

---

## METODOLOGÍAS CUALITATIVA Y CUANTITATIVA: DOS ENFOQUES COMPLEMENTARIOS

*Miguel Ángel Rosado Chauvet*<sup>1</sup>  
*José Régulo Morales Calderón*<sup>2</sup>

### Resumen

*Actualmente y a partir de diferentes insumos las metodologías cualitativa y cuantitativa son complementarias. Los métodos cuantitativos son más utilizados para conductas probables, con capacidad de predecir y que permiten control; los métodos cualitativos son más utilizados para conductas posibles, con capacidad de anticipar y que permiten prevenir.*

*El texto se centra en las posibilidades de la categorización hasta la medición que presentan los métodos cuantitativos y las posibilidades de observación, natural cuando ya está dada, o inducida cuando se produce a través de una entrevista, una exhibición (v.g: películas) o un desempeño de papeles.*

*Sin embargo, hay aspectos del entorno o del comportamiento que son difíciles de observar como son las ondas de radio, televisión, sonido, luz, telefonía, etc. la influencia de las galaxias sobre el sistema solar y, en especial, la Tierra. Por otro lado, la imaginación y el asombro de un niño o el amor que le tiene su madre, o el poder de la mente sobre la salud física de las personas.*

*Hay muchos aspectos de los que los seres humanos son capaces de darse cuenta, sin poder explicar qué es o por qué ocurre, así como otros aspectos en los que el número ya es “real” para ellos mismos por la adap-*

---

<sup>2</sup> Profesor investigador del Área de Economía Política. Departamento de Economía. Universidad Autónoma Metropolitana–Unidad Iztapalapa.

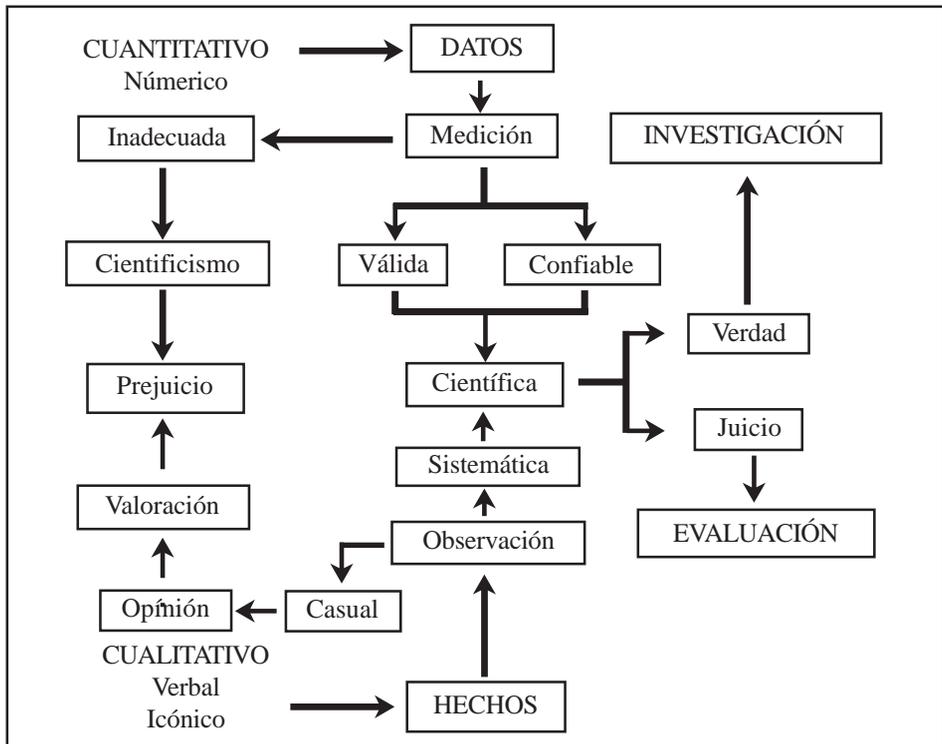
<sup>2</sup> Profesor investigador del Área de Economía Política. Departamento de Economía. Universidad Autónoma Metropolitana–Unidad Iztapalapa.

*tación perceptual y cognoscitiva a la construcción teórica en la cual se basaron inicialmente.*

*Por lo anterior, no se puede elegir una de las dos formas de investigación, sino aprender de ambas y aplicarlas de acuerdo con las características del problema a solucionar; de no ser así, en cualquier caso se limitaría en la comprensión o en la solución, por la adopción por parte del investigador de sus propias fronteras.*

Durante mucho tiempo se han visto las metodologías cualitativa y cuantitativa como mutuamente excluyentes, sin percibirse la importancia de su complementariedad. De hecho, trabajan con insumos diferentes y su uso sirve para soluciones diferentes.

