

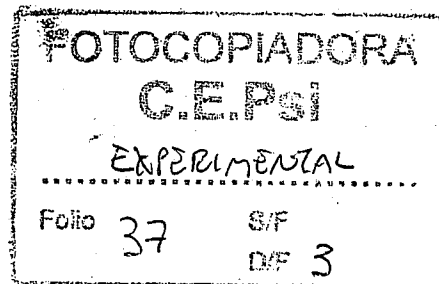
COLECCIÓN  
METODOLOGÍAS

Dirigida por  
Ana Lía Kornblit

ALBERTO CARLI

# LA CIENCIA COMO HERRAMIENTA

Guía para la investigación  
y la realización de informes,  
monografías y tesis científicas



MARÍA JOSÉ SÁNCHEZ VAZQUEZ  
Lic. y Prof. en Psicología - UNLP  
Mgter. Ética Aplicada - UBA  
M.P. 50.865

**Editorial Biblos**  
METODOLOGÍAS

rables. Una tercera manera es dividir la escala en dos y proceder a hacer las mediciones y comparar los resultados de ambas mitades.

Con la validez siempre es importante evaluar teóricamente lo que se quiere medir; otras opciones son convocar a expertos para que opinen, compararla con otras escalas y ver si con criterios diferentes varían o no los resultados.

En una escala además se debe ponderar si todos los ítems controlados tienen el mismo peso relativo dentro de la totalidad de la variable medida.

Las tres escalas más comunes son las de Lickert, la de Guttman y la de Thurstone, sobre las que no abundaremos ya que existen textos, mencionados en la bibliografía, en los que se las trata en extensión.

## El ordenamiento

Antes dijimos cuáles son las condiciones que se deben tener en cuenta en el momento en que se decida ordenar la información recogida. Cumplidas estas condiciones, se deberá proceder a categorizar los datos obtenidos. Esta tarea consiste en descomponer un conjunto o clase en subconjuntos o subclases, según lo que se haya establecido previamente, y en ordenar *los valores de las variables* de manera que cumplan con los criterios de utilidad y significación.

Decimos que son *útiles* cuando tienen potencialidades para descubrir nuevos elementos que hagan evidentes otras cualidades de las variables que pueden permanecer ocultas, en ordenamientos más inclusivos. Si en un estudio en hipertensos se encontrara un alto porcentaje de síndrome metabólico pudiera ser que estuvieran incluidos pacientes con diabetes a los que, en un primer momento, no se les dio la relevancia que luego mostraron.

Decimos que son *significativos* cuando son relevantes en el contexto en el que estamos trabajando. Si la incidencia de diabetes tipo 2 en los pacientes estudiados fuera baja, no tendría sentido generar una categoría que los contemplara y sería conveniente incluirla en una más abarcadora, por ejemplo, *síndrome metabólico*.

En el momento en que se están ordenando los datos es conveniente, sobre todo cuando se utilizan estrategias cualitativas, tener una clara definición conceptual de la variable correspondiente, como ya dijimos al referirnos al dato científico, a la vez que tener bien clara la manera en que se va a codificar la información. En razón de esto surge la necesidad de hacer un *trabajo de supervisión* en el momento de llenar las planillas porque la codificación es un trabajo rutinario y el codificador puede, insensiblemente, introducir cambios.

Es muy importante, además, que los investigadores tengan clara la diferencia entre *análisis e interpretación de los datos*. El análisis consiste en ordenar las observaciones realizadas de manera que respondan a las pregun-

tas de la investigación. La interpretación consiste en buscar significaciones a las respuestas encontradas. Ambas actividades deberán tener como condición un fuerte anclaje en la teoría.

Es interesante saber que *las investigaciones correlacionales y explicativas* y sus hipótesis correspondientes señalan con claridad de qué manera se ordenarán los datos, cómo se cruzarán las variables y qué resultados se esperan encontrar. Las *exploratorias* y las *descriptivas* exigen una mayor agudeza e imaginación para poder establecer esas relaciones.

El ordenamiento de los datos puede realizarse en *tablas* en las que las unidades de análisis están ubicadas en las líneas y las variables, en las columnas.

En una tabla se reconocerán los siguientes elementos: serán numeradas a medida que aparecen en el texto, tendrán un título en el que se consigne el tipo de datos que se presentan y sus unidades de medida, cada hilera y cada columna deben ser claramente identificadas, si los datos son de otra fuente se lo debe mencionar con una nota, las celdas que no tienen valores se señalan con un guión —el cero representa un valor y no carencia de información—, cuando se usen porcentajes se acompañarán de los valores absolutos y los totales deber ubicarse al final o al principio de las hileras y columnas, según la importancia que se les atribuya.

