

La investigación: un diálogo con la naturaleza

Para dar inicio a la conceptualización de elementos de la semántica de la estadística es preciso acudir como apoyo a una metáfora. Con diversas estrategias, los investigadores interrogan a la naturaleza para obtener respuestas a sus preguntas de investigación. Así, de alguna manera, el investigador pregunta y la naturaleza responde.

Para este diálogo el investigador requiere diseñar, organizar y llevar a cabo un interrogatorio, así que debe planear etapas admitidas por la comunidad científica y prever respuestas que le permitan comprender el fenómeno estudiado. Como todo diálogo, la investigación ha de contar con elementos mínimos de comunicación que posibiliten tener un lenguaje común en el que las preguntas sean comprendidas por la naturaleza y las respuestas, a su vez, por el investigador.

Una resonancia magnética puede entenderse de una manera muy simple, como la formulación de una pregunta y la emisión de una respuesta. En este símil, la pregunta se formula sometiendo a un paciente a un campo electromagnético con un imán especial que atrae los protones de sus tejidos, induciendo unos fenómenos físicos que culminan en la liberación de energía que se transforma en imágenes como respuesta. La imagen recibida tiene un significado.

Determinar la calidad de vida del cuidador de un paciente crónico es igualmente análogo a la formulación de una pregunta y la emisión de una respuesta. La pregunta se formula con la aplicación de un cuestionario, previamente construido para tal fin, cuya respuesta se concreta en la asignación de una cifra que tiene una interpretación concertada.

Tomar la temperatura de un paciente es otro ejemplo de interrogar a la naturaleza y de registrar su respuesta. La dilatación del mercurio dentro de un tubo de vidrio, inducida por una actividad molecular que presenta el

paciente y visualizada por medio de una escala graduada, es una respuesta indirecta del paciente cuyo significado se le asigna a él mismo.

Las respuestas de la naturaleza son, por tanto, mensajes que requieren ser decodificados, es decir, interpretados por el investigador. El cuidado de la vida es el mandato moral de enfermería y, en este, son diversos los interrogatorios por desarrollar con la naturaleza, cuyas respuestas se derivarán en evidencias que se traducirán en conductas para mejorar la calidad del cuidado y apoyar la práctica de enfermería.

La medición: proceso de decodificación del mensaje

Decodificar las respuestas de la naturaleza exige que el investigador aplique las reglas adecuadas al mensaje para entenderlo; por tanto, medir es el proceso mediante el cual se decodifica el mensaje.

Medir un atributo de un objeto es asignarle un único elemento de un conjunto de símbolos o caracteres siguiendo unas reglas, convención o protocolo. Si uno de los atributos considerados en la investigación es el tipo de diabetes del paciente, podemos disponer de diversos conjuntos de símbolos para medirlo. Por ejemplo, en la figura 1 se presentan tres posibles grupos de símbolos que son combinaciones de letras, signos gráficos o numerales arábigos, entre otros.

Figura 1. Ejemplo de posibles grupos de símbolos con los que se puede medir el atributo tipo de diabetes

Símbolos	Símbolos	Símbolos	Convención
■	1	DM	Diabetes mellitus
⦿	2	DI	Diabetes insípida
♦	3	MOD	Diabetes tipo MODY
*	4	DR	Diabetes renal
Φ	5	AD	Amino-diabetes
◀	6	FD	Fosfato-diabetes
△	7	PE	Prediabetes

Ahora bien, el proceso requerido para asignar un símbolo a un atributo de un objeto precisa de un recurso; por ejemplo, para medir la presentación cefálica fetal, como cefálica vértice (α), bregmática (β), cefálica de frente (δ) y cefálica de cara (π), es necesaria una observación directa que determine si la cabeza del feto está flectada o deflectada por medio de una ecografía. El ecógrafo no registra la posición fetal, pero a partir

de la emisión de sus imágenes se puede determinar. Dentro de las reglas, convenciones o protocolo de su medición, el ecógrafo es un recurso para lograr medir la presentación cefálica fetal.