FIGURA 1. COMPARACIÓN ENTRE MÉTODOS



---

Los métodos cualitativos se nutren de contenidos verbales o icónicos, entre otras formas de lenguajes, en los que se pueden realizar contrastes arquitectónicos como los presentados por las iglesias, donde la construcción sigue los principios filosóficos o costumbristas de su entorno. La corriente filosófica humanista sugiere que el valor de las personas no se encuentra en la apariencia externa (su físico), sino en lo que las personas tienen dentro (sentimientos, valores, amor, responsabilidad, etc.), y la construcción lo refleja con fachadas simples y a veces toscas, pero con una riqueza ornamental interna en altares con plata y oro, pinturas religiosas de artistas de renombre, como en las iglesias del Siglo XVI. Otro tanto ocurre con los contenidos del romanticismo como los verbales en la poesía (Gustavo Adolfo Bécquer) o acústicos en la música (Fédéric Chopin).

Al comparar aspectos de comunicación no verbal entre países o entre regiones, se puede observar que hay costumbres corporales de proximidad (países Árabes) y distancia (Anglos y Germánicos). En otros aspectos lingüísticos se encuentra el manejo de la lengua como indicador de poder (Norteamericanos que hablan en inglés por todo el mundo), o como una forma de presentar rechazo (Serrat cantando en catalán cuando en España existía la prohibición de hablar en ese idioma).

Muchos de estos indicadores proporcionan información que permite formar categorías, apreciar distancias, observar ordenamientos, afiliar a símbolos, etc., como características típicas de la investigación cualitativa. Quizá sea difícil comprobar una hipótesis, pero permite una producción inagotable de éstas, todo depende de la *flexibilidad*, la *fluencia* y la *originalidad*, características asociadas a la conducta creativa, de quien pretende investigar bajo los cánones del método cualitativo.

Sin embargo, hay que hacer notar que, mientras más controles tenga el investigador, será más factible que éste pueda llegar a conclusiones explicativas más reales y completas. Una de las principales críticas que le atribuyen a métodos como la Investigación-Acción es la falta de control del efecto del investigador sobre el grupo, así como el efecto de *Rata de Laboratorio* en el que es claro para el sujeto que participa en una investigación. El procedimiento metodológico tiene el enfoque de *Solución de Problemas*, más próximo a la toma de decisiones que a la investigación.

Hay varias características que se deben tomar en cuenta en la metodología cualitativa: *a)* el objeto de estudio debe ser *unívoco*, debe representar lo mismo para cualquiera que lo quiera observar o verificar; *b)* la delimitación del objeto de estudio debe ser *exhaustiva*, es decir, debe incluir todo lo que integra el objeto de estudio; *c)* los indicadores de registro deben ser mutuamente *excluyentes*, o sea, si se quiere registrar algún hecho sobre el objeto de estudio no debe haber confusión en cuál de los indicadores deberá registrarse; *d)* finalmente la observación y registro del hecho debe ser *triangulada*, en otras palabras, al menos dos personas deben observar y registrar los hechos relativos al objeto de estudio para contrastar los registros y verificar si estas personas mantienen suficiente relación entre sí, con respecto a lo observado.

Desde luego, se debe planear qué y cómo realizar la observación del enfoque cualitativo. Se han desarrollado diferentes formas para registrar el objeto de estudio sobre hechos y conductas. *No basta con ser un observador casual*. En todos los casos implica la especificación operacional de las conductas a observar, así la confiabilidad y validez del estudio, depende de la cantidad y la frecuencia de conductas, la cantidad de sujetos observados y la precisión de las descripciones de las conductas, así como del acuerdo entre los observadores y la capacidad de observación lograda antes del proceso. En todos los casos de observación ante la presencia del observador, el método incluye varianza de error por intrusión, se sugiere para su control el uso de *puestos de observación* fuera del entorno del observado (v.g.: Cámara de Gesell, cámaras de videocinta ocultas, etc.) o un periodo de convivencia prolongado en forma continua y estable, previo al inicio de las observaciones, hasta llegar a ser percibido por el grupo, como parte integrante de ellos.

Entre los métodos de registro observacional se encuentran<sup>3</sup>:

1. Registro anecdótico
2. Registro de ocurrencia continua
3. Registro de bloque continuo o intervalo
4. Registro de bloque discontinuo o muestreo temporal
5. Registro de duración continua

---

<sup>3</sup> Para la descripción y uso de los diferentes registros puede consultar "Metodología de investigación y evaluación" de M. Rosado, pp. 191-194.