Debe recordarse que *una buena tabla no debe necesitar explicaciones*.

También es común el uso de *gráficos de barra* (histogramas) en los cuales se ubican en el eje vertical (Y) los valores y en el horizontal (X) las frecuencias encontradas.

Una tercera forma de uso habitual es el uso de *representaciones circulares* (las "tortas") en las cuales, para calcular la superficie correspondiente a cada categoría encontrada, se multiplica el valor de la misma (expresada en porcentaje) por 360 y se divide por 100.

## Concepto de prueba estadística: su sentido y justificación. Las principales pruebas

Cuando un investigador recoge información y la estructura como datos científicos suele entrar dentro de un campo conceptual, el del método estadístico. No negaremos que, para quienes no han hecho del tema una práctica cotidiana, presenta dificultades; no es la menor de ellas llegar a entender la relación que se establece entre magnitudes imprecisas y hasta fluctuantes, sobre todo en el mundo de las ciencias humanas, y las rígidas matemáticas.

Frente a esto, que pretende ser una suerte de introducción a un tema que asusta a más de uno, intentaremos clarificar la manera en que un investigador utilizará estas herramientas.

Según se planteó, todo investigador deberá planear un proyecto, que plasmará en un protocolo escrito en el que definirá objetivos, hipótesis, diseño,

datos y, si correspondiera —esto es, que la estrategia fuera cuantitativa—, *el tratamiento estadístico a utilizar*. Reiteramos, una vez más, la idea de lo que de embriogénica tiene la tarea. Antes de empezar a trabajar deberá realizarse una planificación lo más completa posible y esto incluye la idea de haber decidido de qué estrategias estadísticas se hará uso para darle mayor fuerza de verdad a nuestros hallazgos. Dicho de otra manera: de qué manera el investigador medirá la probabilidad de haberse equivocado, de que lo que se afirma sea falso.

Es evidente que *si el diseño es exploratorio no habrá tal tratamiento*.

Cuando el objetivo sea *describir*, esto es atribuir valores a una dimensión de la variable, el interés será conocer el valor de los atributos (variables) de las unidades de análisis del objeto de nuestro interés. En general se trabaja con conjuntos de entes de interés para el investigador. Estos entes estarán expresados por las unidades de análisis seleccionadas. Es claro que lo ideal sería conocer todas las unidades de análisis, lo que se conoce como *población*, pero diversas razones lo hacen altamente dificultoso. Factores económicos (en poblaciones muy numerosas), temporales (la necesidad de contar con resultados en un corto tiempo) o la misma naturaleza del estudio (muerte súbita) pueden hacer que se necesite trabajar con *subconjuntos*. Éstas son las *muestras*, subconjuntos representativos de la población, seleccionados según técnicas llamadas de muestreo o no, según ya se dijo.

Cuando se describe es importante considerar:

- 1) el diseño de la muestra;
- 2) la media aritmética, y
- 3) la desviación estándar.

*La muestra deberá ser representativa*: esto es, estará constituida por las unidades de análisis que caracterizan adecuadamente el problema de investigación. Evidentemente todos aquellos temas que impliquen variabilidad biológica exigirán al investigador un esfuerzo intelectual en la selección de las unidades muestrales. Pero aquí es donde vendrán en su ayuda la *media aritmética* y la *desviación estándar*. La primera será el resultado de sumar todos los valores de la variable considerada y dividirla por el número de unidades estudiadas. Aisladamente considerada, es poca la información que nos brinda porque podría ser el caso de que en una media determinada se encuentren contenidos valores extremos no identificados. Pero si a esta información se le agrega el *desvío estándar* tendremos la noción de la dispersión presente en la muestra. También tendremos que recordar que para que estemos seguros de la adecuada homogeneidad de la muestra, de que sea lo suficientemente representativa, esto es que presente lo que se conoce como *variabilidad del azar* (de la que se ocupa la estadística), la representación gráfica de los casos recolectados tendrá que tener una regularidad conocida como *curva o campana de Gauss*.

Como no escapará al lector, este abordaje exige un tipo de distribución especial, normal. Cuando esto no sea así la medida de posición a considerar será la *mediana*.

