestra seleccionada en el conglomerado **k**. Finalmente, Mos_h (*measure of size*) se refiere al tamaño del conglomerado **h**.

EJEMPLO: PREVALENCIA DE ALGUNAS CONDICIONES DE SALUD EN LA POBLACIÓN PENAL DE PUERTO RICO

Albizu y otros (2012) llevaron a cabo una encuesta en las prisiones de Puerto Rico con el propósito de estimar la prevalencia de varias condiciones de salud entre los confinados. La **Tabla 1** presenta el diseño de muestreo utilizado, así como los pesos de muestreo requeridos.

El diseño de muestreo fue como sigue: las 32 instituciones fueron estratificadas por el nivel de toxicología en las pruebas de orina que se les realizan rutinariamente a los confinados. Dentro de cada uno de estos estratos se hizo una segunda estratificación por el nivel de seguridad de la institución. Dentro de cada uno de estos estratos se seleccionaron dos instituciones (conglomerados) al azar. Finalmente, dentro de cada conglomerado se seleccionó una muestra de confinados al azar. En la tabla, Mos es el tamaño de la población de confinados en la institución y b es el tamaño de la muestra seleccionada. Además de una variable para indicar el estrato y otra para el conglomerado, la tabla incluye el peso de muestreo. Este peso hay que adjudicárselo a cada uno de los miembros de la muestra de una misma institución.

Como ejemplo, el peso de muestreo de la Institución **A** se determinó de la siguiente manera:

$$w_{11} = \frac{\sum Mos}{ab} = \frac{459 + 1311 + 400 + 530}{(2)(60)} = \frac{2700}{120} = 22.5$$

Note que $a = 2$ ya que se seleccionaron al azar dos conglomerados del total de cuatro que había en el estrato de nivel de seguridad mediana. El caso de la institución **O** es uno especial, pues era un conglomerado único en el nivel de seguridad mínima. Con un solo conglomerado no es factible estimar la varianza debido a los conglomerados, por lo que la estrategia consistió en dividir la muestra aleatoriamente en dos mitades de igual tamaño. Estos dos conglomerados fueron llamados institución O-1 e institución O-2.

La **Tabla 2** presenta los resultados obtenidos en esta encuesta. Para cada condición de salud, la tabla presenta la prevalencia, su error estándar y el intervalo de confianza al 95%. Se presenta además el *Deff* y la correlación intraclase. Note que el *Deff* es distinto para cada variable. Esto es así pues la correlación intraclase es distinta para cada condición de salud. Note también que el *Deff* fue menor que 1.0 en algunos casos, lo que indica que se obtuvo mejor precisión de la que se hubiera obtenido con un muestreo aleatorio simple. Los valores mayores de 1.0 oscilan entre 1.48 y 3.56. Por ejemplo, para estimar la prevalencia de Hepatitis B con este diseño se requiere un tamaño de muestra tres veces y media mayor que con el diseño aleatorio simple.

Tabla 1
Diseño de muestra para el estudio de Albizu y otros (2012).

Prevalencia de toxicología positiva en la orina	Nivel de seguridad	Institución	Mos	b	Estrato	Conglomerado	Peso de muestreo
muy alta	mediana	Institución A	459	60	1	1	22.5000
	mediana	Institución B	1311	53	1	2	25.4717
	mediana	No selecc.	400	--			
	mediana	No selecc.	530	--			
	mínima	No selecc.	209	--			
	mínima	Institución C	190	47	2	1	8.0426
	mínima	Institución D	357	45	2	2	8.4000
	mediana	Institución E	979	46	3	1	14.1413
	mediana	Institución F	214	48	3	2	13.5521
	mediana	No selecc.	108	--			
alta	mínima	No selecc.	446	--			
	mínima	Institución G	149	32	4	1	18.6250
	mínima	No selecc.	200	--			
	mínima	Institución H	397	52	4	2	11.4615
	centro de procesamiento	Institución I	508	44	5	1	8.3977
	centro de procesamiento	Institución J	231	47	5	2	7.8617

Tabla 1 (continuación)

Prevalencia de toxicología positiva en la orina	Nivel de seguridad	Institución	Mos	b	Estrato	Conglomerado	Peso de muestreo
normal	máxima	Institución K	280	53	6	1	9.0189
	máxima	Institución L	388	54	6	2	8.8519
	máxima	No selecc.	288	--			
	mediana	Institución M	973	45	7	1	19.8667
	mediana	No selecc.	323	--			
	mediana	Institución N	160	43	7	2	20.7907
desconocida	mediana	No selecc.	332	--			
	mínima	Institución O-1	77	24	8	1	3.2083
	mínima	Institución O-2	77	24	8	2	3.2083
	mediana	No selecc.	6	--			
	mediana	No selecc.	48	--			
	mediana	Institución P	25	16	9	1	10.2500
	mediana	No selecc.	18	--			
	mediana	Institución Q	225	54	9	2	3.0370
	mediana	No selecc.	6	--			
	mínima	Institución R	129	41	10	1	4.0732
	mínima	Institución S	205	48	10	2	3.4792
	Totales:			10248	876		

Tabla 2*Prevalencias de condiciones de salud en la población penal.*

Condición de salud	Prevalencia	Error estándar linearizado	I.C. al 95%	Deff	ρ
LDAD	0.544	0.0438846	(0.430, 0.626)	2.56	0.0180
MD	0.177	0.0112944	(0.152, 0.202)	0.68	-0.0037
GA	0.029	0.0080064	(0.011, 0.046)	1.81	0.0094
PTSD	0.134	0.0169545	(0.096, 0.172)	1.94	0.0109
AADHD	0.164	0.0161124	(0.128, 0.200)	1.48	0.0055
HBV	0.145	0.023748	(0.092, 0.198)	3.56	0.0296
HCV	0.297	0.028813	(0.233, 0.361)	3.11	0.0244
HIV	0.047	0.0043611	(0.037, 0.057)	0.33	-0.0077

Nota. LDAD – Abuso/dependencia a drogas de por vida, MD depresión generalizada, GA-ansiedad generalizada, PTSD-Síndrome de estrés postraumático, AADHD-déficit de atención con hiperactividad en adultos, HBV-hepatitis B, HBC-hepatitis C, HIV-síndrome de inmunodeficiencia adquirida.

RESUMEN

El diseño de una encuesta requiere planificar cómo se presentarán los resultados: su precisión y nivel de confianza. Requiere, además, diseñar el plan de muestreo. Con estos tres elementos especificados, entonces es factible calcular el tamaño de muestra requerido. Si se va a estimar un promedio, entonces para el cálculo del tamaño de muestra se requerirá conocimiento de la desviación estándar de la variable a ser estimada. Esta información se obtiene usualmente de estudios previos de un estudio piloto. La estimación de una proporción, sin embargo, no tiene este requerimiento, ya que el tamaño de la muestra se puede estimar para una variación máxima.

Hay cuatro formas básicas de muestreo: aleatorio simple, sistemático, estratificado y por conglomerados. Estas formas básicas se pueden combinar en diseños complejos que pueden facilitar el recogido de datos, pero en general requieren tamaños de muestra más grandes que para un muestreo aleatorio simple. El factor multiplicativo se conoce como “efecto del factor de diseño” o *Deff* y generalmente tiene valores entre 1.0 y 3.5. El factor de diseño se puede calcular si se conoce la correlación intraclase para la variable que se desee estimar.

El análisis de los diseños complejos requiere el cómputo de pesos de muestreo. Estos pesos se calculan a base de los conglomerados y requieren conocer la cantidad y el tamaño de los conglomerados dentro de los estratos. Con la información acerca del estrato, el conglomerado y el peso, los datos pueden ser procesados utilizando algún programa que tenga los algoritmos para el análisis de datos de encuestas. Es un error craso analizar datos que fueron obtenidos utilizando muestreo estratificado, por conglomerados o un diseño complejo usando paquetes de estadística básicos, los cuales asumen que los datos se obtuvieron mediante muestreo aleatorio simple.

Finalmente es importante recordar que, como se puede apreciar de la **Tabla 2**, los resultados de una encuesta deben incluir, no solo el estimado del parámetro de interés, sino también el intervalo de confianza alrededor de dicho estimado. Se debe informar, además, el nivel de confianza para el cual se construyó el mismo.

REFERENCIAS

- Albizu-García, C. E., Caraballo, J. N., Caraballo-Correa, G., Hernández-Viver, A., and Román-Badenas, L. (2012). *Assessing Need for Medication Assisted Treatment for Opiate Dependent Prison Inmates. Substance Abuse*, 33, 60-69.
- Cochran, W. G. (1977). *Sampling Techniques*, 3rd. Ed. New York, NY: John Wiley & Sons, Inc.
- Lohr, S. L. (1999). *Sampling: Design and Analysis*. Pacific Grove, CA: Brooks/Cole Publishing Company.

Muestras y estimación de parámetros en poblaciones escasas u ocultas

*Julio César Quintana Díaz, Ph.D.
asesor del Centro de Apoyo Estadístico
Pontificia Universidad Católica de Puerto Rico*

INTRODUCCIÓN

En las investigaciones que se realizan en las Ciencias Sociales y en estudios de la conducta humana generalmente se utiliza la obtención de muestras para estimar los parámetros de interés que tenga una población. Se recurre a este procedimiento cuando el tamaño de la población es tan grande que no es factible para los investigadores acceder a todos los miembros que la componen. Existen métodos rigurosos de muestreo aleatorio que permiten obtener estimados de los parámetros, con una confiabilidad y un margen de error determinados. Dependiendo de la estructura y características de la población, del parámetro a estimarse y del interés del investigador, las técnicas de muestreo que más se aplican son: simple al azar, estratificado, sistemático, y por conglomerados. Además, hay investigaciones donde se puede utilizar más de una de estas técnicas.

Para aplicar estos procedimientos se requiere que el marco muestral asociado a la población esté muy bien definido, pues de ese conjunto de unidades muestrales que lo conforman es que la muestra será seleccionada. El diseño básico “simple al azar” requiere que, del marco muestral, se seleccione un grupo de n unidades de tal modo que cada muestra de tamaño n tenga igual probabilidad de ser seleccionada. Cuando el grupo de individuos con la característica de interés para el investigador representa más del 10% de la población, puede usarse este

diseño o diseños derivados de este. Sin embargo, cuando este grupo es muy pequeño, comparado con la población (menos de un 10%), entonces se considera que el grupo es una subpoblación escasa (rara) y se hace muy difícil obtener de ella una muestra simple aleatoria representativa. Es aquí donde se hace necesario implantar métodos no probabilísticos de muestreo (con fundamento matemático y estadístico) para obtener estimados de parámetros de ese tipo de población. Ejemplos de estas técnicas de muestreo son: por bola de nieve, redes, rastreo de enlaces, dirigido por el que contesta (*Respondent-Driven Sampling*); por ruta aleatoria (*random-walk*) o redes egocéntricas de datos.

Existe también otro tipo de poblaciones, las denominadas “ocultas”, cuyos individuos poseen por lo menos una de las siguientes características: a) son estigmatizados socialmente o penalizados; b) ocultan su comportamiento, por lo que se dificulta el acceso y su retención para poder realizar investigaciones; c) realizan actividades ilegales o que se alejan de la corriente principal de la sociedad; y d) por su naturaleza particular, no conforman un marco muestral definido. En estos casos, las estimaciones que se obtengan por los métodos tradicionales de muestreo son muy imprecisas porque, ante una pregunta embarazosa o sensitiva, un individuo con estas características podría no contestarla o contestarla falsamente, aunque se le hayan provisto todas las garantías de confidencialidad y anonimato.

Además de los métodos no probabilísticos mencionados anteriormente, existen unos métodos indirectos de estimación basados en muestras aleatorias. El propósito de estas técnicas es introducir procedimientos matemáticos y estadísticos que les brinden a todos los participantes (estigmatizados o no) la garantía de que las respuestas que cada uno dé a las preguntas sensitivas o embarazosas no serán conocidas por el investigador. Los métodos indirectos más conocidos son: respuesta aleatorizada; pareo de listas incompletas; y diseño de respuestas en bloques aleatorios incompletos. Ocurren situaciones donde estas subpoblaciones tienen ambas características: escasas y ocultas.

En este ensayo se propone exponer, en forma resumida, estas metodologías con el propósito de contribuir al acervo de técnicas no tradicionales de muestreo o estadísticas que la facultad o los

estudiantes graduados de la Institución puedan utilizar para realizar sus investigaciones sobre temas sensitivos o embarazosos.

MÉTODOS DE MUESTREO Y ESTIMACIÓN EN POBLACIONES ESCASAS

MÉTODOS DE MUESTREO POR REDES, REFERENCIAS O ENLACES

De acuerdo con lo expuesto en la Introducción, las subpoblaciones escasas tienen una característica que solo el 10% o menos de la población general comparte, lo que dificulta obtener una muestra representativa aplicando los métodos usuales de muestreo aleatorio. Existe una familia de métodos de muestreo llamada “muestreo por referencia” para acceder a ellas y obtener estimaciones, siendo los principales: bola de nieve, rastreo de enlaces, redes dirigidas por los que responden, redes egocéntricas y ruta aleatoria. Todos estos métodos comparten la característica de que el investigador comienza el proceso con la selección de una muestra por conveniencia, muy pequeña, de participantes que tengan la característica de interés.

Los métodos que se presentarán aquí con más detalle son: a) muestreo por rastreo de enlaces (versión mejorada del muestreo por bola de nieve); b) por multiplicidad; y c) redes dirigidas por los que responden (*Respondent-Driven Sampling*, RDS).

MUESTREO POR “BOLA DE NIEVE” O RASTREO DE ENLACES

Según se indicó, este método tiene la estructura de un muestreo por conveniencia y, en consecuencia, no se pueden hacer inferencias a la población a partir de los estimados. Es particularmente útil en estudios exploratorios. Al final de la discusión de los métodos de muestreo por redes o enlaces se presentarán sus ventajas y desventajas.

El método de muestreo por bola de nieve fue introducido por JS. Coleman (1958) y L. Goodman (1961) como un procedimiento para estudiar redes y organizaciones sociales, siendo el trabajo de Goodman el más conocido y donde se acuñó el nombre del método. En su forma original no estaba orientado a estudiar poblaciones ocultas o escasas,

pues las conexiones que se analizaban ocurrían en poblaciones con marcos muestrales definidos.

El muestreo por rastreo de enlaces (versión mejorada del de bola de nieve) usa muestreo probabilístico para estudiar las redes sociales. (Heckathorn y Cameron, 2017).

Existen tres formas de muestreo por bola de nieve: a) lineal, b) exponencial no discriminatorio, y c) exponencial discriminatorio. En las dos primeras modalidades el objetivo no es estimar la prevalencia de la característica, porque todos en la muestra la tienen, sino crear un perfil (con otra información) de la subpoblación de interés. En el muestreo lineal se selecciona una persona que tenga la característica de interés y se le pide que refiera a una segunda persona con la característica y esta refiere a una tercera y así sucesivamente hasta completar la muestra. En la técnica exponencial no discriminatoria, se inicia nuevamente con una persona con la característica y se le pide que refiera a dos personas (si es exponencial de base 2) que posean la característica y a cada una de ellas se les pide que refieran dos más y así sucesivamente hasta completar el tamaño de la muestra especificada por el investigador. En la modalidad del muestreo exponencial discriminatorio se inicia con una persona que tenga la característica y se le pide que refiera a dos personas, donde por lo menos una tenga la característica. Luego la segunda persona que tenga la característica refiere a dos más, con la misma condición. Los únicos que refieren son los que posean la característica. La ventaja de esta versión es que se pueden crear dos perfiles, el de la subpoblación de interés y el de la subpoblación sin la característica y hacer comparaciones entre ambos. En todos estos procesos se lleva un registro de quién refirió a quién y no se incluyen los referidos recíprocos. A continuación, se presenta un ejemplo de la implantación del muestreo por bola de nieve estratificado.

Sarmiento, W. (tesis no publicada, 2008) demostró, utilizando simulaciones, teoría de grafos y aplicación de algoritmos para obtener estimadores bayesianos de proporciones, que el muestreo por bola de nieve produjo los estimados más precisos y con MSE más pequeños, de la prevalencia de HIV en una subpoblación de adultos al replicar un estudio de Colorado Springs.

Comparó estos estimados con los convencionales. Se resumen aquí algunos de los resultados más importantes. En la **Tabla 1** se presentan los estimados de esta prevalencia al usar un diseño de bola de nieve estratificado, un diseño de bola de nieve simple y el diseño simple al azar. Nótese que la prevalencia real en la población fue de 22.35%.

Tabla 1

Estimados de la prevalencia de HIV en individuos heterosexuales en Colorado Springs (), por diseño de muestreo.*

estimador de la proporción de personas con HIV por tipo de diseño muestral	estimado de la proporción	error cuadrado promedio
Bola de nieve estratificado	0.2225	0.0021
Bola de nieve	0.2221	0.0021
Simple al azar	0.015	0.0435
Proporción en la población	0.2235	

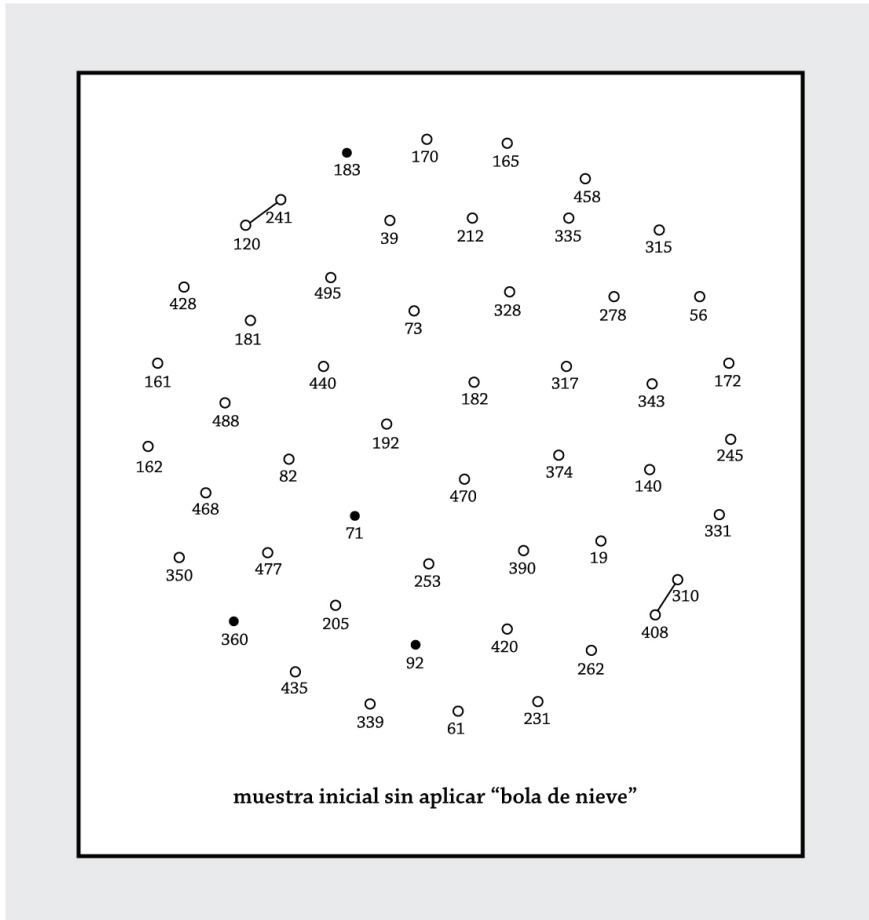
Nota. W. Sarmientos, tesis no publicada, UPRM, 2008

Puede observarse que el método simple al azar produce un estimado muy bajo (0.015) y que los dos estimados por bola de nieve difieren de la prevalencia real en poco menos de una milésima.