---

6. Registro *flash* muestra
7. Registro de actividad planeada (*Pla-check*)
8. Registro *Pla-check flash*
9. Registro *Pla-check* bloque
10. Registro *Pla-check* sucesivo
11. Registro espacial
12. Registro inespecífico
13. Registro S-R o S-R-S
14. Escala S-R-S

Por lo que respecta a los métodos cuantitativos que utilizan datos como principal insumo, para obtener datos se requieren instrumentos que permitan tener: *a*) confiabilidad, y *b*) validez, como elementos necesarios, aunque no suficientes.

La validez, es la característica de un instrumento que permite medir lo que se pretende medir. De ninguna manera se trataría de medir la velocidad con un reloj, aunque un referente pueda ser *kilómetros por hora*. Si se quiere medir lo que se pretende medir, lo primero que se debe hacer es definir muy bien el objeto a medir, por ejemplo los contenidos de una materia de estudio (Unidad de Enseñanza–Aprendizaje), donde un temario indica el total del contenido que se deberá abarcar, ni más ni menos, a pesar de la tan desprestigiada y poco comprendida *libertad de cátedra*.

El temario es algo impuesto, sin personalidad real. El contenido corresponde a un acuerdo temporal con una finalidad específica, que deberá ajustarse periódicamente a las condiciones actuales. En este momento (la era de la fisión nuclear) nadie discutiría si la Tierra es plana o es redonda, o si el átomo es indivisible ( $\alpha = \text{sin}$ ,  $\tau\omicron\mu\epsilon = \text{división}$ ). Los nuevos conocimientos requieren ajustes en los contenidos temáticos de la instrucción, así como la nueva tecnología requiere ajuste de habilidades, como el manejo habitual de los aparatos electrónicos actuales, incluyendo la audición circundante, la visión tridimensional a partir de dos dimensiones, los mundos virtuales, la comunicación por ordenadores, etc. en una velocidad de obsolescencia cada vez mayor. Esto hace que el *constructo* pierda vigencia y deba ser modificado, pero entre tanto ocurre, existe una definición temática que corresponde a las características de ser *exhaustiva* y *unívoca* de los métodos cualitativos, es decir, debe incluir todo lo

que integra el objeto de estudio y representar lo mismo para cualquiera que lo quiera observar o verificar.

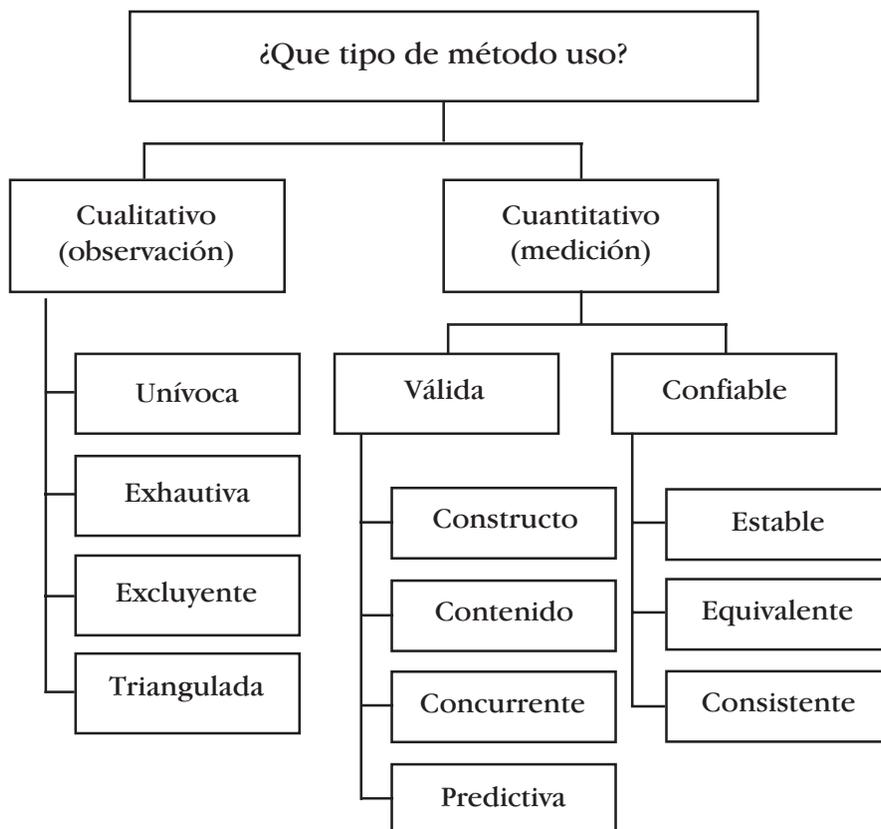
Sin embargo, es difícil contar con una medición completa del *constructo* y, aunque esto fuera posible, se tendría una alta redundancia con aspectos muy similares entre sí. Uno de los principios en la medición en las ciencias sociales consiste en tener suficientes estímulos que tengan una alta correlación con el *constructo* y una baja correlación entre ellos, para evitar redundancia en la medición, lo cual tendría similitud en los métodos cualitativos con la característica de que los indicadores de registro deben ser *mutuamente excluyentes*. En esta condición, si los ítemes de un instrumento son suficientes y representativos de un *constructo* se estará haciendo referencia al *contenido*, como segunda característica de la validez.