Otra información que el investigador deberá contar es el *rango*, que es la diferencia existente entre el valor mayor y el menor, que es otra manera de medir la variabilidad de la muestra estudiada.

Cuando esta regularidad esté presente se utilizarán las *técnicas paramétricas*. Cuando no estén aseguradas el investigador estará obligado a usar técnicas menos precisas, más groseras, las *no paramétricas*. Todo investigador sabe de la variabilidad que tiene lo biológico, lo humano, pero ningún hombre mide seis metros de modo que, de alguna manera, es posible llegar a conocer algún tipo de valor "normal", obviamente logrado con numerosas mediciones que permitan conocer la distribución correspondiente.

Conocer el *desvío estándar* nos servirá para conocer la forma de la curva. Cuanto más pequeña, existirá una menor dispersión de los valores logrados y cuanto mayor sea, deberá considerarse una mayor dispersión de éstos. De manera que se considera que entre dos desvíos a cada lado de la media encontraremos al 95% de los individuos de la muestra y que prácticamente todos (99%) entre los valores comprendidos entre tres desvíos a cada lado de la media aritmética. Si consideráramos un estudio de estaturas de mil individuos con una media de 1,70 metro y un desvío estándar de 0,10 metro esto nos diría, sin conocer otra información complementaria, que entre 1,60 y 1,80 metro tendríamos 950 individuos y que casi la totalidad estarían entre 1,50 y 1,90 metro. Un artificio estadístico es la *varianza*, lograda mediante la elevación al cuadrado del desvío estándar, de amplio uso en las pruebas cuantitativas.

Como vimos con estas dos medidas, *media* y *desvío*, es posible conocer acerca de la variabilidad de lo que estamos midiendo.

Pero, como ya hemos visto en otro lugar, el interés de todo investigador es conocer lo que ocurrirá en la población en general a partir de lo que ha logrado saber en una muestra. Un investigador sabe de la variabilidad de sus resultados pero también que a medida que repita sus mediciones se irá acercando a una mejor representación del fenómeno en estudio pero *siempre con un margen de error*.

Aquí es donde la estadística viene en auxilio del investigador mediante la medición del *error estándar*. Como asimismo hay una falta de precisión en los resultados obtenidos, entonces se establece cuál será la tolerancia para esa falta de precisión, mediante el establecimiento de un *intervalo de confianza*. Como todo esto no es suficiente para lograr que la ciencia obtenga información dentro de ciertos márgenes, se establece una *probabilidad* de que, a pesar de todo el esfuerzo realizado, los valores encontrados escapen de los que se hallarán en la población, esto es, que estén equivocados.

Lo que sustenta esta posición es imaginar que las medidas se toman un número infinito de veces, todo lo cual está demostrado matemáticamente por

el teorema central del límite, que muestra la dispersión hipotética de las medias en contraposición de la dispersión que muestran los valores encontrados en una muestra. Antes dijimos que la dispersión en la muestra está definida por el desvío estándar; en el caso de que la investigación se realizara en repetidas oportunidades esta dispersión estaría mostrada por el error estándar. Así, entre la media real (la de la muestra) y dos veces a cada lado el error estándar encontraríamos las sucesivas medias hipotéticas en 95 de cada 100 veces en que el trabajo fuera repetido. Con esto lo que se intenta es comprender que *la repetición infinita de la medida de muestras es llegar a medir la población completa*.

Así entendemos cómo a partir de un valor medio obtenido en una muestra real si le sumamos y restamos dos veces el error estándar obtendremos el intervalo de confianza de la población. Como sabemos que en 5 de cada 100 ocasiones podría ser que caiga fuera de ese intervalo, entonces la probabilidad de equivocarnos será del 5%, lo que llamamos  $p < 0,05$ . Si en lugar de dos veces tomamos 2,6 veces de cada lado el error estándar entonces el error sería del 1%, esto es,  $p < 0,01$ . Esta misma forma de razonamiento se utilizará en investigaciones cualitativas: en caso de que los resultados fueran expresados como porcentajes lo que se deberá calcular será el *error estándar del porcentaje*.

Como hasta aquí se debe haber comprendido que el gran problema de todo investigador es la imprecisión de sus resultados en cada una de las pruebas o mediciones que haya realizado. Este error se reducirá a medida que se aumente el tamaño de la muestra. Y aquí es bueno recordar un axioma estadístico: *la doble precisión en la población se alcanza cuadruplicando el tamaño de la muestra*.