En un segundo ejemplo, W. Sarmiento, 2008, compara la proporción estimada de elementos con una característica de interés en una muestra de 50 utilizando primero muestreo simple al azar y luego incrementando la muestra a un total de 133 elementos usando redes de enlace. Las figuras y la **Tabla 2** correspondientes se presentan a continuación.

Figura 1

Estimación de individuos con la característica de interés, en una muestra simple al azar de tamaño 50.



Nota. W. Sarmientos, tesis no publicada, UPRM, 2008

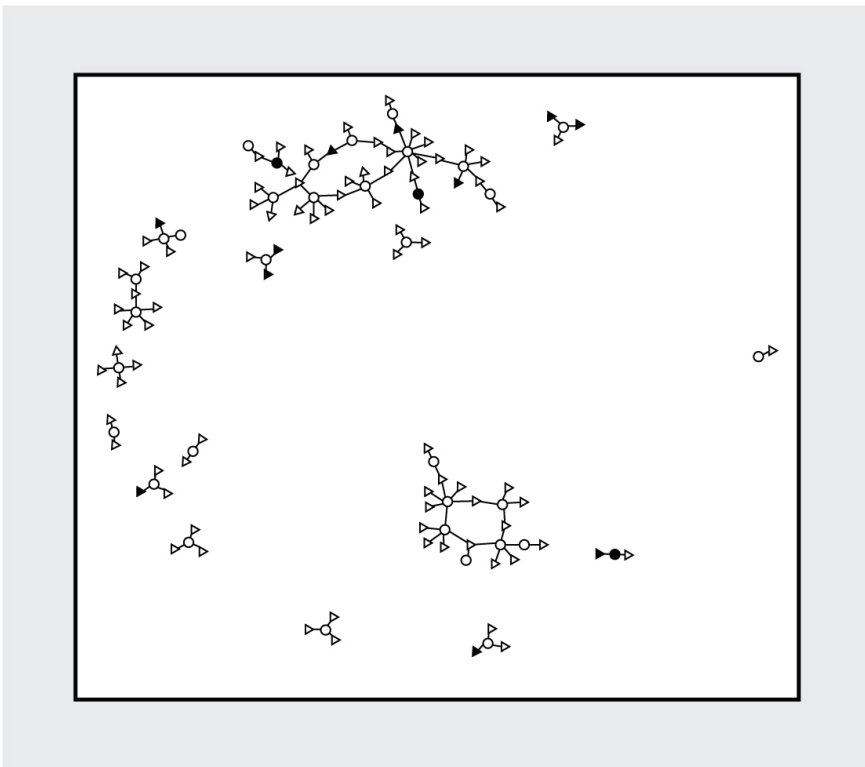
La proporción real de individuos en la población con la característica es 0.10. Los puntos ennegrecidos poseen la característica de interés y solamente hay 4, por lo que un estimado de la proporción es 0.08.

En el segundo ejemplo se presenta una muestra inicial S_0 de 50 individuos donde se añade una muestra S_1 de 83 referidos. En este caso

la proporción de individuos en la población que posee la característica de interés es 0.10. En la gráfica los nodos redondos forman la muestra inicial S_0 . Los ennegrecidos representan los individuos con la característica. Los triángulos forman la muestra de referidos S_1 . Los triángulos ennegrecidos corresponden a individuos con la característica de interés. En la **Tabla 2** se presenta la distribución de la muestra total.

Figura 2

Gráfica de la muestra inicial S_0 de 50 individuos y de la muestra S_1 de 83 referidos, con respecto a la tendencia o no de la característica de int.



Nota. W. Sarmientos, tesis no publicada, UPRM, 2008

Tabla 2*Distribución de la muestra total en muestreo por redes.*

poseen la característica	muestra inicial (S ₀)	muestra de referidos (S ₁)	totales
no	47	72	119
sí	3	11	15
TOTALES	50	83	133

*Nota. En la muestra inicial 3 de 50 tienen la característica (6%)**En la muestra de referidos 11 de 83 tienen la característica (13.3%)**En la muestra total 15 de 133 tienen la característica (11.3%)*

W. Sarmientos, tesis no publicada, UPRM, 2008

MUESTREO POR MULTIPLICIDAD (SIRKEN, 1970)

En este diseño la muestra se genera utilizando las redes de contactos de los participantes de una muestra aleatoria inicial. A cada participante se le pregunta si en su hogar ha ocurrido un evento cuya prevalencia en la población se desee estimar (por ejemplo, presencia de cáncer, emigración de un miembro de la familia, robo domiciliario, etc.). Luego se les pregunta a cuántos contactos de sus redes personales (vecinos, familia, amistades, compañeros de trabajo, miembros de su congregación religiosa, clubes, asociaciones profesionales, etc.) les ha ocurrido el evento de interés. La multiplicidad surge porque un evento puede ser reportado por múltiples fuentes. En este caso, hay que ponderar los estimados por el número de fuentes reportando el evento.

Este método es muy útil en el estudio de poblaciones raras pues aumenta la eficiencia con que las encuestas en hogares pueden estimar la prevalencia de eventos “raros”, pero no se puede implantar en investigaciones de poblaciones ocultas, dado que, por la naturaleza de los mismos, sus miembros no quieren que se conozca su pertenencia a tales grupos.

VENTAJAS DE LOS MÉTODOS DE MUESTREO POR REDES O ENLACES

- El proceso de referidos por enlaces permite al investigador acceder a poblaciones ocultas o escasas que, por otros métodos, no podrían ser estudiadas.
- El proceso es fácil de planificar y conducir.
- No requiere el uso de muchos recursos y es costo efectivo.

DESVENTAJAS DE LOS MÉTODOS DE MUESTREO POR REDES O ENLACES

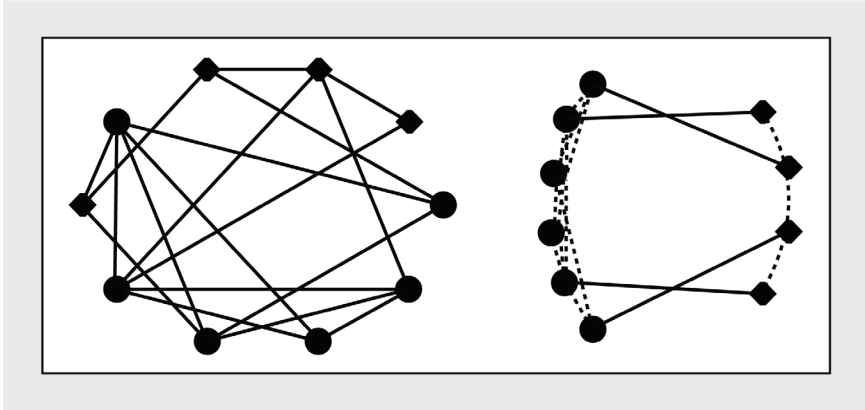
- El investigador tiene poco control del proceso de obtención de la muestra.
- No hay garantías de representatividad de la muestra pues se desconoce la distribución de la población y de la muestra. No pueden hacerse inferencias a la población.
- El riesgo de que la muestra sea sesgada es muy alto porque los participantes tienden a nominar a personas que conocen muy bien y que puedan compartir otras conductas, ideas y preferencias.

MUESTREO *RESPONDENT-DRIVEN SAMPLING* (HECKATHORN, 1997)

Este diseño combina el muestreo por multiplicidad con el de rastreo de enlaces. Tiene la ventaja de que compensa, usando procedimientos matemáticos, el sesgo que resulta de la selección de nuevos contactos por contactos previos (principal desventaja de los muestreos por redes). De este modo, la muestra se torna probabilística. Este método se ha tornado muy popular y se han generado varias variantes. Citando a Heckathorn (2017), en bases de datos científicos se han marcado 642 artículos con RDS como tema, que han sido citados 10,217 veces en 4,897 artículos. Además, los NIH han otorgado más de \$180 millones a 448 proyectos donde se aplica el RDS. Una representación gráfica que ilustra la implantación de este método se muestra en el siguiente ejemplo:

Figura 3

Dos representaciones de las conexiones de una misma población hipotética.



Nota. Tomado de Salganik, M.J. y Heckathorn, D.D, *Sampling and Estimation in Hidden Populations using Respondent Driven-Sampling*, Sociological Methodology, Vol. 34 (2004).

En la segunda figura puede notarse que el 60% de los miembros son círculos y el 40% son diamantes. Los trazos de puntos representan relaciones de amistad dentro de los dos grupos.

SUPUESTOS DEL MÉTODO RESPONDENT-DRIVEN SAMPLING (RDS)

Salganik y Heckathorn, 2004, han establecido los siguientes supuestos para implantar este método:

- Los participantes se conocen unos a otros como miembros de la población objetivo.
- La red de la población objetivo forma un solo componente (patrón para obtener contactos)
- El muestreo se hace con sustitución.
- Los participantes saben cuál es el tamaño de su red personal, es decir el número de los “otros” que cada uno conoce y que reúne los requisitos del estudio (por ejemplo, usuarios de drogas).

- Los participantes reclutan aleatoriamente de sus redes personales. Esto permite estimar el tamaño promedio de la red del grupo, ponderando los participantes por el recíproco del tamaño de su red.
- Cada participante recluta solamente a un contacto, así que el reclutamiento es uniforme a través de los grupos.

VENTAJAS DEL MÉTODO *RESPONDENT-DRIVEN SAMPLING* (RDS)

- Es costo efectivo.
- Su implantación permite un rápido acceso a poblaciones ocultas o escasas.
- Las muestras resultantes son probabilísticas con estimadores asintóticamente insesgados.

MÉTODOS DE MUESTREO Y ESTIMACIÓN INDIRECTA EN POBLACIONES OCULTAS

En la Introducción se expuso la existencia de unos métodos indirectos, basados en muestras aleatorias, de estimación de conductas o características sensitivas o embarazosas que ostentan algunos grupos en la sociedad y que son sancionadas social, moral o legalmente. A estos grupos se les denomina subpoblaciones ocultas. El problema que se plantea es que los estimados obtenidos por el método tradicional de cuestionarios tienden a subestimar la prevalencia de la característica o conducta sensitiva en la sociedad, dado que un individuo que la posea puede optar por dejar este tipo de preguntas sin respuesta o responderla falsamente, aunque se le provean todas las garantías de anonimato y confidencialidad. Con mayor razón, el método de entrevistas no se puede implantar en poblaciones de esta naturaleza.

El propósito de estas técnicas indirectas es introducir procedimientos matemáticos y estadísticos que les brinden a todos los participantes (estigmatizados o no) la confianza y la garantía de que las respuestas que cada uno dé a las preguntas sensitivas o embarazosas no serán conocidas por el investigador. Los métodos indirectos más

conocidos son: respuesta aleatorizada; el de listas no-pareadas y el diseño de respuesta total en bloques aleatorios incompletos. Particularmente, la técnica original de respuesta aleatorizada sugerida por Warner, S. (1965) ha experimentado cambios para hacer las estimaciones de prevalencia más precisas u obtener estimados de frecuencia de la conducta sensitiva o embarazosa.

LA TÉCNICA DE RESPUESTA ALEATORIZADA (RRT)

S. Warner (1965) propuso un procedimiento como el siguiente para estimar la prevalencia de una conducta sensitiva o embarazosa en una población:

- Seleccionar una muestra aleatoria de personas.
- Para efectos de la presentación, designar con la letra A al grupo de personas que tenga la característica sensitiva y con la letra B el grupo que no la tenga.
- Sea π la proporción de personas en el grupo A, los que tengan la característica, y $(1 - \pi)$ la proporción de personas en el grupo B, los que no la tengan.
- Se prepara un conjunto de tarjetas, iguales en apariencia. Una proporción θ de ellas se marca con la letra A y el resto con la letra B. Se mezclan las tarjetas y se apilan, colocando la cara marcada hacia abajo.
- En un cartel se indica que la letra A es para la persona que tenga la característica sensitiva y B para la que no la tenga.
- A una persona de la muestra se le solicita que escoja una tarjeta al azar, vea la letra marcada y que la devuelva al mazo.
- El investigador se coloca de tal manera que no pueda saber qué tipo de tarjeta fue extraída por el participante y le pide que conteste “SÍ” si la tarjeta extraída coincide con su estatus sobre la pregunta sensitiva, por ejemplo, si sacó una tarjeta A y tiene la característica contestará “SÍ”, y si no la tiene, contestará “No”. Si escogió una tarjeta B y no tiene la característica, contestará

“Sí” y si la tiene contestará “NO”. - El investigador llevará un registro de cada “SÍ” informado.

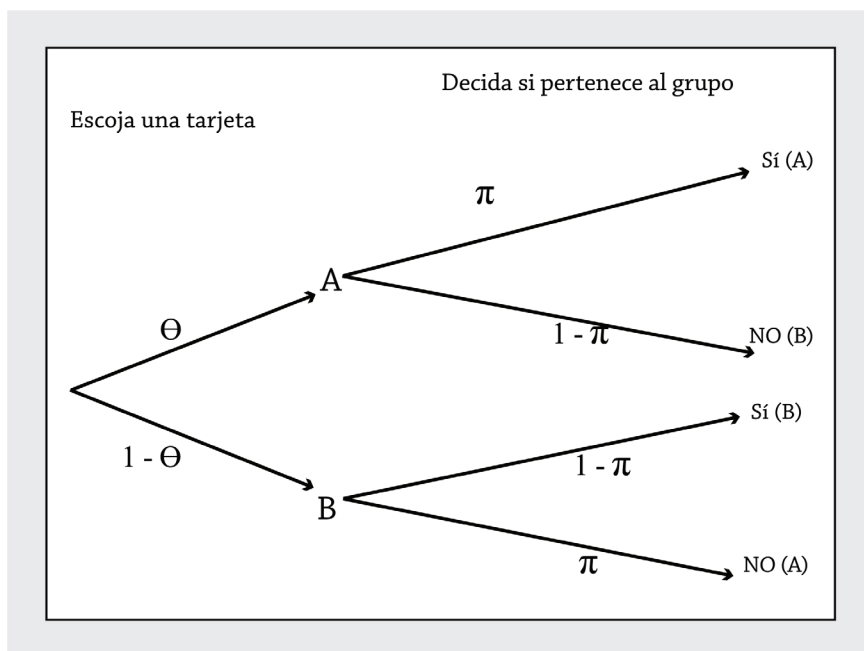
Note que el investigador no tiene conocimiento de qué tipo de tarjeta fue extraído por la persona.

Este procedimiento se realiza con cada una de las personas en la muestra.

En el siguiente diagrama se muestra el procedimiento.

Figura 4

Procedimiento propuesto por S. Warner.



Nota: θ = proporción de tarjetas marcadas A
 $1 - \theta$ = proporción de tarjetas marcadas B
 π = proporción de la población con la característica sensitiva
 $1 - \pi$ = proporción de la población que no tiene la característica sensitiva.

La probabilidad de responder “Sí” está dada por:

$$P(\text{Sí}) = \theta\pi + (1 - \theta)(1 - \pi)$$

por lo que, resolviendo para π se tiene:

$$\pi = \frac{P(Si) + \theta}{2\theta - 1}; \theta \neq \frac{1}{2}, 1$$

Si los n individuos que participan en el proceso pertenecen a una muestra aleatoria, entonces un estimador de $P(Si)$ está dado por:

$$\hat{P}(Si) = \frac{n_{Si}}{n} = \theta \hat{p} + (1 - \theta)(1 - \hat{p})$$

donde \hat{p} es el estimador de la proporción π de individuos que posean la característica sensitiva en la población. De esta ecuación se obtiene el estimador dado por:

$$\hat{p} = \frac{1}{(2\theta - 1)} \left(\frac{n_{Si}}{n} \right) - \left(\frac{1 - \theta}{2\theta - 1} \right), \theta \neq \frac{1}{2}, 1$$

y su varianza estimada,

$$\hat{V}(\hat{p}) = \frac{1}{(2\theta - 1)^2} \left(\frac{1}{n} \right) \left(\frac{n_{Si}}{n} \right) \left(1 - \frac{n_{Si}}{n} \right)$$

a partir de la cual se pueden obtener intervalos de confianza para la proporción π de individuos con la característica de interés en la población usando el estimador \hat{p} .

Un ejemplo (tomado de Blair et al, 2015) usando otro dispositivo aleatorio:

Estos autores utilizaron la técnica de Respuesta Aleatorizada (RR) para estimar la proporción de la población civil que tuvo contacto social con los grupos guerrilleros durante el conflicto armado de Nigeria entre los años 2007 y 2008, asunto de alto riesgo personal para los investigadores. Usaron como dispositivo aleatorio un dado. Las instrucciones a los participantes eran: Si sale el “1” responda Sí; si sale el 6 responda “No” y si sale alguno de los números 2,3,4 o 5, responda con un “Sí” o un “No” a la siguiente pregunta:

“Durante el conflicto armado entre el 2007 y el 2008 hubo algún familiar, un amigo o alguien que fuera militante (guerrillero) con quien Ud. acostumbraba a hablar”

El investigador no tuvo conocimiento, para cada individuo participante, de cuál era el resultado cuando este lanzaba el dado. *A priori* se había estimado que la proporción sería 10%, pero el resultado con la RR fue 26%, con un intervalo de 95% de confianza de (23%, 29%). Participaron 2,448 personas de 204 comunidades afectadas.

VENTAJAS DE LA TÉCNICA DE RESPUESTA ALEATORIZADA (RR):

- Cuando la proporción de personas que tienen la característica y que mienten en un cuestionamiento directo es alta, el método de Respuesta Aleatorizada es mucho más eficiente que el directo.
- El participante tiende a tener más confianza si percibe que el procedimiento garantiza que su respuesta no será conocida por el investigador.

DESVENTAJAS DE LA TÉCNICA DE RESPUESTA ALEATORIZADA (RR)

- Se parte del supuesto de que el participante está respondiendo con la verdad, pero no hay garantía de que lo haga.
- La varianza de la proporción estimada depende de dos elementos importantes: la proporción θ y el tamaño de la muestra. Greenberg, B. et al. (1969) analizan los efectos que tienen estos en la varianza y la eficiencia de los estimadores, por ejemplo: Los valores de θ muy cercanos a 0.5 incrementan la varianza.
- En el caso de la técnica sugerida por Warner, como la tarjeta seleccionada proviene del mazo de tarjetas, si el participante observa que casi todas las tarjetas están marcadas con A (valores de θ muy cercanos a 1) desconfiará del procedimiento.

- En general, los estimados de π tienden a tener mucha variabilidad y a ser ineficientes, aunque en el ejemplo de Blair et al, ese no fue el caso.
- Si las instrucciones no son claras, puede ocurrir que haya participantes que no entiendan cómo se usa el dispositivo aleatorio, introduciendo un elemento adicional en el error de estimación.

REFERENCIAS

- Blair, G., Imai, K. y Yang-Yang, Z., (2015), *Design and Analysis of the Randomized Response Technique*, JASA, Vol. 110, N 511
- Goodman, L., (1961), *Snowball Sampling*, *The Annals of Mathematical Statistics*, 32(1)
- Greenberg, B., Abdul-Ela, A., Simmons, W., and Horvitz, D., *The unrelated question randomized response model: theoretical framework*. JASA, June 1969, 520-539.
- Heckathorn, D. D (1997), *Respondent Driven Sampling: a new approach to the study of hidden samples*, Soc. Probl. 44(2)
- Heckathorn, D. D (2002), *Respondent Driven Sampling II: deriving valid population estimates from chain-referral samples of hidden populations*, Soc. Probl. 44(2)
- Heckathorn, D. D (2017) y Cameron C.J., *Network Sampling: From Snowball and Multiplicity to Respondent-Driven Sampling*, *Annual Rev. of Sociology*, 43
- Salganik, M.J. y Heckathorn, D.D, (2004). *Sampling and Estimation in Hidden Populations using Respondent Driven-Sampling*, *Sociological Methodology*, Vol. 34
- Sarmiento, W. (2008) *Estimación de la proporción de una población oculta a través de muestreo por "Bola de Nieve estratificado"*, Tesis no publicada, UPRM

Sirken, M. G. (1970) *Household Surveys with Multiplicity*, JASA, Vol. 65 (329)

Warner, S., (1965), *Randomized Response: A survey technique for eliminating evasive answer bias*, JASA, Vol. 60 (309)

Estimación de parámetros en poblaciones ocultas (ensayo II)

*Julio César Quintana Díaz, Ph.D.
asesor del Centro de Apoyo Estadístico
Pontificia Universidad Católica de Puerto Rico*

INTRODUCCIÓN

Sobre el tema de las poblaciones escasas, en el ensayo I se planteó la necesidad de aplicar métodos alternos de muestreo para estudiar este tipo de poblaciones. Las razones principales para ello son que no tienen un marco muestral definido, son de difícil acceso, constituyen menos del 10% de la población que las contiene y, en el caso de las ocultas, sus miembros tienen una característica embarazosa que los estigmatiza o se involucran en actividades sancionadas por la sociedad. En lo que concierne a las poblaciones escasas se discutieron con detalle métodos no probabilísticos basados en una muestra inicial por conveniencia que luego se va incrementando al usar las técnicas de muestreo por redes de contactos o por referencia.