Tanto la validez de *construcción*, como la de *contenido*, tienen un enfoque teórico muy vinculado con la metodología cualitativa. Pero los instrumentos también deben tener una validez empírica, es decir, una utilidad práctica. Si la medición ha de servir para algo el instrumento debe ser capaz de permitir un diagnóstico y/o un pronóstico.

Si se va a capacitar a un grupo de trabajadores en una empresa, un instrumento debe proporcionar información diagnóstica sobre los conocimientos, aptitudes, habilidades, etc. que tiene el grupo en ese momento para no perder tiempo ni aburrir a los participantes con los contenidos que ya poseen. Puede ser una prueba de “papel y lápiz” objetiva y con validez teórica, o una prueba de ejecución con procedimientos estandarizados que puedan registrarse en una escala de verificación conductual. De esta manera el diagnóstico permitirá conocer el nivel de inicio y se podrá decir que el instrumento tiene validez *concurrente* con alguna condición actual que poseen los participantes.

También puede servir para predecir la productividad que tendrá un trabajador que se encuentra en un proceso de selección, o la adaptación posible de la persona a las funciones y la seguridad personal si las tareas implican riesgos. Si el instrumento permite predecir con suficiente precisión, se dice que tiene validez *predictiva*, contrastada con un criterio externo al instrumento y a futuro.

FIGURA 2. CARACTERÍSTICAS DE LOS MÉTODOS PRESENTADOS



La característica complementaria para la calidad del instrumento es la confiabilidad. Ésta se refiere a la precisión con la que se puede medir lo que se pretende medir y tiene diferentes formas de verificación.

La primera forma tiene que ver con el paso del tiempo. Si el instrumento se aplica en una ocasión y, sin que medie ninguna información relacionada con los contenidos del instrumento, cuando se aplica en un tiempo posterior relativamente breve se espera que la relación entre la primera y la segunda ocasión guarde una relación *isomórfica* y con cambios no significativos. En caso de ocurrir lo anterior se dice que posee *estabilidad temporal*, como característica de confiabilidad.

Una segunda forma se relaciona con la validez de *construcción*. Si está bien definido el *constructo* y éste ha sido muestreado al azar y de manera suficiente y representativa, puede obtenerse una muestra equivalente a la primera de manera que se obtengan dos instrumentos que incluyan ítemes suficientemente equivalentes para que los instrumentos sean equivalentes también. En este caso, si se aplican ambos instrumentos a las mismas personas, independientemente de cuál se aplique primero y cuál después, la calificación obtenida en ambos instrumentos será equivalente para todos y cada uno del que respondan, por lo que se dice que tiene *equivalencia* como característica de confiabilidad.

En cualquiera de las dos primeras formas se requiere de una doble aplicación, lo cual tiene el inconveniente de la dificultad que representa la reacción de los sujetos, además de que es difícil lograr el *isomorfismo* entre aplicaciones. También, en el primer caso, se pueden estar midiendo actitudes negativas hacia una doble aplicación y, en el segundo caso, se añade la dificultad de garantizar la equivalencia perfecta entre los dos instrumentos.

En la tercera forma se busca que los ítemes midan el mismo *constructo* que el instrumento en su totalidad y su procedimiento implica una sola aplicación. Este tipo de procedimiento permite verificar si un instrumento tiene *consistencia interna*, como característica de confiabilidad. Por las dificultades mencionadas en el párrafo anterior esta tercera forma es la más aceptable, aún cuando tiene algunas limitantes. Con una sola aplicación se puede calcular la correlación entre los ítemes nones y los pares (*non-par*) o se puede correlacionar la primera mitad del instrumento contra la segunda (*odd-even*) para los mismos sujetos; en ambos casos se tendrá la limitante de equivalencia perfecta de las mitades del instrumento por lo que se acepta la *consistencia interna* como coeficiente de mayor calidad, siempre y cuando se calcule a través de lo que se ha llamado *coeficiente Alfa* de Cronbach, que tiene como antecedente el *coeficiente Inter-ítem* de Hoyt; ambos se basan en el Análisis de Varianza (ANOVA).