Otra forma de abordaje estadístico se planteará en aquella circunstancia en que se trate de *comparar la misma variable en distintas muestras*. Bueno será recordar lo antes afirmado: el investigador estará limitado a estudiar muestras reducidas. El problema de evaluar las diferencias existentes entre una mejoría de 50 y 55% con diferentes medicamentos deberá tener una fundamentación que vaya más allá de lo que los números absolutos sugieran. Algo que permita inferir que estos resultados se repetirán en otra experiencia garantizando que mostrarán respuestas diferentes en las poblaciones de las que esas muestras proceden. Se suele considerar que *las respuestas diferentes son estadísticamente significativas* y que, conocida como la medida de la *probabilidad de error*, ésta es aceptada como menor a 5% ( $p < 0,05$ ).

Para que esa diferencia sea aceptada se deberá estar seguro de que las muestras son iguales en todas las condiciones menos en aquella que se está midiendo, investigando. Se dice que *las muestras son homogéneas*. Pero, como ya se mencionó, la ciencia necesita no ser subjetiva, por lo que utiliza el azar, la aleatoriedad, para armar sus muestras, con lo que intenta reproducir el azar presente en la vida misma. Así se tratará de encontrar que las

diferencias halladas no pueden ser atribuidas a la variabilidad que habitualmente está presente en las poblaciones de las que provienen las muestras. No obstante lo afirmado, conviene recordar que una muestra aleatoria no siempre es garantía de rigor científico y puede también presentar *sesgos producidos por el azar*, por lo que el investigador deberá en todo momento mantener lo que se denomina *vigilancia epistemológica*, esto es, que sus resultados no se contradigan con la teoría que los sustenta, en cuyo caso deberá realizar una prolija revisión de sus supuestos. Un momento adecuado y a tiempo para detectar estos sesgos es aquel en el que se realiza la prueba piloto.

Una vez que se asegura la homogeneidad de las muestras el investigador, que tiene alguna hipótesis de trabajo, las negará (*hipótesis nula*), y dirá que *las muestras que ha estudiado no se comportan de manera diferente entre sí*, que lo que encuentra es atribuible al azar. Decir que *no son diferentes no es lo mismo que decir que son iguales*, y aquí deberemos recordar la variabilidad que la vida tiene. La igualdad no se puede probar. Si la experiencia se repitiera muchas veces, existiría una diferencia entre cada uno de los resultados: el error estándar, que también tendría una distribución gaussiana, normal. Por ello volveremos a recordar el beneficio de repetir mediciones (esto es, aumentar el tamaño de la muestra) para disminuir el error. Una prueba de uso habitual, destinada a conocer si las *medias aritméticas obtenidas en dos muestras difieren* significativamente, es la *prueba "t" de student*. Esta prueba será diferente según se trate de *muestras apareadas* (sujetos sometidos a un tratamiento, a los que se les realizó mediciones antes y después) o para *muestras independientes* (unos sujetos sometidos a tratamiento y otros con placebo).

En el caso de que las *muestras a comparar no tengan distribución normal* se recurrirá, como se dijo, a *tests no paramétricos*, en los que *se comparan medianas*. El utilizado en este caso es el test de Wilcoxon para muestras apareadas. En el caso de muestras independientes el test utilizado es el de Mann-Whitney.

Otro problema se presenta *cuando se trata de estudiar la relación entre variables*. En otro lado se mencionó que este tipo de relación podía ser del tipo "A provoca B" (explicativo) o del tipo "A en tanto B" (correlacional). Al margen de que sea de una u otra la forma en que el investigador piense esa relación, lo importante es recordar que será diferente el tratamiento según se traten de variables cualitativas o cuantitativas. En el primero, por ejemplo presencia de dislipemia e infarto de miocardio (variables categóricas), la prueba utilizada (*no paramétrica*) es la denominada CHI cuadrado (se pronuncia *ji*), en la cual lo que se trata es ver si la asociación que se piensa es significativa, en cuyo caso los no dislipémicos deberían tener una incidencia menor de infarto de miocardio.

En el caso de que la investigación se maneje con *variables cuantitativas* el indicador matemático de que la relación entre las variables está presente

es el *coeficiente de correlación* a partir del cual se obtienen fórmulas con las cuales lograr las *ecuaciones de regresión* que permitan estudiar el comportamiento y las relaciones entre variables.