En general, las poblaciones ocultas comparten varias características con las poblaciones escasas. Sin embargo, el énfasis al estudiarlas se hace sobre el uso de métodos indirectos para la estimación de sus parámetros. Estas técnicas están orientadas a proveerles confianza y garantías a sus miembros de que los investigadores no podrán saber de manera alguna cuál o cuáles han sido sus respuestas a las preguntas embarazosas. Cuando se usan métodos directos para la estimación de parámetros en preguntas sensitivas se proveen garantías de anonimato y confidencialidad, pero pueden ser insuficientes y el participante podría optar por no contestar o por contestarlas falsamente, haciendo que los

estimados sean espurios. Ya al final del ensayo I se presentó el primer método indirecto de estimación de parámetros en poblaciones ocultas, el de “respuesta aleatorizada”, desarrollado por S. Warner (1965) pionero en este enfoque. En este segundo ensayo se presentan métodos mejorados del modelo de “respuesta aleatorizada”; una lista de estudios donde se aplicó esta técnica o una de sus modalidades y el método de “listas no pareadas” de J. D. Miller (1984) como otra alternativa de estimación indirecta de respuestas a preguntas sensitivas.

OTROS MÉTODOS DE ESTIMACIÓN INDIRECTA DE PARÁMETROS EN POBLACIONES OCULTAS

LA TÉCNICA DE “PREGUNTA NO RELACIONADA EN EL DISEÑO DE RESPUESTA ALEATORIZADA”

- Todas las investigaciones posteriores al modelo propuesto por Warner, S. (1965): Horvitz, et al; Moors, R.; Greenberg, B. et al; Edgell, S. et al., y muchos otros, se dedicaron a proveer marcos teóricos para estas técnicas o introducir nuevas versiones del modelo original con el propósito de mejorar la eficiencia del método. Fue de gran importancia la contribución de Horvitz, et al. (1967;); H. L Simmons et al. (1967) y de Greenberg, et al. (1969) al incorporar al modelo de Warner una pregunta no relacionada con la característica sensitiva y así mejorar la eficiencia del modelo, porque en el diseño original de “respuesta aleatorizada” las dos opciones de preguntas que debía responder un participante estaban relacionadas con la característica sensitiva, una era la opuesta de la otra, lo que tenía el efecto de que de una u otra manera siempre estuviera ante el participante la pregunta sensitiva. La sustitución de la afirmación “No tengo la característica sensitiva” por otra irrelevante al tema tiene dos ventajas: le brinda más confianza al participante y la estimación del parámetro es más eficiente.

El procedimiento para implantar esta técnica es el siguiente:

- Se selecciona una muestra simple al azar de n personas.

- El instrumento que asigna aleatoriamente la respuesta a la pregunta sensitiva puede ser una moneda, un dado, una ruleta, cuya distribución de probabilidad sea conocida.
- Debe establecerse de antemano la regla de asignación del participante a la pregunta que debe responder, dependiendo del resultado al usar el instrumento.
- La respuesta a la pregunta no relacionada debe ser dicotómica (sí o no) y con distribución de probabilidad conocida.

Un ejemplo hipotético: se desea estimar la proporción de personas en Puerto Rico que subestiman sus ingresos al radicar sus planillas de contribuciones.

A tales efectos, a una muestra aleatoria de 300 personas se les pide que lancen una moneda. ***El resultado es desconocido para el investigador.*** Si cae “cara” el participante contesta la pregunta sensitiva A, solamente con un “sí” o un “no”. Si cae “cruz” contesta la pregunta no sensitiva B, solo con un “sí” o un “no”.

Las preguntas sensitivas e irrelevantes al tema sensitivo son, respectivamente:

- a. Subestimo mis ingresos en las planillas de contribuciones.
- b. Soy mayor de 34 años.

Además, es importante saber la proporción de personas mayores de 34 años en Puerto Rico para así cumplir con el cuarto requisito para aplicar esta técnica. El último estimado registrado en la *Encuesta de la Comunidad* del Censo Federal del por ciento de personas mayores de 34 años en Puerto Rico fue .556 en 2016. El siguiente es el diagrama ramificado del proceso. Siguiendo este diagrama, la probabilidad de responder “sí” está dada por:

$$P(\text{SÍ}) = \frac{1}{2}\pi + \frac{1}{2}(0.556)$$

donde π es la proporción en la población que subestima sus ingresos en las planillas.

Un estimador de π es

$$\hat{P} = 2P(Sí) - 0.556$$

v un estimador de $P(Sí)$ es

$$\frac{n_{Sí}}{n}$$

Si de la muestra aleatoria de 300 personas, 120 respondieron “sí”, entonces un estimado de $P(SÍ)$ es $120/300 = 0.40$, por lo que un estimado de π es $\hat{P} = 0.244$, es decir que en este ejemplo hipotético se estima que el 24.4% de los contribuyentes subestiman sus ingresos en las planillas. La varianza estimada de \hat{P} está dada por

$$\hat{V}(\hat{p}) = 4 \left(\frac{1}{n} \right) \left(\frac{n_{Sí}}{n} \right) \left(1 - \frac{n_{Sí}}{n} \right)$$

Sustituyendo los valores correspondientes en la fórmula se tiene que $\hat{V}(\hat{P}) = 0.0032$ y el error en estimación es $\varepsilon = 0.111$ por lo que un intervalo de un 95% de confianza para la proporción de personas que subestiman sus ingresos en las planillas sería (0.133 , 0.355).

ESTUDIOS A GRAN ESCALA DONDE SE APLICÓ EL MÉTODO DE “RESPUESTA ALEATORIZADA”

La siguiente es una lista no exhaustiva de estudios a gran escala que se han realizado en distintos países donde se utilizó este método para la estimación de parámetros en temas sensitivos:

- Encuesta Nacional para estimar el número de mujeres teniendo abortos durante un período de 12 meses (1978); IM Shimizu.
- Aplicación a la administración de parques marinos con el propósito de evaluar el cumplimiento con los permisos. *Environmental Management* (1985); M. Chaloupka.
- Estudio de la reducción de la ansiedad del entrevistado al usar la RRT *Advances in Consumer Research* (1988)

- Encuesta sobre evasión de impuestos en Australia usando RRT, *The Australian National University* (2001)
- Estudio usando RRT sobre las conexiones sociales con los grupos armados en Nigeria. Blair, G., Imai, K et al. *JASA* (2015)
- Aplicación de la RRT en una encuesta nacional en Georgia (exrepública soviética) sobre el uso de sustancias en la población general, *Drug Alcohol Depend*; Kirtadze, E. (2015, 2018)
- El uso de la RRT y el método *network scale-up* para estimar el tamaño de la población de trabajadoras del sexo. Jing, L., Lu Q., Cui, Y., Yu, H. y Wang, T. (2018)

LA TÉCNICA DE LISTAS NO PAREADAS

En 1984 J.D. Miller, en su disertación doctoral no publicada, *A New Survey Technique for Studying Deviant Behavior*, introdujo la “técnica de conteo no pareado” como método alternativo a la de “respuesta aleatorizada”.

Consiste en seleccionar dos muestras aleatorias (preferiblemente grandes) de la misma población, una servirá como grupo control y la otra como grupo experimental. Al grupo control se le dará una lista con cierto número de enunciados que no son sensitivos, son infrecuentes, pero no imposibles. A cada participante de este grupo se le pide que reporte en la hoja de la lista **cuántos (no cuáles)** de esos enunciados le aplican al participante.

Al grupo experimental se le dará la misma lista de enunciados no sensitivos del grupo control, pero se le añadirá el enunciado sensitivo y se le pedirá que haga el mismo proceso que hacen los del grupo control. En ninguno de los dos grupos los participantes tienen conocimiento de cuál cuestionario están contestando.

Se calculan los promedios respectivos de los totales reportados por el grupo control y el experimental y la diferencia entre ellos, la varianza estimada de esta diferencia e intervalos de confianza para obtener la verdadera diferencia. Si las muestras fueron seleccionadas estrictamente al azar y se presume que los participantes del grupo

experimental contestaron con la verdad, la existencia de una diferencia estadísticamente significativa entre los promedios de los dos grupos se podría atribuir a la presencia del enunciado sensitivo en la lista del grupo experimental.

EJEMPLOS PARA ILUSTRAR EL USO DE LA TÉCNICA DE “LISTAS NO PAREADAS”

En 1991 Sniderman, P., Tetlock, P. y Piazza, T. publicaron un estudio titulado *National Race and Politics Survey* cuya cobertura fue en los 48 estados (hubo una réplica en 2014). Consistió de una encuesta telefónica (aunque luego una proporción sustantiva de participantes accedió a contestar además un cuestionario escrito). Se realizaron 2,223 entrevistas completadas (tasa de respuesta de un 65.3%). La encuesta constaba de múltiples preguntas sobre raza y política en Estados Unidos y se utilizó la técnica de “listas no pareadas”.

Tabla 1

Ejemplo de pregunta y sus respectivas listas no pareadas (traducción libre).

Pregunta al grupo control: “Ahora le voy a leer tres asuntos que a veces hacen que las personas se sientan enojadas o incómodas. Luego de que le lea los tres, dígame SOLAMENTE CUÁNTOS LE MOLESTAN, NO CUÁLES”	Pregunta al grupo experimental: “Ahora le voy a leer cuatro asuntos que a veces hacen que las personas se sientan enojadas o incómodas. Luego de que le lea los cuatro, dígame SOLAMENTE CUÁNTOS LE MOLESTAN, NO CUÁLES”
<ul style="list-style-type: none"> • El aumento del impuesto a la gasolina que aprobó el gobierno federal • Los atletas profesionales que obtienen salarios millonarios • La contaminación ambiental que causan las grandes corporaciones 	<ul style="list-style-type: none"> • El aumento del impuesto a la gasolina que aprobó el gobierno federal • Los atletas profesionales que obtienen salarios millonarios • La contaminación ambiental que causan las grandes corporaciones • Una familia negra que se muda al lado de su residencia

Sniderman, P. et al. en la misma encuesta, pero en 1994, compararon usando las técnicas de “listas no pareadas” y del “cuestionamiento directo” la proporción de personas que se sentían enojadas o molestas cuando los líderes negros le solicitaban al gobierno acción afirmativa.

Uno de los hallazgos fue que esta proporción entre blancos demócratas usando el método de listas se estimó en un poco más del 60% y en el cuestionamiento directo fue 20%.

Los procedimientos estadísticos para analizar los datos y hacer inferencias son los usuales para comparación de promedios en dos muestras independientes.

VENTAJAS DE LA TÉCNICA DE “LISTAS NO PAREADAS”

Es más fácil de aplicar a las muestras de participantes.

En el grupo experimental el participante confía más en que su respuesta no será conocida por el “entrevistador”.

Los procedimientos estadísticos que se aplican (por lo menos en la forma original de la técnica) no son sofisticados. No es un requisito la presencia física de los participantes, dado que las muestras aleatorias pueden obtenerse por teléfono usando rutinas ya establecidas para ello, por ejemplo, el programa CATI en Berkeley (aunque en la actualidad el uso de celulares ha tenido efecto sobre el uso de teléfonos regulares).

DESVENTAJA DE LA TÉCNICA DE “LISTAS NO PAREADAS”

Los auspiciadores de proyectos a gran escala tienen recelo de la implantación de esta técnica porque, en el caso de instrumentos impresos, consideran un gasto innecesario reproducir cuestionarios con preguntas no sensitivas que no guarden relación con el tema sensitivo que se quiere estudiar.

Si se van a utilizar cuestionarios impresos esta técnica es muy costosa, pues tienen que reproducirse dos tipos de cuestionarios.

Si hay varias preguntas sensitivas cuyas respuestas se desean analizar, solamente se podrá hacer una a la vez, lo que redundará en el aumento de alternativas no sensitivas, tanto en el grupo control como en el experimental.

REFERENCIAS

- Blair, G., Imai, K. y Yang-Yang, Z., (2015), *Design and Analysis of the Randomized Response Technique*, JASA, Vol. 110, N 511
- Dalton, D. R., Daily, C. M., Wimbush, J.C., (1997) *Sensitive Data in Business Ethics Research: A case for the Unmatched Count Technique (UCT)*, Journal of Business Ethics, Vol. 16.
- Imai, K., (2010) *Multivariate Regression Analysis for the Item Count Technique*, JASA, Vol. 106, No. 494(June)
- Imai, K. (2011) *Statistical Analysis for the Item Count Technique*, presentación en PowerPoint, Departamento de Política, Princeton University
- Kuha, J. y Jackson, J. (2014) *The Item Count Technique for Sensitive Questions: Modelling Criminal Behaviour*, Journal of the Royal Statistical Society, Vol. 63 Issue 2.
- Kowalczyk, B. y Wieczorkowski, R. (2017) *Comparing proportion of sensitive items in two populations when using Poisson and Negative Binomial Item Count Technique*, Qualitative Methods in Economics, Vol. XVIII, No. 1
- Miller, J.D. (1984), *A New Survey Technique for Studying Deviant Behavior*", Unpublished Ph.D. dissertation, George Washington University
- Simmons, W. R. (1969) en Greenberg, B.G., Abul-Ela, A. L., Simmons, W.R., Horvitz, D. *The unrelated question Randomised Response Model: Theoretical Framework*; JASA. Taylor & Francis 64(326).
- Sniderman, P., Phillip, E., and Piazza, T. (2014) *National Race and Politics Survey 1991*, Inter universities Consortium for Political and Social Research (Version 4)
- Stephenson, M. (2017) *Using the Unmatched Count Technique to Elicit More Accurate Answers on Corruption Experience Surveys*, GAB | The Global Anticorruption Blog

Introducción al Análisis mediante Ecuaciones Estructurales

*José Noel Caraballo, Ph.D.
asesor del Centro de Apoyo Estadístico
Pontificia Universidad Católica de Puerto Rico*

INTRODUCCIÓN

El análisis mediante ecuaciones estructurales es una metodología estadística general que integra áreas que tradicionalmente se consideraban separadas: análisis de regresión, análisis de trayectorias (*path análisis*) y análisis factorial. Así pues, mediante el análisis de ecuaciones estructurales se pueden llevar a cabo estos tres tipos de análisis. El análisis mediante ecuaciones estructurales, sin embargo, permite llevar a cabo otros tipos de análisis tales como: análisis de modelos causales, análisis factorial confirmatorio, análisis de curvas de crecimiento, análisis de clases latentes, análisis de modelos mixtos, entre otros.

Podría pensarse en un modelo estructural como un conjunto de ecuaciones de regresión que haya que resolver de forma simultánea. Entre las ventajas del análisis estructural se encuentra el que es posible modelar el error en la medición cuando las variables bajo análisis son variables latentes (no observadas). Esto permite obtener mejores estimados de los parámetros del modelo, eliminando sesgos por errores de medición y disminuyendo los estimados de las varianzas de error.

En los ejemplos que se dan a continuación se presenta la representación gráfica, así como la matemática de los modelos estructurales.

REPRESENTACIÓN GRÁFICA

Los análisis mediante ecuaciones estructurales se pueden representar gráficamente, lo que facilita su comprensión. En este escrito se considerarán cinco modelos básicos: regresión, análisis de trayectorias, análisis factorial, análisis de regresión con variables latentes y modelos estructurales generales.

En la representación gráfica, una flecha indica la dirección de una relación causal, una flecha doble representa una correlación o covarianza, un óvalo o un círculo representan una variable latente (variable no observada o constructo) y un rectángulo o un cuadrado representan una variable observada o manifiesta. La terminología de ecuaciones estructurales incluye además los términos de variables endógena y exógena, tomados del análisis de trayectorias. En un diagrama, una variable exógena se distingue porque no hay flechas sencillas apuntando hacia ella. De una variable exógena pueden salir una o más flechas. Una variable endógena se distingue porque solo tiene flechas que apuntan hacia ella.

ANÁLISIS DE REGRESIÓN

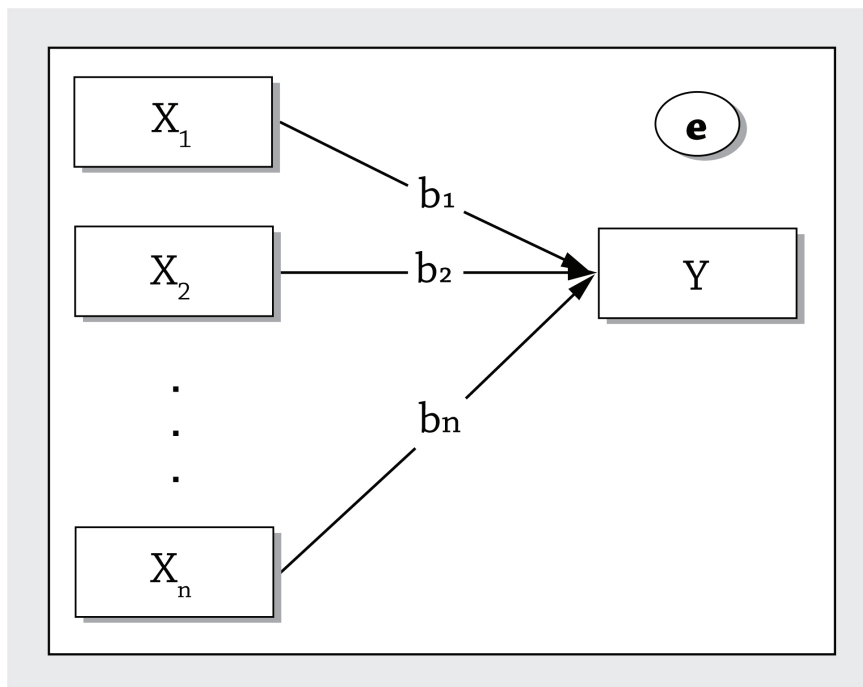
En el análisis de regresión clásico, debe haber por lo menos una variable independiente (exógena) y una sola variable dependiente (endógena). El modelo de regresión se representa matemáticamente mediante la ecuación:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon$$

En el modelo de regresión, tanto las variables independientes como la dependiente son variables observadas o manifiestas, por lo que se representarán gráficamente mediante rectángulos. El término de error, ϵ , sin embargo, es una variable no observada o latente, por lo que se representará mediante un círculo. Tanto $\beta_0, \beta_1, \dots, \beta_n$ como ϵ representan los parámetros a ser estimados usando un conjunto de datos. La **Figura 1** ilustra la representación general de un modelo de regresión.

Figura 1

Representación gráfica del modelo de regresión clásico.