El caso de las dos mitades estaría relacionado con la *triangulación* del método cualitativo, pero con la consistencia interna ya no se comparan instrumentos, por lo que se sugiere el uso de la metodología de

*replicación directa*. Como ya se sugirió que hay problemas para una doble aplicación, si el instrumento tiene una *consistencia interna* aceptable se puede aplicar a dos muestras aleatorias de sujetos, con suficientes casos para garantizar equivalencia entre las dos muestras ( $n > 30$ ). Si en ambas aplicaciones el valor del coeficiente de *consistencia interna* es muy similar, se puede concluir que el instrumento tiene confiabilidad en sus resultados.

Si la *replicación directa* se hace sobre los mismos casos con dos instrumentos se tendría el mismo caso de la *equivalencia* y, si se realiza con el mismo instrumento en dos grupos de sujetos, el problema consistiría en garantizar suficiente equivalencia entre los grupos, lo que requeriría aplicarle el instrumento a un mínimo de 90 casos (de preferencia  $N = 384$ ) bajo las mismas condiciones de lugar, fecha, condiciones y horarios, y después dividir al estricto azar el total en tres grupos iguales de 30 casos (de preferencia  $n = 128$ ). Una aplicación con doble replicación permitiría una máxima seguridad en los resultados a través de la metodología de *contrabalanceo* que controle también el efecto de orden de aplicación.

|       |          | N  | n  | Prueba |   |   | n  | Prueba |   |   | n  | Prueba |   |   |
|-------|----------|----|----|--------|---|---|----|--------|---|---|----|--------|---|---|
| Grupo | A        | 30 | 10 | D      | M | F | 10 | F      | D | M | 10 | M      | F | D |
|       | M        | 30 | 10 | M      | F | D | 10 | D      | M | F | 10 | F      | D | M |
|       | B        | 30 | 10 | F      | D | M | 10 | M      | F | D | 10 | D      | M | F |
|       | $\Sigma$ | 90 | 30 |        |   |   | 30 |        |   |   | 30 |        |   |   |

Como ejemplo, se supone una prueba de 10 temas con 12 ítems por tema y se ordena la dificultad de los ítems de más fácil a más difícil en cada tema, tomando los cuatro más fáciles de cada tema para formar una prueba fácil de 40 ítems (F), los cuatro ítems de dificultad regular de cada tema para formar una prueba regular de 40 ítems (R) y los cuatro ítems más difíciles de cada tema para formar una prueba difícil de 40 ítems (D). Por otro lado se ordenan los 90 sujetos en orden descendente de aciertos totales y los se dividen en tres grupos de 30, que incluyen a los 30 de mayor cantidad de aciertos en el grupo alto (A), a los de la cantidad media de aciertos en el grupo medio (M) y a los de menor cantidad de aciertos en el grupo bajo (B).

|       |   | Prueba |      |      |      |
|-------|---|--------|------|------|------|
|       |   | F      | R    | D    | Σ    |
| Grupo | A | 997    | 870  | 759  | 2626 |
|       | M | 663    | 581  | 447  | 1691 |
|       | B | 377    | 296  | 207  | 880  |
|       | Σ | 2037   | 1747 | 1413 | 5197 |
|       |   |        |      |      |      |

Se puede observar que, independientemente de la dificultad de la prueba, sí hay diferencia entre los grupos: alto (A), medio (M) y bajo (B); además se puede observar que, independientemente de los grupos, sí hay diferencia entre las pruebas: fácil (F), regular (R) y difícil (D). Algo más que podemos observar es que el grupo de cantidad media de aciertos (M) con la prueba regular (R) obtiene una puntuación de 581 y, por tratarse de 10 casos el promedio es de 58.1 que se encuentra muy cercano al 60 esperado como promedio de la prueba de 120 ítems; es decir, que la prueba de 40 ítems de dificultad media representa bien al promedio esperado de respuestas, y el grupo de cantidad media de aciertos representa también al promedio general esperado para los sustentantes.

Ahora, se presentan las preguntas ¿por qué en cada casilla hay 10 casos? ¿por qué se requieren 30 casos por grupo? ¿por qué las pruebas se integran por 4 ítems de cada tema y no sólo se ordenan los ítems de más fáciles a más difíciles y se divide entre tres grupos de ítems? ¿por qué tres grupos y tres pruebas?