En el primero de los casos se utiliza el *coeficiente de correlación de Pearson* y el segundo consiste en un artificio matemático mediante el cual se da a conocer la influencia de una variable (independiente) sobre otra (dependiente). Cuando esta relación se grafica en un eje de coordenadas se ubica, por acuerdo, en el eje horizontal la variable independiente ( $x$ ) y en el vertical la dependiente ( $y$ ). Sería redundante repetir las consideraciones que ya hemos realizado acerca de las fuertes connotaciones teóricas exigibles para asumir las condiciones de estas variables como respectivamente independiente y dependiente, y los errores resultantes de una inadecuada caracterización.

En el caso en que de lo que se trate es de valorar estadísticamente la *influencia existente entre dos o más variables categóricas* (de clase o categoría: sexo, profesión, etc.) que consideramos independientes sobre una que consideramos dependiente (medida en niveles de razón: peso, estatura; o de intervalos: mayor, menor, etc.) se utilizará el *análisis factorial de varianza* (ANOVA).

Cuando se desee evaluar el efecto que tienen, sobre una variable dependiente, variables independientes conformadas por continuas y categóricas y se quiera "limpiar" la influencia de estas últimas, se utilizará el *análisis de covarianza*.

## El análisis y la interpretación de los datos

En aquellas circunstancias en las que se trate de estudiar una sola variable el investigador se limitará a describir. Pero el análisis podrá ser *divariado*; en este caso se trabajará con dos variables simétricas, esto es, ninguna de las dos está en función de la otra. En aquella circunstancia en la que no podamos establecer una relación temporal (antes-después) entre variables, también se tratará de una relación *descriptiva*, al no ser posible establecer una asociación secuencial.

Cuando se piense que existe una *asociación secuencial* será necesario partir del supuesto de que existe una secuencia temporal. Los métodos para establecer relaciones entre variables que cumplen con este requisito pueden ser:

- 1) El *método de la concordancia* consiste en verificar si dos o más fenómenos tienen una circunstancia en común, que sería la causa.
- 2) El *método de la variación concomitante* propone que un fenómeno varía de cualquier manera siempre que varíe otro fenómeno, porque está conectado con este último por algún nexo de causalidad.

Se deberá recordar que no existen métodos que garanticen la condición de

verdad. Todo deberá remitirse al plan lógico de estudio y la capacidad del investigador.

Puede ser que con un adecuado cruce de variables se obtengan resultados satisfactorios pero, en caso de tratar de alcanzar algún tipo de explicación, convendrá recurrir al análisis de otras variables.

En investigaciones de mayor nivel de complejidad se realizarán *análisis multivariados*, pero también este tipo de análisis se decide cuando se intenta un mayor acercamiento a la realidad.

Cuando dos variables se relacionan de alguna forma es preciso que el investigador intente conocer en qué consiste esa relación. Para ello existen procedimientos, uno de los cuales es el de la *covariación*: introducir, previa a la variable independiente, una tercera, el *factor de prueba* (del que se sepa que está relacionado con ambas, independiente y dependiente) y si se constata que con su presencia la relación desaparece se dice que la relación es parcial. Si la relación se mantiene, serviría para confirmar la relación existente. De todas maneras se recordará que podría darse el caso de que existieran variables a las que no se tomó en consideración y que modifiquen, de alguna manera, la relación entre ellas. Se deberá recordar que la utilización del factor de prueba es conveniente para poner de manifiesto la existencia de variables no consideradas.

Cuando la relación se confirma se puede plantear una pregunta adicional acerca de cómo se interpreta la relación entre la variable independiente y la dependiente. Para ello se utiliza otra variable denominada *interviniente*, posterior a la independiente y previa a la dependiente, correlacionada con ambas, y constatar lo que ocurre en su presencia.

Para probar si esa relación se mantiene en otras circunstancias se puede hacer una prueba que da mayor *especificación* y que consiste en examinarla en otro grupo en condiciones diferentes, observar si subsiste o no y en qué grado se mantiene. Aquí no se toma en cuenta la relación temporal de las variables usadas como factor de prueba. Se trata de incluir en el diseño preguntas que se refieran a fenómenos relacionados y, mediante tablas en las que se cruzan los datos, establecer la frecuencia con que los individuos clasificados de determinada manera, en un caso, aparecen en la nueva variable en juego. Como quiera que sea, siempre se trata de realizar cruces que estén dentro de la lógica de un teoría determinada.