En esta gráfica la dirección de las flechas va de las variables independientes (X_1, \dots, X_n) hacia la variable dependiente (Y). Note también que hay una flecha que va desde e hasta Y , lo que implica que hay una (o más) variables no observadas que afectan a la variable dependiente. Note, además, que en este diagrama se han utilizado las letras latinas b_1, \dots, b_n y e , las cuales representan los estimados obtenidos de los parámetros en la ecuación de regresión. Observe que para las variables independientes no hay términos de error, esto debido a que en el análisis de regresión clásico se presume que las variables independientes son medidas sin error. Como se verá más adelante, en el análisis de regresión latente, un caso especial de modelo estructural, es posible estimar el error en las variables independientes.

ANÁLISIS DE TRAYECTORIAS

La representación de un análisis de trayectorias es similar al del análisis de regresión. La diferencia básica es que en el análisis de trayectorias puede haber variables mediadoras. Además, es posible que haya más de una variable dependiente. La **Figura 2** muestra un análisis de trayectorias que incluye una variable moderadora y una variable dependiente.

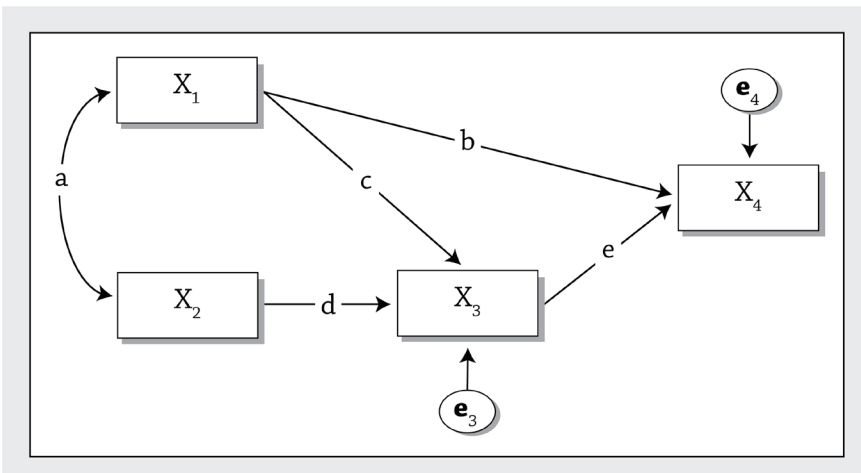
La variable moderadora, X_3 , se reconoce porque hay flechas sencillas que apuntan hacia ella y flechas sencillas que salen de ella. Las variables moderadoras se clasifican como variables exógenas también. Note que, como X_3 depende de X_1 y X_2 , también hay que incluir un término de error para ella (e_3). La razón para esto es evidente si se representa este modelo mediante un conjunto de ecuaciones de regresión:

$$X_3 = \beta_{00} + \gamma X_1 + \delta X_2 + \epsilon_3$$

$$X_4 = \beta_{01} + \beta X_1 + \varepsilon X_3 + \epsilon_4$$

Figura 2

Ejemplo de un modelo de trayectorias.



La relación entre los coeficientes en estas ecuaciones y la gráfica son los siguientes: **b** es el estimado de β , **c** es el estimado de γ , **d** es el estimado de δ y **e** es el estimado de ε . Note que el modelo incluye además una correlación (o covarianza) entre X_1 y X_2 , cuyo estimado está representado por la letra **a**.

En la **Figura 1**, las variables X_1 y X_3 tienen un efecto directo sobre la variable X_4 . La variable X_2 también tiene un efecto sobre X_4 , pero es indirecto. Es decir, X_2 influye sobre X_3 y esta última influye sobre X_4 . X_2 se correlaciona con X_1 (flecha doble), así que X_2 afecta a X_4 por otras rutas: $X_2 \rightarrow X_1 \rightarrow X_4$ y también $X_2 \rightarrow X_1 \rightarrow X_3 \rightarrow X_4$. El efecto total de X_2 sobre X_4 está dado por $d \cdot e + a \cdot b + a \cdot c \cdot e$. De igual forma, el efecto total de X_1 sobre X_4 es $b + c \cdot e + a \cdot d \cdot e$.

ANÁLISIS FACTORIAL

Mediante el análisis de ecuaciones estructurales es factible modelar un gran número de tipos de análisis factoriales. La **Figura 3** muestra la representación gráfica de varios tipos de análisis factoriales: de un factor, dos factores, dos factores correlacionados y un factor con covarianzas entre algunos de los errores de las variables observadas. Las variables X_1 a X_6 representan indicadores o ítems de una escala y son variables observadas, por lo cual se dibujan como rectángulos. El factor F es una variable latente cuyo valor se infiere por medio de los indicadores. Por esta razón se dibuja en un óvalo.

El análisis factorial también se puede representar por medio de un conjunto de ecuaciones estructurales. Por ejemplo, para el caso de un factor ilustrado en la **Figura 3(b)** está:

$$X_1 = \beta_{01} + \beta_{11}F + \epsilon_1$$

$$X_2 = \beta_{02} + \beta_{12}F + \epsilon_2$$

$$X_3 = \beta_{03} + \beta_{13}F + \epsilon_3$$

$$X_4 = \beta_{04} + \beta_{14}F + \epsilon_4$$

$$X_5 = \beta_{05} + \beta_{15}F + \epsilon_5$$

$$X_6 = \beta_{06} + \beta_{16}F + \epsilon_6$$

Figura 3

Representaciones gráficas de varios tipos de análisis factoriales.

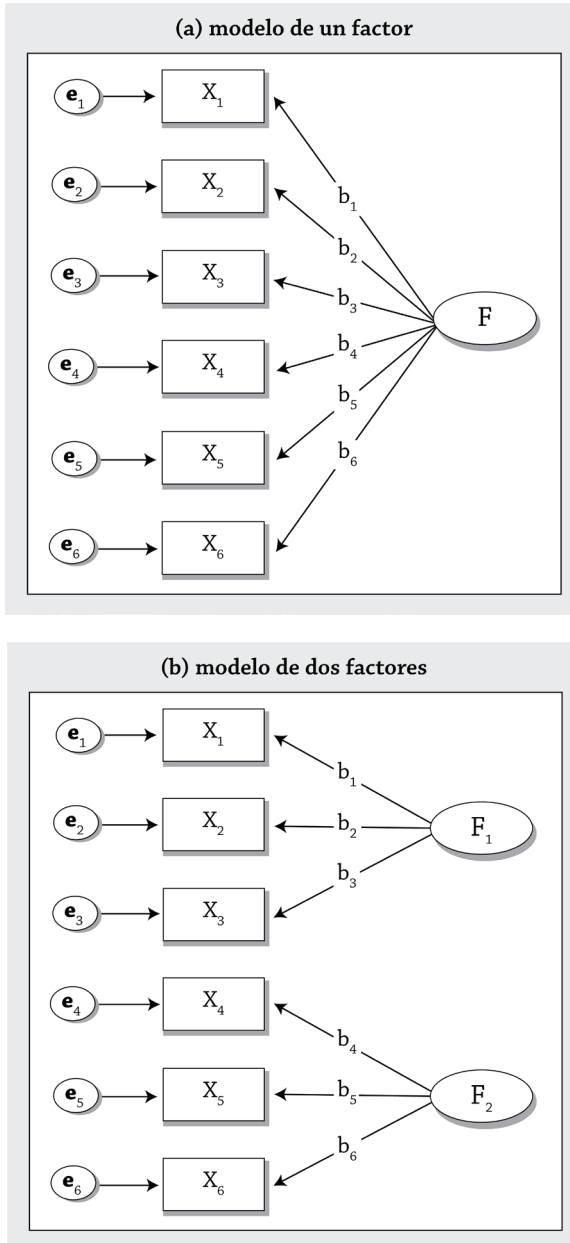
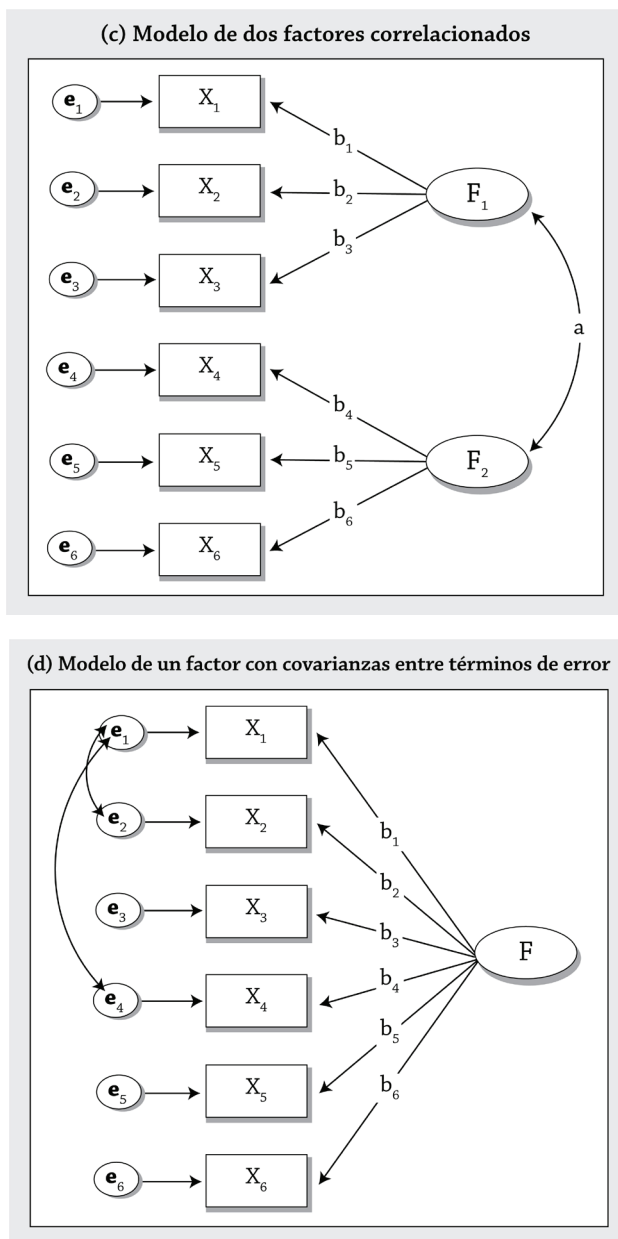


Figura 3 (continuación)



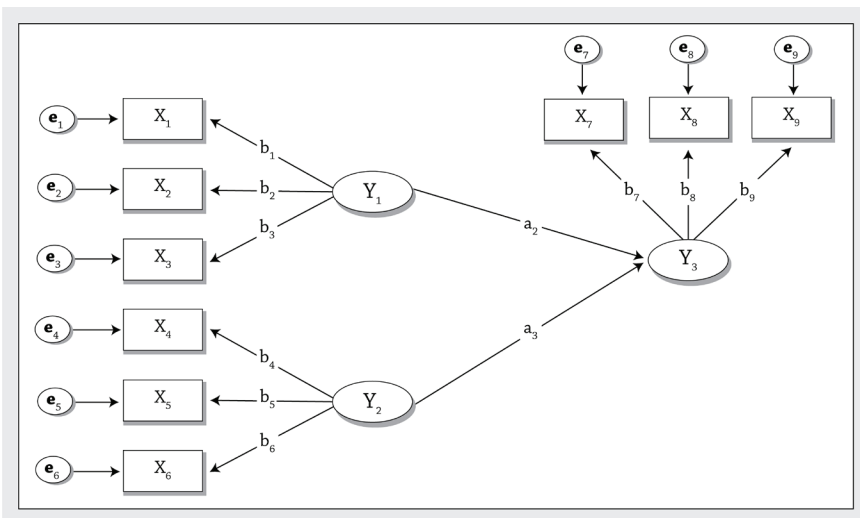
Como se puede apreciar en la **Figura 4(d)**, el análisis de ecuaciones estructurales permite modelar la covarianza entre los errores de los indicadores. Esta es una ventaja sobre el análisis factorial tradicional en el que se asume que los errores de medición de los indicadores son independientes unos de los otros. De forma análoga se pueden construir modelos para analizar la invarianza métrica de una escala.

ANÁLISIS DE REGRESIÓN CON VARIABLES LATENTES

Si una o más variables independientes son variables latentes, por ejemplo, constructos medidos a través de un conjunto de ítems, entonces es posible llevar a cabo un análisis de regresión con variables latentes. La representación gráfica de este modelo estructural es una combinación del modelo de regresión clásico y el análisis factorial, como se ilustra en la **Figura 4**.

Figura 4

Representación gráfica de un modelo de regresión con variables latentes.

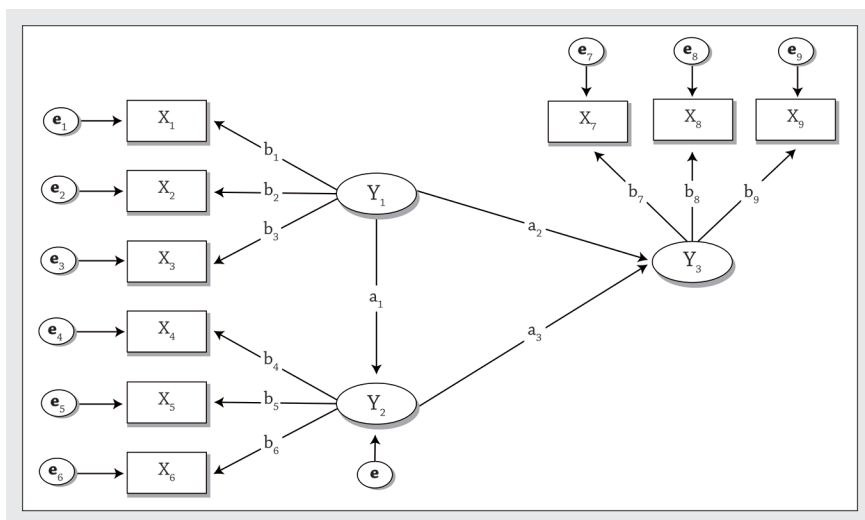


En esta figura las variables independientes Y_1 y Y_2 son variables latentes, y las variables X_1 a X_3 y X_4 a X_6 son sus indicadores, respectivamente. La variable dependiente Y_3 es también una variable latente con indicadores X_7 a X_9 .

MODELO ESTRUCTURAL GENERAL

Un modelo estructural general es similar a un modelo de trayectorias en las cuales una o más variables exógenas son variables latentes. El modelo será similar al de una regresión con variables latentes con la adición de una o más variables mediadoras (véase la **Figura 5**). Recuerde que el modelo estructural general puede incluir más de una variable endógena.

Figura 5
Modelo estructural general.



ETAPAS EN EL ANÁLISIS MEDIANTE ECUACIONES ESTRUCTURALES

El análisis estructural se lleva a cabo, generalmente, en unas cinco etapas:

1. Especificación del modelo. Es decir, establecer concretamente las relaciones entre las variables bajo estudio, incluyendo los coeficientes de regresión, correlaciones y covarianzas. Típicamente el modelo especificado se representa gráficamente como se ilustró en la sección anterior. Identificación del modelo.

2. Identificación del modelo. Esto se refiere a determinar si hay suficiente información como para poder obtener estimados de todos los parámetros (medias, varianzas, coeficientes de regresión, correlaciones y covarianzas) en el modelo. Existen varias reglas generales para saber si un modelo está especificado, pero los programas de computadora especializados hacen este análisis de forma automática, por lo cual no es necesario abundar sobre ellas.
3. Estimación de parámetros. La estimación de parámetros es la médula del análisis estructural. En esta etapa es que se obtienen los estimados de las medias, varianzas, correlaciones o covarianzas y los coeficientes estructurales o coeficientes de regresión. Existen varios métodos para obtener estos estimados, incluyendo la minimización de la suma de cuadrados, minimización general de la suma de cuadrados, estimación no distribucional y el método de verosimilitud máxima (*maximum likelihood estimation* -MLE). Este último es el método más utilizado porque produce estimados no sesgados y con mínima varianza, sin embargo, requiere de muestras relativamente grandes.
4. Evaluación del ajuste del modelo. Existen varios métodos para evaluar la adecuación del modelo estructural. Adecuación se refiere a cuánto se ajustan los datos al modelo propuesto para ellos. Se puede llevar a cabo una prueba de significación mediante la estadística Ji-cuadrada, sin embargo, esta está influenciada por el tamaño de la muestra y tiende a resultar en diferencias estadísticamente significativas entre el modelo propuesto y el que se infiere de los datos. Por este motivo, comúnmente se hace uso de índices de ajuste, los cuales típicamente tienen un rango de valores entre cero (cero ajuste) a 1.0 (ajuste completo).
5. Reespecificación del modelo. Cuando la evaluación de los resultados indica que no hay un buen ajuste de los datos al modelo, típicamente se realizan modificaciones al mismo hasta obtener un buen ajuste. Las modificaciones incluyen la eliminación de relaciones causales entre variables en el modelo

(flechas de una dirección), o la inclusión de nuevas relaciones causales, la inclusión de covarianzas entre las variables, a menudo en los términos de error de los indicadores, entre otras. Una vez que se reespecifica un modelo, se procede a repetir los últimos cuatro pasos de este procedimiento.

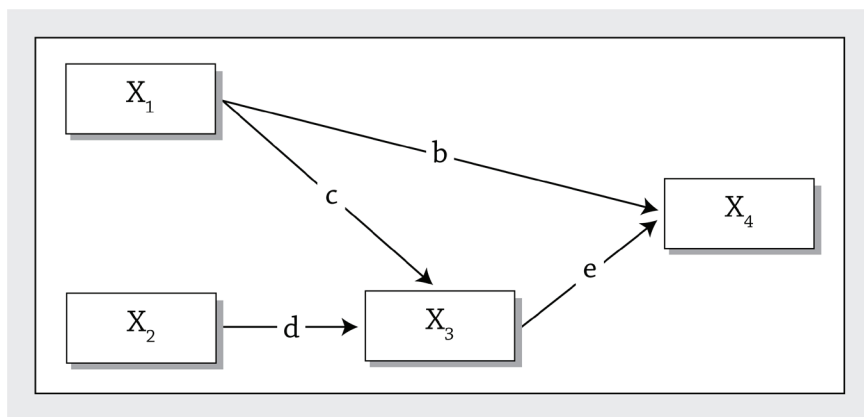
MEDIDAS DE AJUSTE DEL MODELO

JI-CUADRADA

La medida de ajuste basada en la prueba de Ji-cuadrada se basa en una comparación entre la matriz de varianza-covarianza obtenida con los datos del estudio, llamada la matriz de varianza covarianza empírica, y la matriz de varianza-covarianza que se infiere del modelo estructural, llamada la matriz de varianza-covarianza reproducida. Considere el modelo estructural simple ilustrado en la **Figura 6**. En ese modelo hay cuatro variables observadas X_1 a X_4 . La matriz de varianza-covarianza empírica, S , es la que se obtiene al calcular las covarianzas entre las variables:

Figura 6

Un modelo estructural simple (análisis de trayectorias).



Nota. Los términos de error en X_3 y X_4 fueron omitidos deliberadamente.

$$S = \begin{bmatrix} s_1^2 & s_{12} & s_{13} & s_{14} \\ s_{21} & s_2^2 & s_{23} & s_{24} \\ s_{31} & s_{32} & s_3^2 & s_{34} \\ s_{41} & s_{42} & s_{43} & s_4^2 \end{bmatrix}$$

En esta matriz los elementos s_{ij} se refieren a las covarianzas entre cualquier par de variables X_i y X_j . Por supuesto, la covarianza de una variable consigo misma es la varianza, es decir $s_{ii} = s_i^2$. Por ejemplo, s_{12} es la covarianza entre X_1 y X_2 . La covarianza no es otra cosa que una correlación no estandarizada: $s_{12} = r_{12}s_1s_2$, donde r_{12} es la correlación entre X_1 y X_2 y s_1 y s_2 son las desviaciones estándar de X_1 y X_2 .