1. Se sugiere un mínimo de 10 casos por casilla para poder trabajar con algunos de los modelos estadísticos, como es el caso de la regresión, el cual es de vital importancia en la predicción y prospección.
2. Se sugiere un mínimo de 30 casos aleatorios para tener una aproximación adecuada a la distribución normal.
3. Se requiere la división por tema y no por la prueba total, porque debe haber representatividad temática en cada uno de los instrumentos y equivalentes en todos y cada uno de los temas.

4. Se sugieren 3 grupos y 3 pruebas porque se ha comprobado experimentalmente que una aplicación con doble replicación es la que controla la mayor parte del error de medición debido a la falta de equivalencia, tanto de grupos como de instrumentos.

Las pruebas no pueden ser demasiado amplias, a no ser que se apliquen en sesiones diferentes, porque llega el momento en que se está incluyendo un error de medición originado por tedio, cansancio, aburrimiento, etc. En el caso de la Universidad Autónoma Metropolitana el examen de admisión era originalmente de 180 ítems, divididos en 50 de razonamiento verbal, 40 de razonamiento numérico y 90 de conocimientos específicos. En algún momento se presentó la propuesta de cambiar su estructura y se integró con 120 ítems, divididos en 60 de conocimientos específicos y 60 de razonamientos; a su vez los razonamientos se dividieron en 30 de razonamiento verbal y 30 de razonamiento numérico. Los temas fueron tres por categoría, de 20 ítems de conocimientos cada uno y de 10 ítems de razonamientos cada uno. Inicialmente, la confiabilidad del instrumento era de  $r_{tt}=0.60$  a  $r_{tt}=0.65$  y con el cambio de estructura se lograron índices de  $r_{tt}=0.90$  a  $r_{tt}=0.95$  en las diferentes pruebas y aplicaciones.

Para el caso de los sujetos siempre será mejor contar con una mayor cantidad, pero puede llegar el momento en que el costo de una investigación sea demasiado alto por la cantidad de casos. Aunque se puede trabajar con un solo caso, la generalización de los hallazgos no podrá ir más allá de ese solo caso, y sólo se sugiere este procedimiento cuando no importa la generalización sino el mayor control, y se trata de casos raros y aislados. Entre uno y otro extremo se encuentra la sugerencia del muestreo aleatorio irrestricto con poblaciones finitas e infinitas, donde se sugiere trabajar con una tolerancia de error del 5% ( $\alpha \leq .05$ ), que es el límite donde se controla mejor tanto el error tipo  $\alpha$  (rechazar la  $H_0$  siendo verdadera) como el error tipo  $\beta$  (aceptar la  $H_0$  siendo falsa). Las fórmulas para poblaciones finitas ( $N < 3000$ ) y para poblaciones infinitas ( $N \geq 3000$ ) son:

$$N_f = \frac{NZ^2pq}{Ne^2 + Z^2pq} \quad N_\infty = \frac{Z^2pq}{e^2}$$

donde:

- N = Número de casos de la población.
- z = Valor probabilístico de la distribución normal. Generalmente se ha trabajado con un error a de .05 siendo 1.96 su valor en z.
- p = Valor que representa la ocurrencia de cualquier característica en la población. Si se ignora la proporción, o no se desea calcularla, puede trabajarse con un valor de 0.5 para que proporcione la mayor muestra posible y cubra todos los requisitos de representatividad.
- q = Valor que representa la no ocurrencia de cualquier característica en la población. Si se ignora la proporción, o no se desea calcularla, puede trabajarse con un valor de 0.5 para que proporcione la mayor muestra posible y cubra todos los requisitos de representatividad.
- e = Grado en el que se permite la oscilación (error de medida permitido) de los valores encontrados. Generalmente se ha trabajado con un 5% de oscilación.

| N   | N'.05 | %   | N'.01 | %   | .01/.05 |
|-----|-------|-----|-------|-----|---------|
| 1   | 1     | 100 | 1     | 100 | 100     |
| 14  | 14    | 100 | 14    | 100 | 100     |
| 18  | 17    | 94  | 18    | 100 | 106     |
| 19  | 18    | 95  | 18    | 95  | 100     |
| 30  | 28    | 93  | 29    | 97  | 104     |
| 40  | 36    | 90  | 38    | 95  | 106     |
| 50  | 44    | 88  | 46    | 92  | 105     |
| 75  | 63    | 84  | 67    | 89  | 106     |
| 100 | 79    | 79  | 87    | 87  | 110     |
| 200 | 132   | 66  | 154   | 77  | 117     |

| N    | N'.05 | %  | N'.01 | %  | .01/.05 |
|------|-------|----|-------|----|---------|
| 300  | 168   | 56 | 207   | 69 | 123     |
| 400  | 196   | 49 | 250   | 63 | 128     |
| 500  | 217   | 43 | 285   | 57 | 131     |
| 750  | 254   | 34 | 352   | 47 | 139     |
| 1000 | 278   | 28 | 399   | 40 | 144     |
| 1500 | 306   | 20 | 460   | 31 | 150     |
| 2000 | 322   | 16 | 498   | 25 | 155     |
| 2500 | 333   | 13 | 524   | 21 | 157     |
| 3000 | 341   | 11 | 543   | 18 | 159     |
| ∞    | 384   |    | 664   |    | 173     |