En el caso particular de las *técnicas cualitativas* podría ocurrir que el investigador se encuentre con casos que se salen de lo esperado, con categorías no pensadas; en esa circunstancia no debe apartarlos sin tratar de entender su significado, pues es posible que presenten otras variables no consideradas relevantes y será el momento de buscar otras relaciones, otros supuestos teóricos. Una vez realizado el análisis y la categorización correspondientes se procederá de la misma manera que con los datos estructurados. Por supuesto que las respuestas que se recogen en un cuestionario abierto contendrán datos que no responden a los objetivos de la investigación, por lo cual se los

dejará de lado aunque se los puede mostrar de manera ilustrativa para que quede claro cuáles son los temas de interés concreto y cuáles no.

Una vez ordenados los datos es importante avanzar y tratar de interpretarlos, lo cual consiste en "leerlos" a la luz de la teoría y extraer conclusiones que serán la base para la elaboración de nuevas teorías o para refutar o reforzar las existentes. Se tratará de establecer alguna inferencia, tal como vimos en el capítulo 3, y, mediante una adecuada articulación lógica, generar una estructura que satisfaga las preguntas de investigación originales. Debe recordarse que *una investigación intenta responder sólo a las preguntas-problemas que la originaron.*

#### PROYECTO DE INVESTIGACIÓN

*En el proyecto de investigación que usted está desarrollando es el momento de planear el ordenamiento y el análisis a que someterá los datos planeados para su futura investigación.*

La obtención de la *muestra* deberá tener como objetivo a cumplir ser lo más semejante a la población a la que pertenece. A tal fin se deberán explicitar las técnicas de muestreo, con la utilización (*probabilísticas o aleatorias*) o no (*no probabilísticas*) del azar.

Se expresarán los criterios usados para estimar el *tamaño de la muestra*, y los niveles que se aceptarán para el *error de tipo I* (aceptar una hipótesis falsa y rechazar una hipótesis nula, esto es, aceptar que hay diferencias cuando en realidad no las hay, un *falso positivo*) y el *error de tipo II* en el que no hay pruebas suficientes para rechazar la hipótesis nula (recordar que esto no significa que no haya una verdadera diferencia), un *falso negativo*. Para el primero, habitualmente, se acepta un nivel del 0,05 y para el segundo entre 0,10 y 0,20. Recuerde que el *error muestral* disminuye con el aumento del tamaño, pero éste no elimina los sesgos.

#### 5. Materiales y métodos

Este punto también podría incluir lo descrito en el punto 3. No obstante, se ha preferido presentarlos separados, en razón de la importancia teórica que tiene, a los fines de reforzar la validación científica del proyecto a realizar. Una investigación deberá informar adecuadamente de la manera en que ha sido realizada, de modo que se faciliten posibles discusiones en el seno de la comunidad científica de pertenencia. Así es como los autores se preocuparán de clarificar acerca de:

**5.1. Variables relevantes.** Son aquellas características del objeto de estudio y de las unidades de análisis que, en razón de los objetivos de la investigación y del marco teórico de referencia, los investigadores eligen por considerarlas útiles para estudiar el fenómeno de su interés. Se deberá considerar que las variables tienen *dimensiones* que son pertinentes y otras no, todo esto también condicionado por los objetivos, el marco teórico y las condiciones de factibilidad en su abordaje, ya que es necesario su *operacionalización* para que sean *medidas*. La medición, o generación de valores o categorías, deberá respetar dos tipos de condiciones: a) *lógicas*, según las cuales las categorías generadas deberán ser *exhaustivas* (no deberá quedar ningún individuo fuera) y *exclusivas* (ningún individuo pertenecerá a más de una categoría), y b) *epistémicas*, con un *criterio único* y preciso, con *discriminabilidad*, esto es que debe ser útil a los fines planeados y con *fertilidad heurística*, para que sirva al descubrimiento de nuevas relaciones dentro de una misma variable.

**5.2. Variables no relevantes** (también llamadas *confusionales*) son cualidades del objeto en estudio que no se han considerado para su construcción pero que *existen*, están ubicadas en el contexto, y pueden dar lugar a resultados falsos o sesgados. Se controlan mediante distintas técnicas que deben ser explicitadas en este punto: *exclusión* (eliminando a aquellos individuos que las presenten), *estratificación* (ubicando en diferentes grupos, según la variable que nos interese controlar, a la población en estudio), *registro* de la presencia de la variable que nos interese controlar para analizar luego los resultados en función de sus diferentes valores.