La matriz de varianza-covarianza reproducida es la que se obtiene utilizando los estimados de los parámetros:

$$\hat{\Sigma} = \begin{bmatrix} s_1^2 & 0 & cs_1^2 & (b+ec)s_1^2 \\ 0 & s_2^2 & ds_2^2 & eds_2^2 \\ cs_1^2 & ds_2^2 & c^2s_1^2 + d^2s_2^2 & bcs_1^2 + eds_2^2 \\ (b+ec)s_1^2 & eds_2^2 & bcs_1^2 + eds_2^2 & (b^2 + 2bec)s_1^2 + de^2s_2^2 \end{bmatrix}$$

Una diferencia importante entre estas dos matrices son las celdas con ceros. Note, por ejemplo, que en esta matriz $S_{12} = S_{21} = 0$. La razón para ello es que en el modelo especificado en la **Figura 6** no existe relación alguna (flechas sencillas o dobles) entre las variables X_1 y X_2 . De esta manera el modelo impone restricciones a los posibles valores de la matriz reproducida.

Si se busca la diferencia entre estas dos matrices, obtendrán otra llamada la matriz de residuales, R:

$$R = \hat{\Sigma} - S$$

Una función, llamada la función de ajuste, \mathbf{F} , evaluada en \mathbf{R} se caracteriza porque su distribución es la de la estadística Ji-cuadrada:

$$(n - 1)F(R) = \chi_{n_p - c - n_9}^2$$

La función \mathbf{F} depende, a su vez, del método de estimación utilizado. Por ejemplo, para el método de minimización de suma de cuadrados ordinario (OLS), para el método de minimización de suma de cuadrados mínimos generalizado (GLS) y el método de verosimilitud múltiple (MLE), \mathbf{F} se define como:

$$F_{ML} = tr(S\Sigma^{-1}) - n_p + \ln|\Sigma| - \ln|S|$$

$$F_{GLS} = \frac{1}{2} tr \left[(S - \Sigma)S^{-1} \right]^2$$

$$F_{OLS} = \frac{1}{2} tr \left[(S - \Sigma)^2 \right]$$

En estas ecuaciones \mathbf{tr} se refiere a la suma de los elementos de la diagonal principal de las matrices, n_p es el número de parámetros del modelo y \mathbf{ln} se refiere al logaritmo natural.

ÍNDICES DE AJUSTE

La literatura sobre los índices de ajuste para los modelos estructurales es vasta y contiene un sinnúmero de dichos índices. Estos se clasifican en cuatro familias:

- De incremento Tipo I (ejemplos: NFI, BL86) - Comparan el modelo propuesto con el modelo nulo.
- De incremento Tipo II (ejemplos: TLI, NNFI, BL89) - Asumen normalidad y estimación mediante verosimilitud máxima (MLE).
- De incremento Tipo III (ejemplos: BFI, CFI) - Comparan el modelo propuesto con el modelo saturado.
- Absolutos (ejemplos: GFI, AGFI, AIC, Mc, SRMR, RMSEA) - Calculan la proporción de las varianzas y covarianzas observadas que explique el modelo. Útiles solo cuando se utiliza el método de estimación de máxima verosimilitud.

Schreiber y otros (2006) recomiendan reportar los valores de solo tres de estos índices: CFI, TLI y RMSEA. Los índices CFI y TLI tienen un rango de valores entre 0 y 1, donde 1 representa un ajuste perfecto. Por otra parte, el índice RMSEA, que es similar a un error estándar, tiene un rango de valores de cero hasta infinito. Hu y Bentler (1999) han establecido los puntos de corte para estos índices: 0.95, 0.96 y 0.06. Es decir, que para afirmar que un modelo tenga un buen ajuste a los datos, tanto CFI como TLI, deben ser mayores o iguales a 0.95, mientras que RMSEA debe ser menor o igual a 0.06.

TAMAÑO DE LA MUESTRA

El tamaño de la muestra requerido para un análisis estructural recomendado es de por lo menos 250 casos, aunque la realidad es que este número depende, entre otras cosas, de la complejidad del modelo. Otro aspecto que se debe tomar en consideración es que si se hicieran ajustes al modelo, no habría garantía alguna de que estos ajustes fueran significativos en otra muestra tomada de la misma población. Si el investigador planifica hacer ajustes a un modelo, entonces debe considerar tomar una muestra de mayor tamaño, y dividirla aleatoriamente en dos muestras, usando una para crear un modelo ajustado y la otra para confirmar si los ajustes son significativos.

Muthen y Muthen (2002) recomiendan el uso de simulaciones para establecer el tamaño de la muestra para un modelo dado. En su artículo presentan código en el lenguaje M+ para llevar a cabo este análisis.

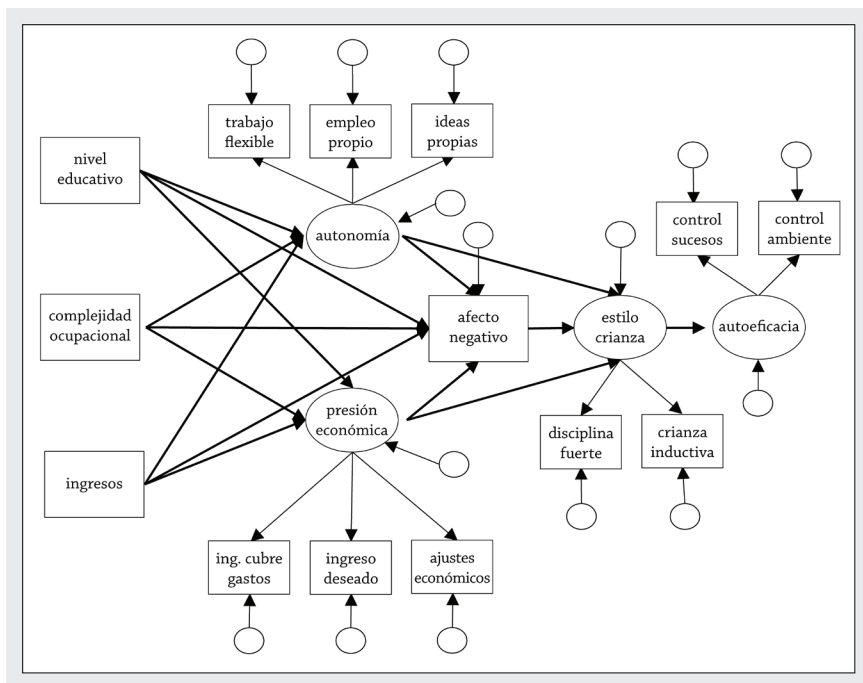
EJEMPLO DE MODELO ESTRUCTURAL

El siguiente ejemplo es tomado del artículo de Whitbeck y otros (1997) en el cual se estudia la relación entre las condiciones de trabajo de los padres y la necesidad económica del hogar sobre las prácticas de crianza y la autoeficacia de los hijos. El modelo hipotetizado está representado gráficamente en la **Figura 7**. Este modelo inicialmente resultó no ajustarse a los datos del estudio, por lo que hubo que realizar varios ajustes al mismo. El resultado se ilustra en la **Figura 8**. En esta figura se han eliminado del diagrama las relaciones (flechas) en las cuales los parámetros no fueron significativamente distintos de cero. Para este

modelo ajustado, la estadística Ji-cuadrada tuvo un valor de 41.56 con un valor de p de 0.97 la cual no es estadísticamente significativa. Los datos obtenidos se ajustan al modelo modificado. Dado que el tamaño de la muestra de este estudio fue bajo ($N = 158$) no fue posible validar el modelo final con una muestra independiente.

Figura 7

Modelo estructural propuesto por Whitbeck y otros (1997).



Los análisis en este ejemplo fueron llevados a cabo en el programa M+. El código utilizado se presenta en la **Tabla 1**.

Tabla 1

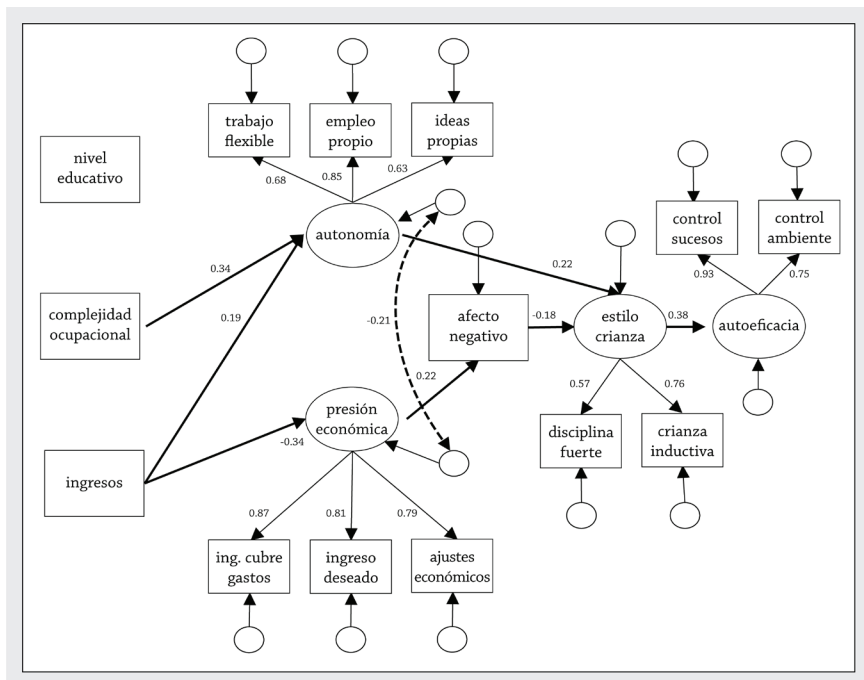
Código de M+ utilizado para el análisis de los datos de este ejemplo.

```

TITLE:                               Family Conditions
DATA:                                 FILE IS fam_cond.prn;
                                     TYPE IS COVARIANCE;
                                     NOOBSERVATIONS = 158;
                                     NAMES ARE v01-v14;
VARIABLE:                             USEVARIABLES v01-v14;
MODEL:                                Auto BY v04 v05 v06;
                                     Pres BY v07 v08 v09;
                                     Esti BY v11 v12;
                                     AuEf BY v13 v14;
                                     Auto ON v01 v02 v03;
                                     Pres ON v01 v02 v03;
                                     v10 ON v01 v02 v03 Auto Pres;
                                     Esti ON Auto v10 Pres;
                                     AuEf ON Esti;
OUTPUT:                               STANDARDIZED; MODINDICES;
    
```

Figura 8

Modelo final ajustado a los datos obtenidos por Whitbeck y otros.



RESUMEN

El análisis de ecuaciones estructurales es una metodología de análisis flexible con la cual se puede llevar una gran variedad de modelos estadísticos. En este escrito se ha visto cómo se representa gráficamente un modelo estructural, incluyendo el análisis de regresión, análisis de trayectorias, análisis factorial, análisis de regresión con variables latentes y modelos estructurales generales. Se discutieron, además, los pasos generales para llevar a cabo un análisis estructural y las estadísticas de ajuste del modelo a los datos empíricos. Se presentó un ejemplo de un modelo estructural y el código utilizado para el análisis en el lenguaje M+.

Es posible utilizar el modelaje mediante ecuaciones estructurales para llevar a cabo análisis equivalentes a la prueba t, análisis de varianza, análisis longitudinales (medidas repetidas) y curvas de crecimiento. Existe una gran variedad de programas para llevar este tipo de análisis tales como M+, LAVAAN (paquete en el lenguaje R), Stata, EQS, AMOS (SPSS), entre otros.

REFERENCIAS

- Hu, L. and Bentler, P. (1999). "Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives". *Structural Equation Modeling: A Multidisciplinary Journal*, 6, 1-55.
- Muthen, L. and Muthen, B. (2002). "How to Use a Monte Carlo Study to Decide on Sample Size and Determine Power", *Structural Equation Modeling: A Multidisciplinary Journal*, 9:4, 599-620.
- Schreiber, J. B., Stage, F. K., King, J., Nora, A., and Barlow, E. A. (2006). "Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review". *The Journal of Educational Research*, 99(6), 323-337.
- Whitbeck, L. B., Simons, R. L., Conger, R. D., Wickrama, K. A. S., Ackley, K. A., y Elder, G. H. (1997). "The Effects of Parent's Working Conditions and Family Economic Hardship on Parenting Behaviors and Children's Self-Efficacy". *Social Psychology Quarterly*, 60(4), 291-303

Árboles de clasificación y regresión

*José Noel Caraballo, Ph.D.
asesor del Centro de Apoyo Estadístico
Pontificia Universidad Católica de Puerto Rico*

INTRODUCCIÓN

Los métodos de modelaje clásicos se basan en el supuesto de que los datos fueron generados mediante un proceso estocástico. Un modelo típico especifica la relación entre las variables de respuestas (variables dependientes) y los predictores (variables independientes) en términos de algunos parámetros (desconocidos) y algún término de error. Los datos obtenidos se consideran como generados aleatoriamente por este modelo. Las pruebas de bondad de ajuste o los análisis de residuos son utilizados luego para determinar cuán bueno es el modelo en explicar los datos observados. Una estrategia alterna de modelaje consiste en tratar de encontrar un conjunto de reglas de decisión que opere sobre las variables independientes para predecir las dependientes sin hacer supuestos acerca de los mecanismos que generen los datos (Breiman, 2001). Este “modelaje algorítmico” está ejemplificado por los árboles de decisión, los cuales generan el conjunto de reglas de decisión haciendo una partición recursiva de los datos.

Los árboles de decisión, que incluyen a los árboles de clasificación y a los de regresión, son, junto al análisis de regresión y al análisis de conglomerados, uno de los pilares del campo de la inteligencia artificial o, más específicamente, del aprendizaje de máquinas. En este escrito se discuten el algoritmo de árboles de clasificación y regresión (CART) de Breiman y los conceptos básicos de esta forma de modelaje. Finalmente, el método se ilustra con un ejemplo tomado del área de la educación superior.

PARTICIÓN RECURSIVA

La partición recursiva es un procedimiento mediante el cual un conjunto de datos se divide en subconjuntos que sean cada vez más homogéneos. Esta estrategia es particularmente útil para tratar problemas de clasificación, pero también podría ser utilizada en situaciones que normalmente se tratarían usando modelos de regresión. Típicamente los resultados de este procedimiento se presentan como un dendograma o árbol de decisión invertido. Este tipo de árbol se caracteriza por ser binario: cada grupo se divide en dos y solo dos subgrupos.

Considere el grupo de quince elementos presentados en la **Figura 1**. Estos varían en tres características: color (rojo, azul y amarillo); forma (círculo, cuadrado y triángulo); y tamaño (grande y pequeño). Para subdividirlo en conjuntos más homogéneos en términos de su color (variable dependiente), se podría tomar cualquiera de las otras dos variables (variables independientes) para segmentarlos. Por ejemplo, si se usara “tamaño”, el grupo de objetos quedaría dividido en dos grupos: objetos grandes y objetos pequeños (véase la **Figura 2**). Note que Grupo 1 es predominantemente azul, es decir, más homogéneo en cuanto el color. En el Grupo 2, sin embargo, aún no predomina ningún color.

Figura 1

Conjunto inicial de objetos.

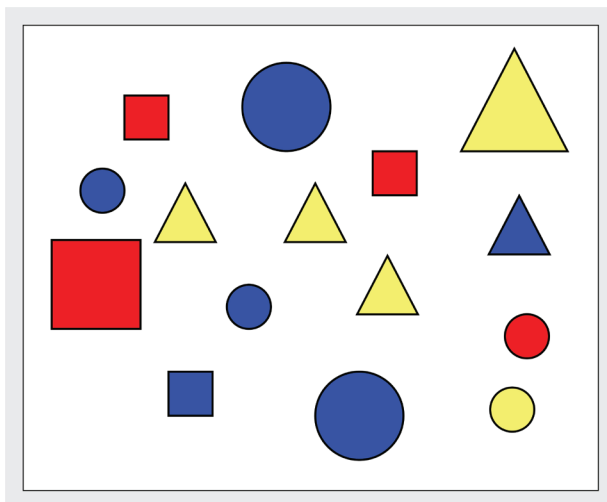
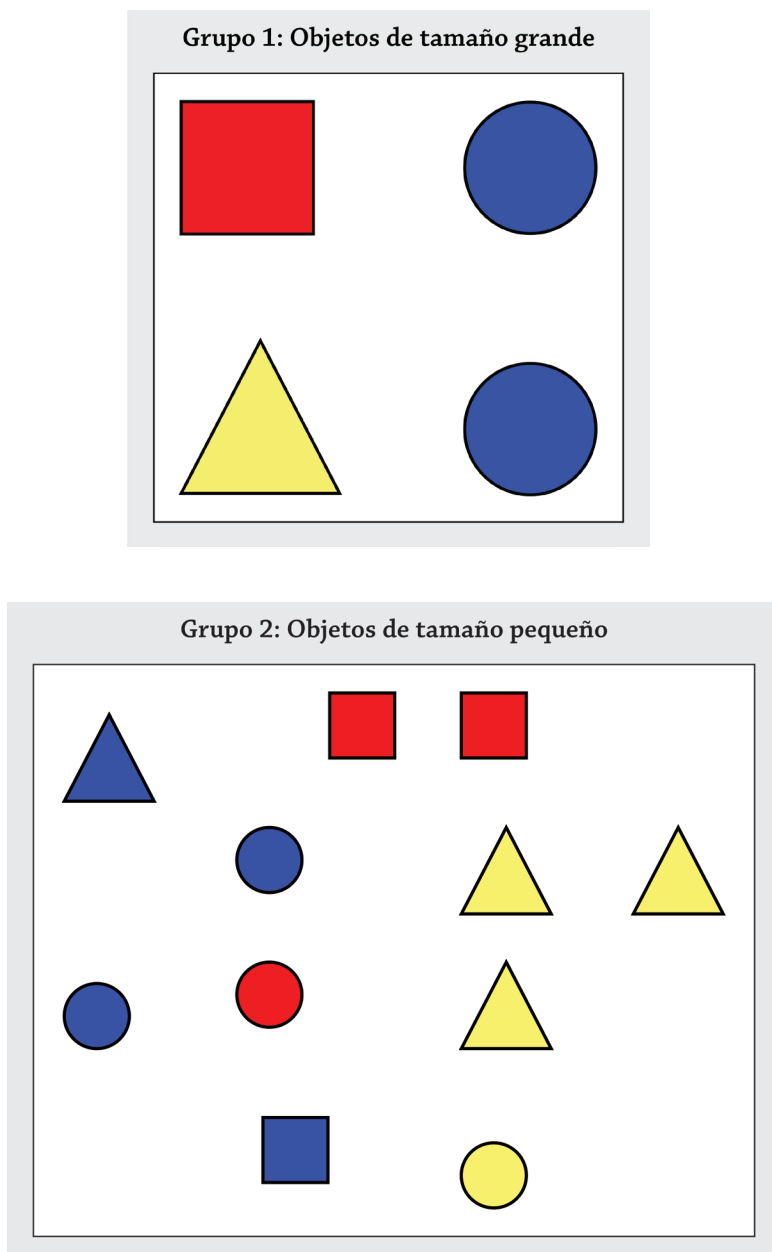


Figura 2

Partición de los diez objetos de acuerdo a su tamaño.