Hasta 14 casos la muestra es igual al total de casos y es igual para  $\alpha = .05$  y  $\alpha = .01$ ; con 100 casos se deben tomar 79 casos para  $\alpha = .05$  y 87 casos para  $\alpha = .01$  que representa un 10% más para ganar un 4% de precisión, con una población al infinito se toman 384 casos para  $\alpha = .05$  y 664 casos para  $\alpha = .01$  que representa aumentar la muestra en un 73%

para ganar un 4% de precisión. Esto también se debe tomar en cuenta si se desea aumentar la muestra sin una razón de más peso que incrementar la precisión para la hipótesis.

Por último, es necesario hacer notar que, si bien había existido una dificultad para probar hipótesis a través de métodos cualitativos, en muchos casos ha sido posible describir sucesos, hechos, objetos, etc. mediante la utilización de frecuencias y porcentajes, dando paso a la estadística *descriptiva* con varios de los modelos actuales que permiten reducir la distancia entre lo cualitativo y lo cuantitativo, algunas de cuyas similitudes se han expuesto hasta este momento, sin que se pretenda agotarlas, sino más bien concordarlas en lo posible.

## Bibliografía

- Adkins, D. (1983). *Elaboración de tests*, Trillas, México.
- Bakeman, R. & Gottman, J. M. (1989). *Observación de la interacción: introducción al análisis secuencial*, Morata, Madrid.
- Bastin, G. (1965). *Los tests sociométricos*, Kapeluz, Buenos Aires.
- Best, J. W. (1982). *Cómo investigar en educación*, Morata, Madrid.
- Bloom, B. et al. (1990). *Taxonomía de los objetivos de la educación: Clasificación de las metas educacionales*, El Ateneo, Buenos Aires.
- Campbell, D. T. & Stanley, J. (1973). *Diseños experimentales y cuasiexperimentales en la investigación social*, Amorrortu, Buenos Aires.
- Castro, L. (1978). *Diseño experimental sin estadística*, Trillas, México.
- Cochran, W. G. & Cox, G. M. (1976). *Diseños experimentales*, Trillas, México.
- C.E.E.B. [Puerto Rico] (1979). *El análisis de items*, Materiales del Seminario sobre los fundamentos y prácticas de las admisiones universitarias. Hato Rey: The College Board.
- Cook, T. D. & Reichardt, Ch. S. (1997). *Métodos cualitativos y cuantitativos en investigación evaluativa*, Morata, Madrid.
- Cortada de K., N. (1968). *Manual para la construcción de tests objetivos de rendimiento*, Paidós, Buenos Aires.
- Cortés del Moral, R. (1985). *El método dialéctico*. Trillas, México.
- De Schutter, A. (1981). *Investigación participativa*, CREFAL, México.

- Dieppa, J. J. (1979). *La evaluación de la conducta en el área psicomotor*, Materiales del Seminario sobre fundamentos y prácticas de las admisiones universitarias. Hato Rey: The College Board.
- Edwards, A. L. (1957). *Techniques of attitude scale construction*, Appleton-Century-Crofts, New York.
- Edwards, A. L. (1968). *Experimental design in psychological research*, Holt, Rinehart & Winston, New York.
- Elliot, J. (1994). *La investigación-acción en educación*, Morata, Madrid.
- Ezekiel, M. & Fox, K. A. (1959). *Methods of correlation and regression analysis*, Wiley, New York.
- Ferguson, G. A. (1959). *Statistical analysis in psychology and education*, McGraw-Hill, New York.
- Gagné, R. M. (1973). "Domains of learning", *Interchange*. 3: 3.
- Gronlund, N. E. (1982). *Elaboración de tests de aprovechamiento*, Trillas, México.
- Kerlinger, F. N. (1975). *Investigación del comportamiento. Técnicas y metodología*, Interamericana, México.
- Krathwohl, D. R., et al. (1964). Taxonomy of educational objectives, Handbook II: Affective domain. David McKay, New York.
- Kuder, G. F. & Richardson, M. W. (1937). "The theory of the estimation of test reliability", *Psychometrika*, 2, 151-160.
- Lawshe, C. H. (1953). *Pruebas psicotécnicas en la selección de personal*, Rialp, Madrid.
- Likert, R. A. (1932). "A technique for the measurement of attitudes", *Archives of Psychology*, 140, 44-53.

- Loevinger, J. (1947). "A systematic approach to the construction and evaluation of tests of ability", *Psychol. Monogr.*, 61, núm 4.
- Magnusson, D. (1975). *Teoría de los tests*, Trillas, México.
- Marín, G. (1977). *Manual de investigación en psicología social*, Trillas, México.
- Martínez M., M. (2004). *Ciencia y arte en la metodología cualitativa*, Trillas, México.
- Meehan, E. (1968). *Explanation in social sciences: A system paradigm*, Dorsey Press, Homewood.
- Muñiz F., J. (1990). *Teoría de respuesta a los ítems. Un nuevo enfoque en la evolución psicológica y educativa*, Pirámide, Madrid.
- Osgood, C. E. & Tannenbaum, P. H. (1955). "The principle of congruity in the prediction of attitude change", *Psychological Review*, 62, 42-55.
- Poch, A. (1976). *Curso de muestreo y aplicaciones*, Aguilar, Madrid.
- Rheault, J. P. (1979). *Teoría de las decisiones*, Limusa, México.
- Rojas S., R. (1991). *Guía para realizar investigaciones sociales*, Plaza y Valdez, México.
- Rosado, M. A. (1973). *Prueba de aptitudes múltiples para obreros de máquinas-herramienta y ayudantes de las mismas*, Tesis presentada para obtener el grado de licenciatura. Facultad de Psicología. UNAM, México.
- Rosado, M. A. (1996). "Cinco estudios de estadística para la evaluación", *Cuadernos del CENEVAL*, Junio.
- Rosado, M. A. (1998). *Hacia un modelo de evaluación de la docencia en la educación superior*, Tesis presentada para obtener el grado de Maestría. Facultad de Psicología. UNAM, México.

- Rosado, M. A. (1999). *Análisis de ítemes. Teoría Clásica y Teoría de Respuesta al Ítem. Comparación y aplicaciones de tres modelos*, Tesis presentada para obtener el grado de Doctorado. Departamento de posgrado en Educación. Universidad Autónoma de Tlaxcala, México.
- Rosado, M. A. (2003). *Metodología de investigación y evaluación*, Trillas, México.
- Rosenblueth, A. (1993). *El método científico*, La Prensa Médica Mexicana, México.
- Scriven, M. (1976). *The methodology of evaluation*, En R. E. Stake [dir.] Curriculum evaluation. Monograph series on evaluation 1, American Educational Research Association. Rand McNally, Chicago.
- Sidman, M. (1978). *Tácticas de investigación científica*, Fontanella, Barcelona.
- Siegel, S. & Castellan, N. J. (1995). *Estadística no paramétrica. Aplicada a las ciencias de la conducta*, Trillas, México.
- Summer, B. & Summer, R. (2001). *La investigación del comportamiento. Una guía práctica con técnicas y herramientas*, Oxford, México.
- Summers, G. (1976). *Medición de actitudes*, Trillas, México.
- Thurstone, L. L. (1925). "A method of scaling psychological and educational tests", *J. Educ. Psychol*, núm. 16, págs. 433-451.
- Tuckman, B. W. (1979). *Evaluating instructional programs*, Allyn & Bacon, Boston.
- Tuckman, B. W. (1978). *Conducting educational research*, Harcourt Brace Javanovich, New York.
- Tuckman, B. W. (1975b) *Measuring educational outcomes (Fundamentals of testing)*, Harcourt Brace Javanovich, New York.

- Weiss, C. H. (1980). *Investigación evaluativa*, Trillas, México.
- Werkmeister, W. H. (1948). *An introduction to critical thinking*, Johansen Publishing, Lincoln, Nebr.
- Whyte, W. F. & Hamilton, E. (1964). *Action research for management*, Homewood, Ill.: Dosey Press. pp. 209-221.
- Winer, B. J. (1971). *Statistical principles in experimental design*, McGraw-Hill, New York.
- Wood, G. (1981). *Fundamentals of psychology research*, Little, Brown & Co, Toronto-Boston.
- Wright, B. & Stone, M. (1998). *El diseño de mejores pruebas utilizando la técnica de Rasch*, CENEVAL, México.
- Yela, M. (1997). *La técnica del análisis factorial*, Biblioteca Nueva, Madrid.
- Zetterberg, H. (1965). *On theory and verification in sociology*, Bedminster Press, New York.