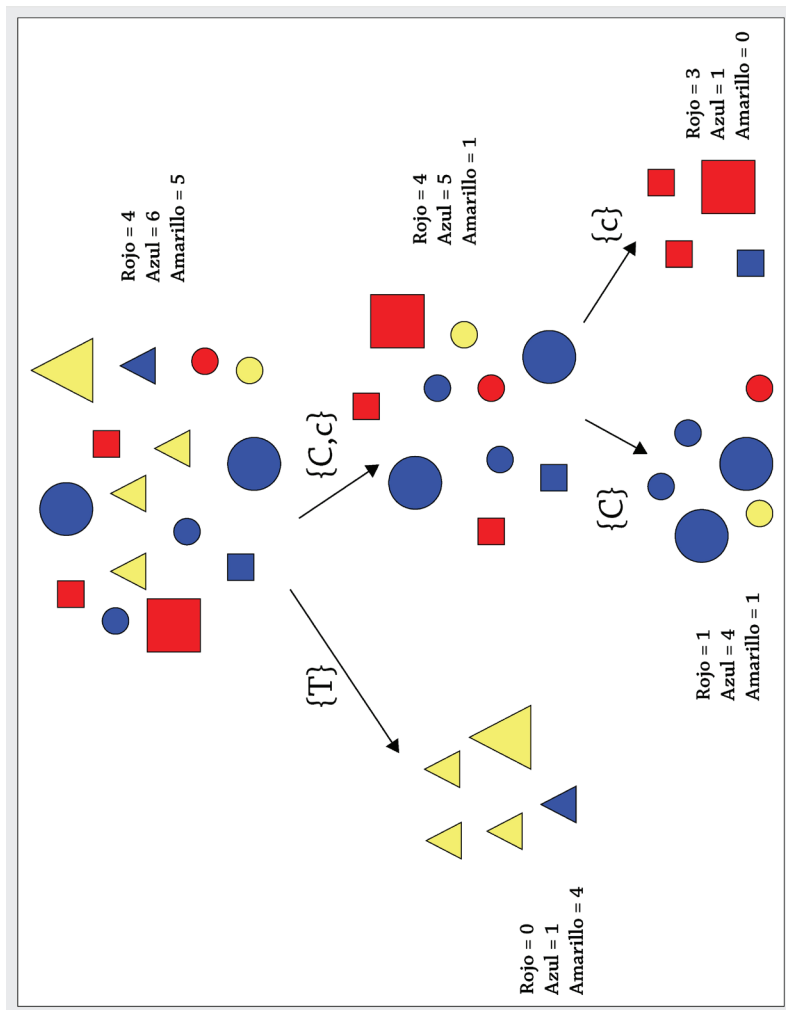
Ahora se podría proceder a subdividir cada uno de estos dos grupos hasta utilizar todas las variables disponibles, llegar a un punto en donde el subconjunto fuera totalmente homogéneo en su color, o simplemente detenerse en algún punto basado en algún criterio establecido *a priori*. La **Figura 3** muestra un posible árbol de clasificación que utiliza “forma” para realizar la primera partición. En este diagrama se puede apreciar cómo los subgrupos se hacen cada vez más homogéneos. Observe que junto a las flechas se ha escrito, entre corchetes, la regla utilizada para la subdivisión. Para una variable categórica el número de formas en que se puede dividir un conjunto de datos de forma binaria es $2^{k-1} - 1$ en donde k es el número de categorías de esta variable. Por ejemplo, para “forma” hay tres categorías: triángulo {T}, círculo {C} y cuadrado {c}. En este caso $k = 3$, por lo que $2^{k-1} - 1 = 2^{3-1} - 1 = 4 - 1 = 3$. Hay entonces tres formas de dividir el conjunto inicial: {T}{C,c}, {C}{T,c} y {c}{T,C}. En la **Figura 3** la regla utilizada para la primera división fue {T}{C,c}, es decir, separar los triángulos de las círculos y cuadrados.

En la nomenclatura de árboles de decisión se dice que cada grupo del diagrama del cual sale o entra una flecha es un “nodo”. Los subgrupos que no se dividan más se llaman “nodos terminales”. En la **Figura 3** hay tres nodos terminales. Es en estos nodos en donde se lee la clasificación predicha (la de mayor frecuencia). Así que, en este ejemplo, los datos se dividieron en tres grupos: rojo, azul y amarillo.

¿Cuáles son las reglas de clasificación en este caso? Hay tres: 1) si el objeto es un triángulo, se predice que su color será amarillo; 2) si el objeto no es un triángulo, pero es un círculo, se predice que su color será azul; y 3) si el objeto no es un triángulo, pero es cuadrado, se predice que su color será rojo.

Figura 3

Árbol de clasificación para predecir el color de un conjunto de figuras geométricas.



El árbol de regresión se construye de igual manera. La diferencia consiste en que, en vez de categorías en un nodo, se tendría un conjunto de valores numéricos para la variable dependiente. Los valores en un nodo se promediarían para obtener el valor predicho.

EL ALGORITMO DE CART

El algoritmo de árboles de clasificación (Breiman, Friedman, Olshen y Stone, 1984) llamado CART es una de las implementaciones más conocidas que utiliza un método no paramétrico para “crecer” los árboles. El algoritmo de CART consiste en un conjunto de reglas para: a) dividir cada nodo, b) decidir cuándo está completo un árbol, y c) asignar una categoría o un valor predicho a cada nodo terminal. La división de un conjunto de datos se basa en la minimización de una función de impureza y continúa hasta que los datos en cada nodo sean homogéneos o queden muy pocos casos disponibles en el nodo. Este árbol “máximo” entonces se poda hasta obtener un árbol con mínimo error de clasificación.

Para un conjunto de datos que consista en una variable dependiente Y y J variables independientes X_j , CART considera todas las posibles divisiones para todas las variables independientes X_j y las ordena a base de la calidad del criterio de división (Gini o Twoing para variables dependientes categóricas y mínimos cuadrados o desviaciones absolutas para variables continuas.) Para una variable independiente continua CART divide los datos $n - 1$ veces, donde n es el número de valores distintos de la variable en el conjunto de datos. Cada uno de los $n - 1$ divisiones se llevan a cabo en un punto x_{in} que queda en medio de dos valores consecutivos $x_{j(m)}$ y $x_{j(m+1)}$. Para una variable categórica con k categorías, CART examina todos los posibles $(2^{k-1} - 1)$ subconjuntos posibles de los valores de X_j .

Considere el conjunto de datos presentado en la **Tabla 1**. Aquí Y_1 y Y_2 representan variables dependientes categóricas y continuas, respectivamente; las variables, X_1 y X_2 son variables independientes categóricas; y X_3 es una variable independiente continua. Como X_1 tiene cuatro categorías, habrá $2^{4-1}-1 = 7$ divisiones a explorar. De igual forma para X_2 habrá $2^{2-1}-1 = 1$ división. La variable continua X_3 tiene siete valores distintos ($\{4,5,6,7,8,9,10\}$), así que tendrá $7-1 = 6$ divisiones.

Para construir un árbol de clasificación, la variable dependiente debe ser categórica, como por ejemplo Y_1 en la **Tabla 1**. Una posible función de impureza es el índice de diversidad Gini el cual se define para un nodo t como $i(t) = 1 - \sum_{j=1}^J p^2(j|t)$. Aquí $p(j|t)$ es la proporción de la clase o categoría j en el nodo t . Por ejemplo, el nodo raíz contendrá todos los veinte datos (véase la **Tabla 1**). La categoría 1 aparece siete veces, la 2 cinco veces y la 3 ocho veces, por lo que $p(1|\text{raíz})=7/20$, $p(2|\text{raíz}) = 5/20$, y $p(3|\text{raíz}) = 8/20$. Si se utiliza la variable X_2 para hacer la división se verá que solo hay una división posible para dividir los datos de Y_1 : $\{1,1,1,1,2,2,2,2,3,3,3\}$ y $\{1,1,1,2,3,3,3,3,3\}$. Para los datos en el primero de estos dos grupos o nodos tiene que $p(1|1) = 4/11$, $p(2|1) = 4/11$, y $p(3|1) = 3/11$, así que $i(1) = 1-(4/11)-(4/11)-(3/11) = 0.66$. De forma similar, para el segundo nodo $i(2) = 1-(3/9)-(1/9)-(5/9) = 0.57$.

De entre todas las posibles divisiones de los valores de las variables dependientes el criterio GINI selecciona aquel que minimiza a $\Delta i = i(t) - p_1 i(1) - p_2 i(2)$. En el ejemplo anterior $\Delta i = 0.4 - (11/20)(0.66) - (9/20)(0.57) = 0.036$. La **Tabla 2** presenta los valores de Δi para las 14 posibles divisiones en el nodo raíz. En esta tabla puede verse que la variable X_3 será seleccionada para hacer la división en el nodo raíz, con la división hecha para $X_3 \leq 7.5$. La función de impureza GINI es una de varias funciones que se pueden utilizar para hacer las divisiones en un árbol de clasificación. El lector interesado puede consultar a Loh y Shih (1997) para una discusión de otras alternativas.

Para construir un árbol de regresión para la variable dependiente continua Y_2 la mejor división es la que minimiza la varianza pesada de los dos subnodos: $p_1 s_1^2 + p_2 s_2^2$. Como se puede apreciar en la última columna de la **Tabla 2**, X_3 será seleccionada como la variable para llevar a cabo la división, la cual será $X_3 \leq 5.5$.

Tabla 1

Conjunto de datos utilizados en el ejemplo (N = 20).

Y_1	Y_2	X_1	X_2	X_3	Y_1	Y_2	X_1	X_2	X_3
1	0.367	1	1	10	3	0.550	1	2	9
2	0.434	1	2	7	3	0.300	1	2	10
1	0.653	2	1	5	2	0.530	1	1	7
2	0.607	2	1	7	2	0.558	2	1	9
3	0.573	3	2	8	1	0.640	3	2	6
3	0.449	3	2	8	3	0.510	3	2	9
3	0.413	2	1	8	3	0.380	3	1	8
3	0.520	2	1	9	2	0.466	2	1	6
1	0.650	4	1	6	1	0.690	4	2	5
1	0.803	4	1	5	1	0.740	4	2	4

Este procedimiento de división se repite para cada subnodo hasta que se obtiene el árbol máximo, esto es, hasta que no es posible llevar a cabo ninguna otra división porque el nodo es puro, o contiene un número de casos menor que un valor preespecificado. El árbol máximo claramente tiende a sobreajustarse a los datos, así que CART comienza un proceso de poda o eliminación de ramas hasta que produce todos los posibles subárboles. Haciendo uso de una muestra de prueba independiente se puede entonces calcular la tasa de error de cada subárbol. El árbol con el error más pequeño es declarado el “mejor” árbol. Cuando el conjunto original de datos es muy pequeño, no resulta práctico dividirlo en un conjunto de aprendizaje y otro de prueba, se utiliza entonces un procedimiento de validación cruzada. Este procedimiento consiste en dividir el conjunto original aleatoriamente en pedazos de igual tamaño, usualmente entre 10 y 20. Cada pedazo se utiliza entonces como un conjunto de datos de prueba y los restantes como el conjunto de aprendizaje.

Para un árbol de clasificación, la tasa de error se calcula por medio del estimado de resustitución (R.E.), o proporción de casos mal clasificados. El cálculo de R.E. utilizando los datos en el conjunto de aprendizaje es un estimado inocente y no se recomienda. La **Figura 4** muestra el árbol de

clasificación construido por CART para los datos de la **Tabla 1**. Hay tres casos mal clasificados, así que R.E. será igual a $3/20$, o 0.15 . Como se mencionó antes, un estimado más certero se puede obtener si se utiliza un conjunto de datos de prueba o por medio de validación cruzada. Para estos datos, el valor obtenido mediante validación cruzada es de 0.571 que es sustancialmente mayor que el estimado inocente de 0.15 .

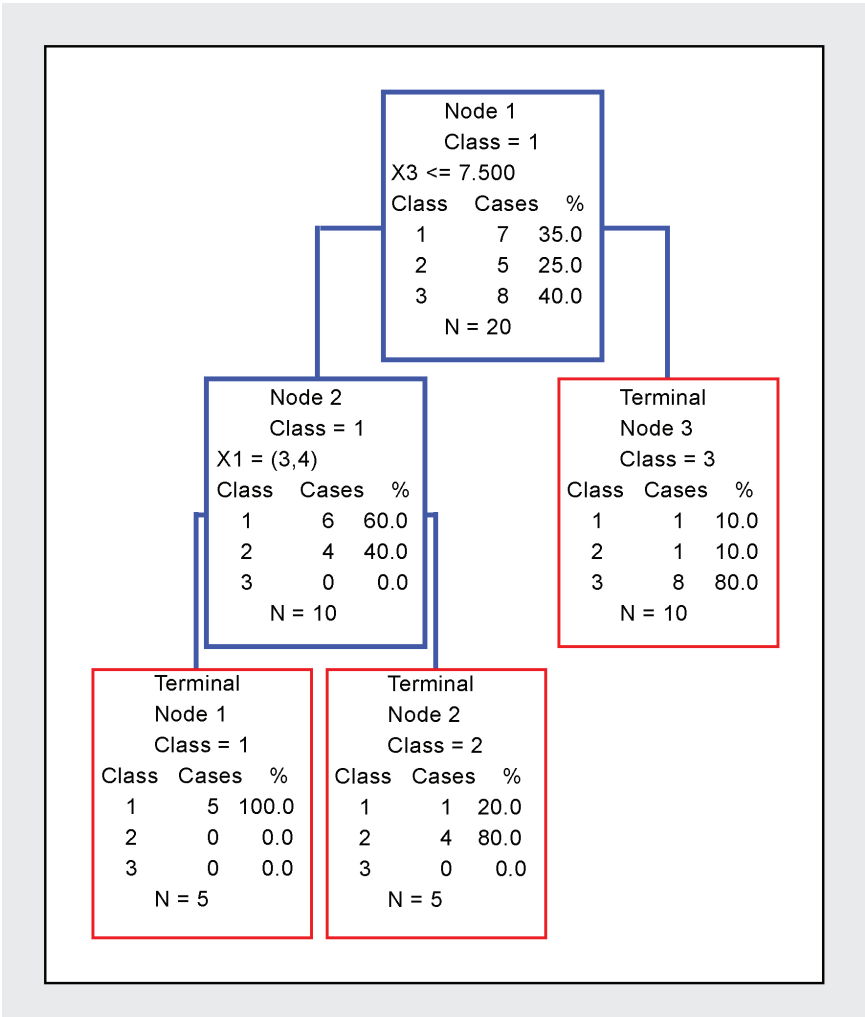
Tabla 2

Criterios para dividir los valores de las variables dependientes Y_1 and Y_2 de la Tabla 1.

<i>Splitting Variable</i>	<i>Split</i>	Δ_i	<i>Weighted variance</i>
X_1	1-234	0.015	0.0140
	2-134	0.043	0.0179
	3-124	0.082	0.0176
	4-123	0.149	0.0091
	12-34	0.087	0.0148
	13-24	0.065	0.0130
	14-23	0.061	0.0179
X_2	1-2	0.036	0.6972
X_3	≤ 4.5	0.034	0.0151
	≤ 5.5	0.161	0.0090
	≤ 6.5	0.231	0.0094
	≤ 7.5	0.395	0.0112
	≤ 8.5	0.048	0.0155
	≤ 9.5	0.011	0.0124

Figura 4

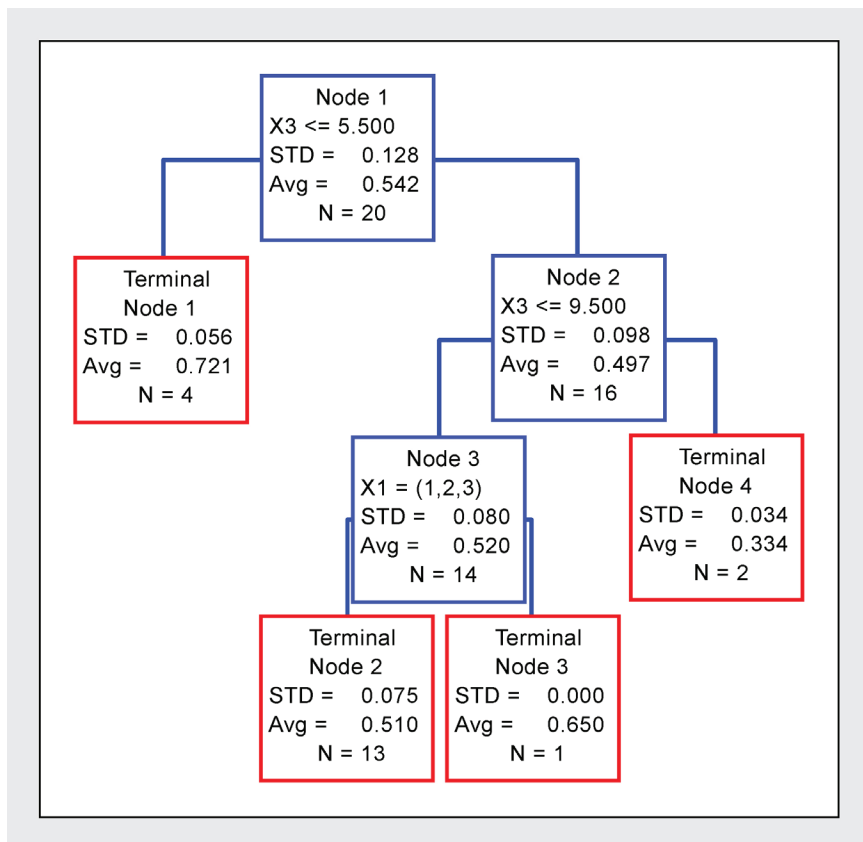
Árbol de clasificación para los datos de la Tabla 1.



Para un árbol de regresión, la tasa de error es el error cuadrado medio (*mean square error*). De nuevo, si se calcula con el conjunto de datos de aprendizaje se obtiene un estimado inocente. El árbol de regresión para estos datos está presentado en la **Figura 4**. En este ejemplo el estimado inocente es 0.029, mientras que el obtenido mediante validación cruzada es de 0.767.

Figura 5

Árbol de regresión para los datos de la Tabla 1.



EJEMPLO: FACTORES QUE DETERMINAN SI UN ESTUDIANTE COMPLETA SU GRADO EN SEIS AÑOS O MENOS

Este ejemplo es tomado de un estudio cuyo propósito era predecir si un estudiante completará su grado en seis años y medio o menos. En el mismo participaron 5,240 estudiantes universitarios de todas las disciplinas. Como predictores se utilizaron el índice de escuela superior (GPA_ADM), las puntuaciones en el examen del College Board en aptitud matemática (MATH_APT), aptitud verbal (VERBAL_APT), aprovechamiento en inglés (ENGLISH_ACH), aprovechamiento en

español (SPANISH_ACH) y aprovechamiento en matemáticas (MATH_ACH); el nivel educativo del padre (FATHER_EDU), el nivel educativo de la madre (MOTHER_EDU); el ingreso familiar (FAMILY_INCOME); el tipo de escuela: pública o privada (SCHOOL_TYPE); el programa de estudios al que fue admitido inicialmente (DEPARTMENT), y si era estudiante universitario de primera generación (FIRST_GENERATION). La variable dependiente era una dicótoma: si completó o no completó el grado en 6.5 años. Dado que la variable dependiente en este caso es categórica, el análisis a realizarse fue uno de árboles de clasificación.

El árbol de regresión resultante se presenta en la **Figura 5**. En el nodo raíz se puede ver que para el análisis se consideraron solo 4,174 estudiantes que tenían la información completa. De estos, 2,545 (61%) no completaron su grado en 6.5 años o menos y 1,629 (39%) sí lo hicieron (véase la **Figura 7**). La primera regla para subdividir este conjunto de datos es que el índice de ingreso (GPA_ADM) sea menor o igual a 3.365. De todas las variables y todos los posibles puntos de corte, esta es la regla que produce, en primera instancia, los dos subgrupos más homogéneos posibles.

Si se considera el subgrupo en la rama de la izquierda, se puede observar que de los 2,131 estudiantes 1,512 (71%) no completaron su grado en 6.5 años o menos, mientras que 619 (29%) sí lo hicieron (véase la **Figura 8**). La regla para subdividir este nuevo grupo es el índice de ingreso que fuese menor o igual a 2.915. La **Figura 8** muestra la composición del segundo subgrupo y su correspondiente regla de partición.

Note en la **Figura 6** que hay doce nodos terminales y que en la mayoría de ellos predomina una de las dos categorías, ilustradas con la banda de color rojo y azul. De las **Figuras 6** y **7** también se puede apreciar que las reglas de partición utilizan, en general, distintas variables independientes en cada nivel. Es decir, el árbol muestra que existen interacciones entre las distintas variables independientes.

En este ejemplo, el error en la clasificación inocente es de 27%. El estimado del error, utilizando el método de validación cruzada es de 47%. La **Figura 4** muestra cómo varía el error de clasificación con la estructura del árbol. Note que el valor de 47% es el error mínimo y ocurre para el árbol ilustrado en la **Figura 5**. Aun cuando el error es

relativamente alto, note que en varios de los nodos terminales los errores de clasificación son pequeños. Así que, aunque un árbol pueda tener un error de clasificación alto, es posible que para algunos nodos el modelo haga buenas predicciones para algunas combinaciones de predictores

DISCUSIÓN

Los árboles de clasificación y regresión pueden ser una alternativa a métodos paramétricos tales como el análisis discriminante y análisis de regresión. Entre las ventajas de este método de análisis se encuentra el que, por ser un método no paramétrico, no hay supuestos acerca de las distribuciones de las variables bajo análisis. Además, la precisión de este método tiende a ser igual o mejor que el de los métodos paramétricos. Los árboles de clasificación y regresión tienen la ventaja de ser fáciles de interpretar y utilizar. Para hacer una predicción en un caso nuevo, bastaría con aplicar las reglas de decisión del árbol las cuales son simples comparaciones. En los métodos de regresión habría que sustituir los valores en una fórmula matemática que pudiera ser más o menos compleja.

Este método de análisis maneja las variables categóricas fácilmente, sin importar cuántas categorías tenga. Por ejemplo, la variable categórica DEPARTAMENTO, utilizada en el ejemplo, es una variable que tiene cerca de 80 categorías. Para usarla en un análisis de regresión o discriminante habría que crear 79 variables nuevas, llamadas *dummy variables*. De la misma forma, tampoco importa cuántas categorías tenga la variable dependiente. Esta es una gran ventaja sobre el análisis discriminante, para el cual hay que crear $n - 1$ funciones discriminantes, donde n es el número de categorías de la variable dependiente.

Finalmente, en los árboles de regresión se pueden visualizar fácilmente las interacciones entre las variables predictoras. Las interacciones entre variables están presentes cada vez que la regla de decisión en un nivel hace uso de una variable distinta a la del nodo en el nivel superior o en el inferior. De esta forma, aun cuando el investigador hubiera decidido usar un análisis paramétrico, sería conveniente hacer un análisis de clasificación o regresión para determinar así qué interacciones entre variables incluir en su modelo.

Figura 6
 Arbol de clasificación para predecir si un estudiante completará su grado en 6.5 años o menos.

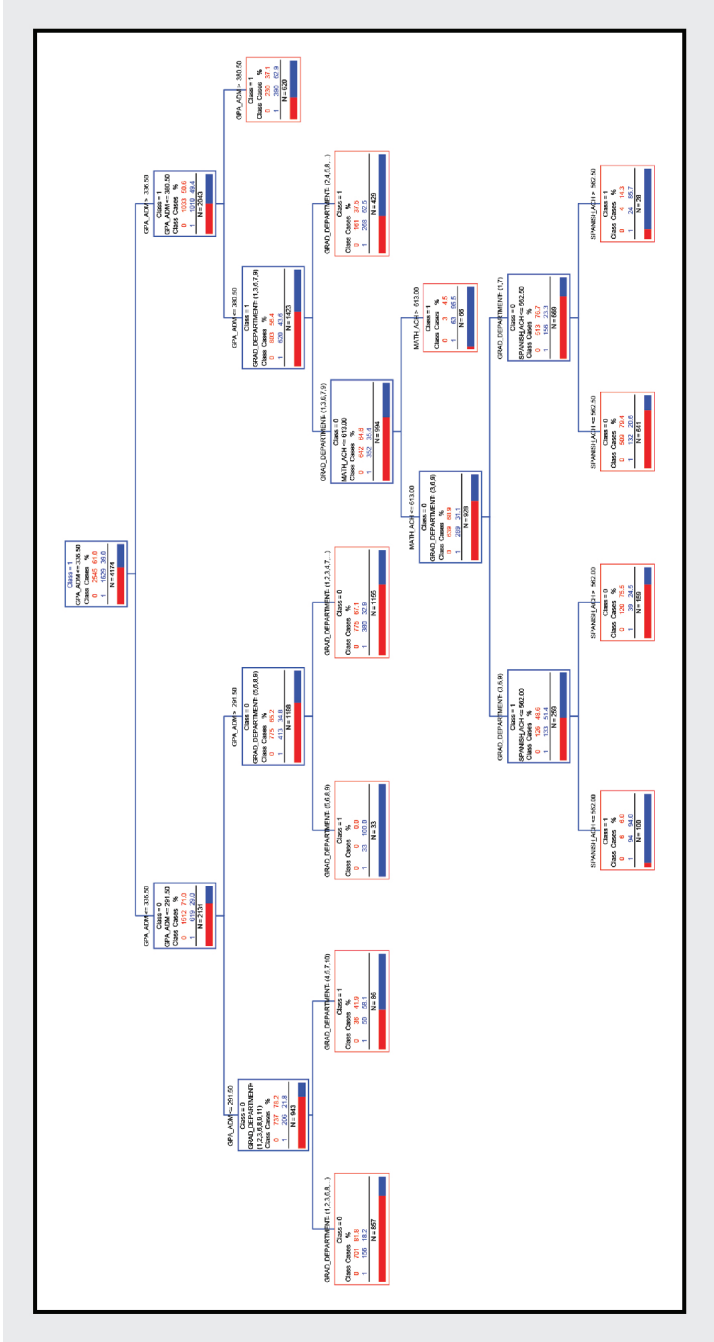


Figura 7
 Nudo raíz.

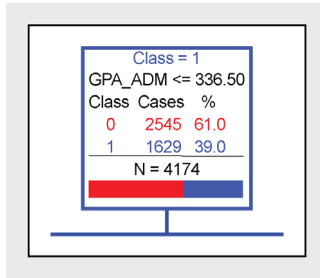


Figura 8
 Rama izquierda del árbol de clasificación.

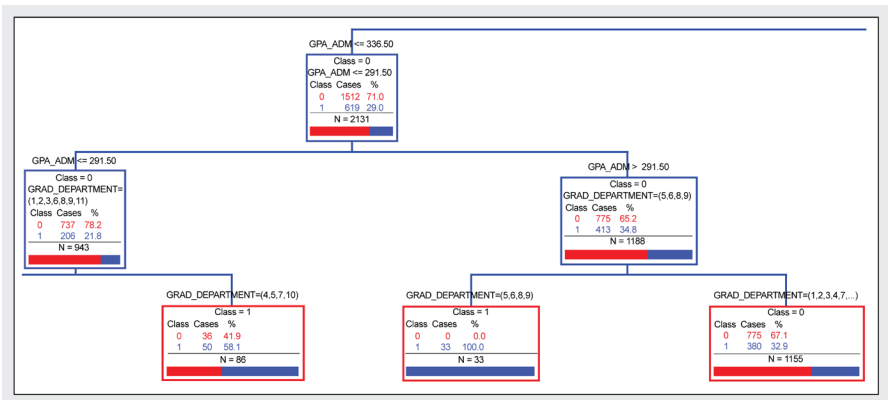
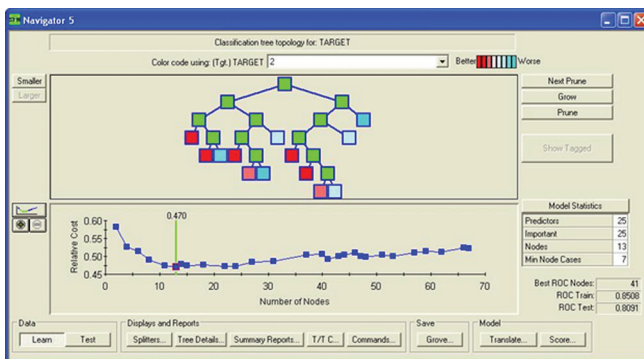


Figura 9
 Gráfica del error de clasificación versus la complejidad del árbol de clasificación.



REFERENCIAS

- Breiman, L., (2001). *Statistical Modeling: The Two Cultures*. *Statistical Science*, 16(3), 199-231.
- Breiman, L., Friedman, J., Olshen, R. and Stone, C. (1984). *Classification and Regression Tree*. Belmont, CA: Wadsworth.
- Loh, W.-Y. and Shih, Y.-S. (1997). *Split selection methods for classification trees*. *Statistica Sinica*, 7, 815-840.

Introducción a la estadística bayesiana

*José Noel Caraballo Ríos, Ph.D.
asesor del Centro de Apoyo Estadístico
Pontificia Universidad Católica de Puerto Rico*

La estadística trata sobre la recopilación de datos, su análisis y su interpretación con el propósito de realizar inferencias acerca del proceso que dio lugar a los datos. Este proceso se representa mediante un modelo matemático que incluye uno o más parámetros desconocidos tales como la media y la desviación estándar poblacional de una variable de interés. La idea general consiste en hacer uso de los datos y del modelo matemático para inferir los valores de los parámetros del modelo. Existen dos acercamientos distintos para el problema de la inferencia: la estadística clásica o frecuentista y la estadística bayesiana. La mayoría de las personas que toma un curso de estadística por vez primera usualmente aprende sobre la estadística clásica. Los métodos de análisis bayesianos, por su parte, usualmente se estudian en cursos más avanzados.

El nombre de este método proviene de Thomas Bayes (1702-1761) quien era un matemático británico y sacerdote presbiteriano. El fundamento de este método es un teorema, llamado Teorema de Bayes, que permite calcular la probabilidad de lo que se conoce como la probabilidad inversa. En este escrito se tratarán algunas de las ideas principales de la estadística bayesiana y se ilustrará sobre su uso con un ejemplo de análisis de regresión utilizando el lenguaje de programación R.

LA INTERPRETACIÓN DE LA PROBABILIDAD EN LA TEORÍA CLÁSICA

La distinción más importante entre la teoría clásica y la bayesiana es en cómo se interpreta la probabilidad de que ocurra un suceso. En la

teoría clásica, o estadística frecuentista, la probabilidad de que ocurra un suceso se refiere a la fracción de veces que ocurre el suceso cuando un experimento se repite un número grande de veces. Por ejemplo, suponga que se desea determinar la probabilidad de obtener un 3 al lanzar un dado de 6 caras. En la estadística frecuentista, la manera de calcular esta probabilidad sería lanzando un dado un número grande de veces y determinando la frecuencia relativa con la que el suceso de interés, obtener un 3, ocurra (véase la **Tabla 1**).

Tabla 1

Cálculo de la probabilidad de obtener un 3 al lanzar un dado, utilizando la idea de frecuencia relativa.

cuatro lanzamientos:	{6,3,2,6}	$p = 1/4 = 0.25$
diez lanzamientos:	[6,3,2,6,5,5,2,5,3,1]	$p = 2/10 = 0.20$
muchos lanzamientos	[6,3,2,6,5,5,2,5,3, ...]	$p = \frac{\text{Total de '3'}}{\text{Total de lanzamientos}} = 0.16\bar{6}$

Suponga que se lleva a cabo un experimento para determinar si un tratamiento experimental es efectivo (dos grupos) y se encuentra que la diferencia entre los promedios favorece al tratamiento y que el valor obtenido de la probabilidad es de 0.035. La interpretación frecuentista de esta probabilidad es:

Si este experimento se repite un número grande de veces, asumiendo que la hipótesis nula es la correcta, en el 3.5% de los casos se esperarían obtener una diferencia entre las medias igual o mayor a la observada en el experimento realizado.

De igual forma, la interpretación de un intervalo de confianza se hace en términos de frecuencias relativas. Por ejemplo, suponga que en un estudio se encuentra que la proporción de personas que votará por el partido X en las próximas elecciones está dentro del siguiente intervalo de confianza (al 95%): $0.35 < P < 0.40$. La interpretación de este intervalo de confianza es:

Si se realizara el estudio un número grande de veces y se crearan los intervalos de confianza al 95% cada vez, entonces en el 95% de las veces, estos intervalos contendrían el valor “verdadero” de la proporción de votantes que votarán por el partido X.

LA INTERPRETACIÓN DE LA PROBABILIDAD EN LA TEORÍA BAYESIANA

A diferencia de la teoría clásica, la probabilidad en la teoría bayesiana se interpreta como una creencia subjetiva, como cuando se está en una situación en la que se hará una apuesta. En ese caso, la probabilidad se basa en cuánto conocimiento tenga la persona y cuánto estaría dispuesta a apostar. La probabilidad subjetiva es la que permite contestar preguntas tales como ¿irá a llover hoy? En estos dos ejemplos, el estimado de la probabilidad depende de cuánta información previa se tenga. Por ejemplo, si se mira por la ventana y se ve el cielo nublado el estimado de la probabilidad de lluvia sería mayor que si el cielo estuviera claro. La probabilidad de que ocurran dos sucesos simultáneamente se conoce como probabilidad conjunta y es la base de la teoría bayesiana.

Considere la probabilidad conjunta de dos sucesos, X y Y, como por ejemplo tener cáncer y fumar. Esta probabilidad se representa por:

$$p(X, Y)$$

ó

$$p(\text{cáncer}, \text{fumar})$$

La probabilidad conjunta está dada por:

$$p(X, Y) = p(X | Y)p(Y)$$

donde $p(X | Y)$ representa la probabilidad condicionada, es decir, la probabilidad de que X ocurra si Y ha ocurrido antes. Note que el orden de X y Y en la probabilidad conjunta no importa:

$$p(X, Y) = p(Y, X)$$

así que podría escribirse también.

$$p(Y, X) = p(Y|X)p(X) \quad 3$$

De esta forma, en este ejemplo

$$p(\text{cáncer}, \text{fumar}) = p(\text{cáncer} | \text{fumar})p(\text{fumar})$$

o

$$p(\text{fumar}, \text{cáncer}) = p(\text{fumar} | \text{cáncer})p(\text{cáncer})$$

Igualando las probabilidades anteriores se encuentra que:

$$p(\text{cáncer} | \text{fumar})p(\text{fumar}) = p(\text{fumar} | \text{cáncer})p(\text{cáncer})$$

Rearreglando términos se obtiene:

$$p(\text{cáncer} | \text{fumar}) = \frac{p(\text{fumar} | \text{cáncer})p(\text{cáncer})}{p(\text{fumar})}$$

ó

$$p(\text{fumar} | \text{cáncer}) = \frac{p(\text{cáncer} | \text{fumar})p(\text{fumar})}{p(\text{cáncer})}$$

Se puede afirmar que $p(\text{cáncer} | \text{fumar})$ es la probabilidad inversa de $p(\text{fumar} | \text{cáncer})$. En general, estas dos probabilidades son distintas. En estas fórmulas las probabilidades $p(\text{cáncer})$ y $p(\text{fumar})$ se conocen como “probabilidades marginales” o “probabilidades previas”.

Como ejemplo, suponga que en cuatro ocasiones en las cuales los caballos A y B han corrido juntos, A ha ganado 3 veces y B una vez. Si usted fuera a apostar en la próxima carrera en que ambos caballos compitieran, lo lógico sería que usted apostara al caballo A. Suponga ahora que dicen que cuando el caballo B ganó estaba lloviendo, mientras que las tres veces en las que el caballo A ganó, llovió una sola vez y que se pronostica lluvia para la próxima carrera. ¿A qué caballo apostará ahora?

Note que:

$$p(A) = \frac{3}{4}$$
$$p(B) = \frac{1}{4}$$
$$p(\text{llover}) = 1$$

Además,

$$p(\text{llover}|A) = \frac{1}{3}$$
$$p(\text{llover}|B) = 1$$

Así que:

$$p(A|\text{llueve}) = \frac{(llueve|A)p(A)}{p(\text{llover})} = \frac{(1/3)(3/4)}{1} = 0.25$$
$$p(B|\text{llueve}) = \frac{(llueve|B)p(B)}{p(\text{llover})} = \frac{(1)(1/4)}{1} = 0.25$$

Por lo tanto, los caballos A y B tienen la misma probabilidad de ganar el día de la próxima carrera.

EL TEOREMA DE BAYES

La ecuación:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

es la expresión matemática del Teorema de Bayes. El mismo permite determinar el valor de $p(\theta|y)$ si se conoce la probabilidad “inversa” $p(y|\theta)$. En esta ecuación y representa el conjunto de datos observados, mientras que θ representa los parámetros de una distribución. La diferencia fundamental entre la inferencia frecuentista y la bayesiana tiene que ver con la naturaleza de los parámetros θ : En la primera se asume que θ es desconocida, pero tiene un valor fijo (único) que se pretende estimar. En

la segunda, θ se considera desconocida, pero es una variable aleatoria con una distribución que refleja la incertidumbre que se percibe sobre su valor verdadero.

INFERENCIA BAYESIANA

Dado que, tanto los datos y los parámetros θ se presumen aleatorios, se podría modelar la probabilidad conjunta de los parámetros y los datos como una función de la distribución condicionada de los datos, dados los parámetros:

$$p(\theta|y) \propto L(y|\theta)p(\theta)$$

De acuerdo con esta ecuación, la incertidumbre percibida en torno de los parámetros del modelo, según expresada mediante la distribución previa, está siendo pesada por los datos actuales $p(y|\theta)$ (o $L(y|\theta)$), dando lugar a un estimado actualizado de la incertidumbre percibida, según expresada por la distribución posterior $p(\theta|y)$.

La función $L(y|\theta)$, se conoce como la función de verosimilitud (*likelihood function*) y se utiliza para resumir los datos. El principio de verosimilitud afirma que toda la información en una muestra está contenida en la función de verosimilitud, y que esta función está indizada por los parámetros del modelo. Note que no hay que hacer alusión a repetición alguna de experimentos: todo lo que importa es la muestra que está a la mano.

Observe que en la ecuación anterior la función de verosimilitud está siendo actualizada por la distribución previa, es decir, por la información que se tiene de antemano. La razón (filosófica) principal por lo que la estadística bayesiana es atractiva es que, precisamente, se tiende a pensar que el progreso en las ciencias se obtiene cuando se incorpora a los estudios actuales el conocimiento obtenido en estudios previos.

En el caso en que haya información previa, para la distribución previa se selecciona una distribución que sea conjugada a la función de verosimilitud (véase la **Tabla 2**). Una distribución conjugada es una distribución que, al combinarlas con la función de verosimilitud resulten estar en la misma familia que la distribución previa. Por ejemplo, la

distribución conjugada de una distribución normal es también, bajo ciertas condiciones, una distribución normal.

Tabla 2

Distribuciones de probabilidad y sus funciones conjugadas.

distribución de los datos	distribución previa conjugada
normal	normal; uniforme
Poisson	Gama
Binomial	Beta
Multinomial	Dirichlet

Por otro lado, cuando no hay información previa, los métodos de análisis bayesianos se pueden llevar a cabo haciendo uso de una distribución no informativa, tal como la distribución uniforme $U(a,b)$. Esta distribución esencialmente indica que se entiende que el valor del parámetro de interés se encuentra en el intervalo entre a y b , y que todos los valores dentro de ese intervalo tienen la misma probabilidad.

ESTIMACIÓN DE PARÁMETROS

En estadística, se interesa obtener estimados de los parámetros de un modelo, tales como medias, desviaciones estándar o coeficientes de regresión. Para obtener algún estimado, tal como la media, en la estadística frecuentista se construye el valor esperado del parámetro mediante una ecuación de la forma:

$$E[g(y)] = \int_{-\infty}^{+\infty} g(y)p(y)dy$$

Por ejemplo, para obtener estimados de la media y la varianza, esta ecuación tendrá la forma

$$E[y] = \int_{-\infty}^{+\infty} yp(y)dy$$

y

$$var(y) = E[y - E[y]] = \int_{-\infty}^{+\infty} (y - E[y])^2 p(y) dy$$

En la estadística bayesiana, el valor esperado toma la forma

$$E(g(\theta)|y) = \int_{-\infty}^{+\infty} g(\theta)p(\theta|y)dy$$

Así que los estimados de la media y la desviación estándar estarían dados por

$$E(\theta|y) = \int_{-\infty}^{+\infty} \theta p(\theta|y) d\theta$$

y

$$var(\theta|y) = \int_{-\infty}^{+\infty} (\theta - E(\theta|y))^2 p(\theta|y) d\theta$$

Como se puede apreciar, la diferencia entre los dos métodos estriba en que en la estadística bayesiana se hace uso de la distribución condicionada $p(\theta|y)$ en vez de $p(y)$.

Además de las estimaciones puntuales mencionadas arriba, a menudo es de interés construir un intervalo de confianza alrededor de un parámetro. Los intervalos de confianza en la estadística bayesiana son similares a los intervalos de confianza en la estadística frecuentista, pero difieren en cuanto a que en la estadística bayesiana se presume que los parámetros tienen una distribución de probabilidad (son variables aleatorias). Otra diferencia importante es en su interpretación. La interpretación frecuentista del intervalo de confianza al 95% es:

Si este experimento se repite un número grande de veces y se construye el intervalo de confianza, en el 95% de las ocasiones estos contendrán el valor real del parámetro q .

Mientras que la interpretación bayesiana es una lógica y simple:

Hay un 95% de probabilidad de que el intervalo contenga el valor real del parámetro.

SELECCIÓN DE MODELOS

En las situaciones en las que se desee modelar algún proceso mediante un análisis de regresión, a menudo el investigador encuentra varios posibles modelos, por lo que requiere de algún criterio para escoger el “mejor” de ellos. Entre las opciones disponibles en la estadística frecuentista se encuentran R^2 , R^2 ajustado y C_p de Mallows, entre otros. Es posible, además, llevar a cabo una prueba de hipótesis, la prueba de razón entre verosimilitudes (*Likelihood Ratio*) para determinar si dos modelos anidados difieren significativamente entre sí.

En la estadística bayesiana existe un método análogo que hace uso de las razones entre las verosimilitudes, llamado el Factor de Bayes. Este provee una forma de cuantificar los *odds* de que los datos favorezcan un modelo sobre otro. Una ventaja de este método es que los modelos que se comparan no tienen que estar anidados. El Factor de Bayes se define por

$$\frac{p(M_1|y)}{p(M_2|y)} = \frac{p(y|M_1)p(M_1)}{p(y|M_2)p(M_2)}$$

La probabilidad $p(M|y)$ es la probabilidad de que el modelo M sea correcto, dados los datos y . De esta forma, el factor de Bayes es la razón entre la probabilidad de que el modelo M_1 sea favorecido por los datos entre la probabilidad de que el modelo M_2 sea favorecido por los datos. Este método no es una prueba de hipótesis, por lo que su interpretación es un tanto subjetiva. La **Tabla 3** muestra los criterios típicamente utilizados para interpretar el Factor de Bayes. Cuando no hay información previa se puede presumir que $p(M_1) = p(M_2) = 1/2$.

Tabla 3*Interpretación del Factor de Bayes.*

Factor de Bayes	interpretación
> 100	Evidencia extrema a favor de M_1
30 – 100	Evidencia muy fuerte a favor de M_1
10 – 30	Evidencia fuerte a favor de M_1
3 – 10	Evidencia moderada a favor de M_1
1 – 3	Evidencia anecdótica a favor de M_1
1	Falta de evidencia
1/3 – 1	Evidencia anecdótica a favor de M_2
1/3 – 1/10	Evidencia moderada a favor de M_2
1/10 – 1/30	Evidencia fuerte a favor de M_2
1/30 – 1/100	Evidencia muy fuerte a favor de M_2
< 1/100	Evidencia extrema a favor de M_2

Como alternativa al Factor de Bayes se puede utilizar el Criterio de Información Bayesiano (BIC, por sus siglas en inglés). El BIC es una aproximación al Factor de Bayes y está definido por:

$$BIC = -2 \log(\hat{\theta}|y) + p \log(n)$$

Para comparar dos modelos M_1 y M_2 , se busca la diferencia

$$\Delta BIC = BIC(M_1) - BIC(M_2)$$

El valor obtenido para ΔBIC se interpreta de acuerdo con la **Tabla 4**.

Tabla 4*Interpretación del Factor de Bayes.*

ΔBIC	evidencia contra M_2
0 – 2	pobre
2 – 6	positiva
6 – 10	fuerte
> 10	muy fuerte

EJEMPLO: ANÁLISIS DE REGRESIÓN BAYESIANO

Como ejemplo del análisis bayesiano, se presenta un análisis de regresión siguiendo el ejemplo presentado en el texto de Kaplan (2014). Para el mismo se hace uso de la base de datos de la prueba de aprovechamiento *Programme for International Student Assessment* (PISA), en su administración del año 2009. La variable independiente en este análisis son las puntuaciones en la prueba de lectura (*rcomb1*) y las variables independientes son variables de trasfondo y actitudes (*gender*, *native*, *slang*, *ESCS*, *JOYREAD*, *DIVREAD*, *MEMOR*, *ELAB* y *CSTRAT*).

Se considera el efecto de las distribuciones previas en el análisis. Primero, se tomará el caso de funciones de distribución previas no informativas para los coeficientes de regresión y, en el segundo, se hará uso de la distribución previa (informativas) normal multivariada para los coeficientes de regresión y la distribución Gama inversa para la varianza de error, utilizando los estimados obtenidos de estos parámetros con los datos de la administración del 2000. Finalmente, se compararán con el análisis de regresión utilizando la teoría frecuentista. Los análisis fueron llevados a cabo utilizando el lenguaje de programación R. El código utilizado en los análisis se presenta en el Apéndice de este escrito.

Los resultados de los análisis haciendo uso de previas no informativas e informativas se presentan en las **tablas 5 y 6**. Como se puede apreciar, los intervalos de confianza para la variable *gender* y *native* incluyen el cero. Es decir, con excepción de *gender* y *native*, todas las variables independientes contribuyen a la predicción de las puntuaciones de lectura.

Como se puede apreciar en la **Tabla 5**, al utilizar las previas informativas, los intervalos de confianza para las variables *gender* y *native* no incluyen el cero, es decir, que estas variables ahora contribuyen a la predicción de las puntuaciones en la sección de lectura. Note además que los estimados de las desviaciones estándar para cada coeficiente son, en todos los casos, menores que cuando se utilizan las previas no informativas. Como resultado de esto, los intervalos de confianza tienden a ser más estrechos. Es decir, al hacer uso de distribuciones previas informativas se obtiene mayor precisión en los estimados de los coeficientes de regresión. Además, es posible que un predictor no

contribuya a la predicción de la variable dependiente cuando se utilicen previas no informativas y lo haga al hacer uso de previas informativas, como es el caso de las variables *gender* y *native* en este ejemplo.

Tabla 5

Resultados del análisis de regresión utilizando distribuciones previas no informativas.

parámetro	media	D.E.	intervalo de confianza al 95%	
(intercepto)	490.40	3.34	483.85	496.95
<i>gender</i>	6.44	2.31	1.92	10.96
<i>native</i>	-5.68	3.85	-13.23	1.86
<i>slang</i>	6.75	4.50	-2.07	15.57
ESCS	30.28	1.30	27.72	32.83
JOYREAD	28.90	1.26	26.43	31.36
DIVREAD	-4.33	1.19	-6.66	-2.00
MEMOR	-18.85	1.30	-21.40	-16.30
ELAB	-14.88	1.25	-17.33	-12.42
CSTRAT	27.39	1.46	24.53	30.26

Tabla 6

Resultados del análisis de regresión utilizando distribuciones previas informativas.

parámetro	media	D.E.	intervalo de confianza al 95%	
(intercepto)	487.94	2.82	482.41	493.47
<i>gender</i>	9.46	1.85	5.83	13.09
<i>native</i>	-6.87	3.28	-13.30	-0.45
<i>slang</i>	9.53	3.56	2.56	16.51
ESCS	31.20	1.04	29.17	33.23
JOYREAD	26.50	1.00	24.53	28.46
DIVREAD	-2.52	0.97	-4.42	-0.61
MEMOR	-18.77	1.09	-20.91	-16.62
ELAB	-13.63	1.06	-15.70	-11.55
CSTRAT	26.06	1.17	23.76	28.35

La **Tabla 7** presenta los resultados del análisis de regresión frecuentista. Como se puede apreciar, los resultados de este análisis son prácticamente idénticos que en el análisis bayesiano con previas no informativas. Es importante recordar que, como se indicó anteriormente, la interpretación de los intervalos de confianza es distinta en los dos métodos de análisis.

Finalmente, se pueden comparar dos modelos con previas informativas, uno que incluye todas las variables (Modelo1) y otro en el que se descartaron las variables *gender* y *native* (Modelo 2). El valor obtenido para el factor de Bayes favorece al Modelo 1 sobre el Modelo 2, con un valor de 5.72. De acuerdo con la **Tabla 8**, la evidencia en favor del Modelo 1 es moderada.

Tabla 7
Análisis de regresión.

	media	D.E.	t	Pr(> t)	intervalo de confianza al 95%	
<i>(intercept)</i>	490.40	3.35	146.35	< 2e-16	483.84	496.97
<i>gender</i>	6.44	2.32	2.77	0.0056	1.88	10.99
<i>native</i>	-5.63	3.85	-1.46	0.1441	-13.18	1.92
<i>slang</i>	6.70	4.54	1.48	0.1397	-2.19	15.59
ESCS	30.29	1.31	23.13	< 2e-16	27.72	32.86
JOYREAD	28.90	1.25	23.08	< 2e-16	26.44	31.35
DIVREAD	-4.32	1.20	-3.59	0.0003	-6.68	-1.97
MEMOR	-18.87	1.31	-14.43	< 2e-16	-21.43	-16.31
ELAB	-14.89	1.24	-11.97	< 2e-16	-17.33	-12.45
CSTRAT	27.42	1.45	18.89	< 2e-16	24.57	30.26

Tabla 8
Resultados del cómputo del Factor de Bayes.

	Modelo1	Modelo2
Modelo1	0	5.72
Modelo2	-5.72	0

DISCUSIÓN

En este escrito se introdujeron los conceptos básicos de la estadística bayesiana y se demostró cómo llevar a cabo un análisis de regresión. Como se pudo observar, el uso de información previa tiene el efecto de mejorar la precisión de los estimados de los parámetros. Además, cabe la posibilidad de encontrar predictores importantes de una variable independiente que podrían resultar que no contribuyeran a la predicción cuando se usan previas no informativas.

El análisis de regresión con previas no informativas tiene resultados muy similares a los de la estadística frecuentista. Una diferencia importante, sin embargo, es la interpretación de los intervalos de confianza para los parámetros. Mientras que en la estadística frecuentista se dice “si realizamos un número grande de repeticiones de este estudio y construimos los intervalos de confianza al 95%, el 95% de estos intervalos contendrán el valor del parámetro”, en la estadística bayesiana se diría “tenemos un 95% de certeza de que el parámetro se encuentra dentro del intervalo”. Finalmente, además del análisis de regresión, se vio cómo se hace uso del Factor de Bayes para seleccionar entre dos o más modelos de regresión.

BIBLIOGRAFÍA

- David Kaplan (2014). *Bayesian Statistics for the Social Sciences*. The Guilford Press. ISBN 978-1-4625-1651-3.
- Kruschke, J.K. (2015). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. Academic Press. ISBN 978-0-12-405888-0.

APÉNDICE

CÓDIGO DE R UTILIZADO EN LOS EJEMPLOS

I. Análisis de regresión bayesiano con previas no informativas e informativas y análisis de regresión frecuentista.

```
# Instalación de los paquetes necesarios para el análisis
install.packages("MCMCpack")
install.packages("BMA")
install.packages("coda")
require(MCMCpack)
require(coda)
require(BMA)

# Leer el archivo de datos PISA2009.csv
datafile <- read.csv(file.choose(),header=T)
PISAdata <- subset(datafile, select=c(rcombl,gender,native,slang,
    ESCS,JOYREAD,DIVREAD,MEMOR,ELAB,CSTRAT))
head(PISAdata)
nrow(PISAdata)
PISAdata <-na.omit(PISAdata)
nrow(PISAdata)

# Modelo bayesiano con previas no informativas
Prev_NI <- MCMCregress(rcombl~gender+native+ slang+ESCS+
    JOYREAD+ DIVREAD+ MEMOR+ ELAB+ CSTRAT,
    data= PISAdata,burnin=5000,mcmc=100000,thin=10,b0=0,B0=0)
summary(Prev_NI)

# Modelos bayesiano con previas informativas
Prev_inf <- MCMCregress(rcombl~gender+native+ slang + ESCS+
    JOYREAD+ DIVREAD+ MEMOR+ ELAB+ CSTRAT,data= PISAdata,
    marginal.likelihood="Chib95",mcmc=10000,b0=c(491.2,12.9,3.4,
    21.4,32.6,22.9,0.01,-18.6,-11.1, 22.8),B0=c(0.0173,0.0932,
    0.0141,0.0216,0.3354,0.3210,0.3245,0.2101,0.2111,0.1707))
summary(Prev_inf)

# Modelo de regresión frecuentista
Reg_freq <-lm(rcombl~gender+native+ slang + ESCS+
    JOYREAD+ DIVREAD+ MEMOR+ ELAB+ CSTRAT,data= PISAdata)
summary(Reg_freq)
```

II. Cálculo del Factor de Bayes.

```
# Construir los modelos que se quieran comparar
Modell1 <- MCMCregress(rcombl~gender+native+slang + ESCS+JOYREAD+
  DIVREAD+ MEMOR+ ELAB+ CSTRAT,data=datfile9,
  marginal.likelihood="Chib95",mcmc=10000,b0=c(491.2, 12.9, 3.4,
  21.4, 32.6, 22.9, 0.01, -18.6, -11.1, 22.8),
  B0=c(0.0173, 0.0932, 0.0141, 0.0216, 0.3354, 0.3210, 0.3245,
  0.2101, 0.2111, 0.1707))

Modell2 <- MCMCregress(rcombl~slang + ESCS+
  JOYREAD+ DIVREAD+ MEMOR+ ELAB+ CSTRAT,data=datfile9,
  marginal.likelihood="Chib95",mcmc=10000,b0=c(491.2, 21.4, 32.6,
  22.9, 0.01, -18.6, -11.1, 22.8), B0=c(0.0173, 0.0216, 0.3354,
  0.3210, 0.3245, 0.2101, 0.2111, 0.1707))

# Calcular el Factor de Bayes
Bayes_Factor <- BayesFactor(Modell1,Modell2)
print(Bayes_Factor)
```


Centro de Apoyo Estadístico (C.A.E)

La Pontificia Universidad Católica de Puerto Rico (PUCPR), en aras de continuar fomentando la cultura de investigación, estableció el Centro de Apoyo Estadístico (CAE). La PUCPR fue pionera entre las universidades de Puerto Rico al establecer en 2012 un proyecto que atendiera la necesidad de proveer apoyo en el área de estadísticas al estudiantado de la Institución, tanto graduado como subgraduado. La reestructuración sistemática y formal de las investigaciones ofrece mayor cobertura y una experiencia positiva para estudiantes y facultativos.

El objetivo del CAE es asegurar la calidad de la investigación cuantitativa, producto del esfuerzo de los estudiantes y de la facultad en general. De este modo, la Institución contribuye al acervo académico-investigativo de la comunidad científica en Puerto Rico, así como al análisis y a la solución de problemas que encara el país. El CAE provee servicios a todos los estudiantes doctorales y facultad de los tres recintos de la PUCPR, en especial en las disciplinas de administración de empresas, educación y ciencias de la conducta.

Los servicios que se ofrecen en el CAE incluyen asesoramiento en:

- planificación de una investigación
- construcción y validación de instrumentos
- diseño de esquemas de muestreo
- diseño de experimentos
- modelación y análisis estadísticos
- análisis estadísticos (univariados, bivariados y multivariados) e interpretación de resultados
- orientación sobre programas apropiados para efectuar el análisis de datos
- seminarios y talleres a estudiantes y facultad de la PUCPR.

El personal del CAE está compuesto por dos consultores quienes se encargan de ofrecer asesoría estadística a estudiantes graduados y a facultad de la PUCPR. Además, cuenta con un enlace que coordina las actividades de capacitación que se lleven a cabo.

ASESORES:



Dr. Julio Quintana Díaz - posee un Ph.D. en Estadísticas de University of Wales, Reino Unido y un M.S. en Matemáticas de la Universidad de Puerto Rico en Mayagüez. Ha sido presidente de 15 comités graduados en estadísticas en el RUM y asesor estadístico de un centenar de disertaciones doctorales en la PUCPR. Además, ha sido asesor estadístico de la Junta de Planificación por tres décadas y asesor estadístico de múltiples proyectos gubernamentales y privados en diversas dependencias tales como el Departamento de Agricultura, la Comisión de Derechos Civiles, el Departamento de la Vivienda, ASPIRA de Puerto Rico, la presidencia de la UPR, el Centro de Detallistas, el Departamento de la Familia, la *Small Business Administration*, y el *Forest and Wild Life Administration*, entre otros. También, ha servido como evaluador externo de los proyectos MSP de la Universidad Interamericana.



Dr. José Noel Caraballo Ríos - posee un Ph.D. en Educación de Penn State, una Maestría en Estadística de Texas A&M y una Maestría en Física de la UPR. Ha sido asesor en metodología de investigación y estadística para más de 50 estudiantes doctorales en las disciplinas de educación, trabajo social, psicología y salud pública. También ha servido como asesor en metodología de investigación y análisis estadístico para el Instituto de Psicología del recinto de Río Piedras de la UPR (antes CUSEP); el Centro de Investigaciones

Sociomédicas del recinto de Ciencias Médicas de la UPR; el Instituto de Investigaciones Interdisciplinarias del recinto de Cayey de la UPR; y el Centro de Nanotecnología, también en la UPR de Cayey. El doctor Caraballo es, además, estadístico profesional acreditado por la *American Statistical Association*.

ENLACE



Dra. Ilia Rosario Nieves - posee un doctorado en Psicología Académica Investigativa (Ph.D.) de la Universidad de Puerto Rico, recinto de Río Piedras. Desde 2005 se desempeña como profesora de estadísticas e investigación en el Colegio de Ciencias de la Conducta y Asuntos de la Comunidad de la Pontificia Universidad Católica de Puerto Rico (PUCPR), recinto de Ponce. En la actualidad dirige la Escuela Graduada de Psicología de la PUCPR, que incluye los programas graduados de Psicología Clínica e Industrial Organizacional. Ha sido reconocida en tres ocasiones por la Pontificia Universidad Católica de Puerto Rico como Investigadora del Año.

Para mayor información pueden comunicarse vía correo electrónico a: centrodeapoyoestadistico@pucpr.edu

Lecturas sobre métodos estadísticos en la investigación. *Publicación del Centro de Apoyo Estadístico de la Pontificia Universidad Católica de Puerto Rico* se compaginó en Adobe InDesign CS6, utilizando la familia tipográfica Chaparral Pro.

HIMNO DE LA PUCPR

Letra:

Carlos Méndez, Milagros Mayoral,
Carmen Acosta y Marisabel Ruiz

Música: José R. Echevarría

Alegres cantemos
a nuestra Alma Mater
este himno,
símbolo de fe.
Celosos pongamos
en alto tu nombre;
orgullosos veamos
en ti nuestro ideal.

De azul, oro y blanco
tu bandera ondeará;
miles de esperanzas
vamos por ti a realizar.

Con paso firme
siempre nos guiarás
a través de la vida
con tu ciencia y verdad.
Nunca, te habremos de olvidar
fuente de amor y de saber
Unida a Dios estarás
Siempre en nuestros corazones.

¡A luchar, Pioneros, adelante!
¡A vencer con la Universidad!