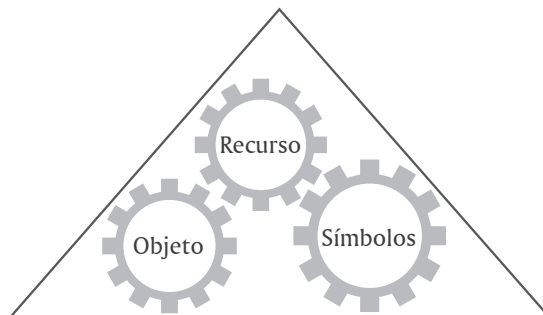
Los recursos pueden tener diferentes niveles de complejidad; por ejemplo, para asignar un símbolo al sexo de un paciente basta solo con observar sus características fenotípicas, mientras que para asignar un símbolo a su presión venosa central se requiere una conjunción de varios elementos, entre los que se encuentran un acceso venoso central, un equipo y una solución salina normal.

De manera corriente, a los recursos se les denomina instrumentos de medición, pues son el medio para obtener respuestas a partir de las cuales el investigador asigna un símbolo pertinente. Por tanto, los instrumentos de medición son el recurso que hace viable la asignación de un símbolo.

Independientemente de su idoneidad y eficacia, un instrumento de medición es más sensible o tiene mayor resolución que otro en la medida en que puede producir un número mayor de señales y, por ende, asignar un mayor número de símbolos para un mismo atributo de interés del objeto. El pH-metro o “peachímetro” es un recurso más sensible que el papel tornasol para determinar la concentración de iones hidrógeno de una sustancia, pues permite asignar símbolos numéricos comprendidos entre 0 y 14, mientras que el papel tornasol únicamente nos permitiría asignar dos: ácido o alcalino. No es que uno sea mejor que el otro, sino que probablemente para una situación particular sea más útil uno de ellos.

En resumen, como medir un atributo de un objeto es hacerle corresponder un único elemento de un conjunto de símbolos, no necesariamente numerales, se puede sintetizar que la acción de medir constituye un proceso que involucra tres componentes: objeto, recurso y símbolos, lo que puede denominarse proceso de medición. La figura 2 esquematiza el engranaje de los tres elementos del proceso de medición.

Figura 2. Elementos del proceso de medición



Cada uno de los componentes del proceso genera conceptos consecutivos, que conformarán en gran medida el contenido de las unidades de este libro. El recurso, como parte inherente a la medición, debe ser un medio eficaz en la transmisión del mensaje que evite interferencias o distorsiones del mensaje. Por ello se precisa el cumplimiento de requisitos orientados a satisfacer exigencias en precisión, exactitud, validez y confiabilidad, entre otros, que acrediten la idoneidad y el desempeño del recurso en su función propia.

En particular, facultar un recurso de la medición como los denominados instrumentos de medición requiere, por ejemplo, el apoyo de procesos de metrología o de conceptos psicométricos, según sea su naturaleza. Las escalas utilizadas en enfermería son recursos especiales tan útiles e importantes como instrumentos de medición, que merecen un espacio especial para su estudio. Sin embargo, dada la finalidad de este libro, los autores declinan ese propósito.

La población y la muestra: conjuntos potenciales y asequibles de interlocutores

La investigación usualmente aborda conjuntos de objetos con el propósito de explorar, describir o inferir algunas de sus propiedades colectivas; en consecuencia, el investigador debe contar con un número plural de estos, ya sea con la totalidad de los objetos de interés o con una parte de ellos elegida de manera apropiada. En cualquiera de las dos situaciones, el proceso de medición se replica tantas veces como objetos se hayan incluido en la investigación.

Para definir con precisión con quiénes se va a desarrollar el diálogo, hay que delimitar un agregado de unidades, es decir, un conjunto de unidades homogéneas, de la misma esencia. Ese agregado de unidades en la ciencia del cuidado, casi siempre definido por unas pautas establecidas en el estudio —como, por ejemplo, los llamados criterios de inclusión—, adopta del lenguaje particular de la estadística el nombre de población, cuya definición se propone más adelante.

La población puede ser un conjunto de turnos de enfermería, de familias, de profesionales de enfermería, de familiares del paciente en cuidado crónico, de instituciones prestadoras de servicios de salud, de pacientes o de encuentros entre paciente y enfermera en la administración de medicamentos, entre otros. El término *población* no es muy acertado para denominar a este conjunto, pues sus acepciones pueden dar motivo a

equivocos conceptuales o metodológicos. Sin embargo, se mantendrá ese vocablo debido a su difusión y permanencia.

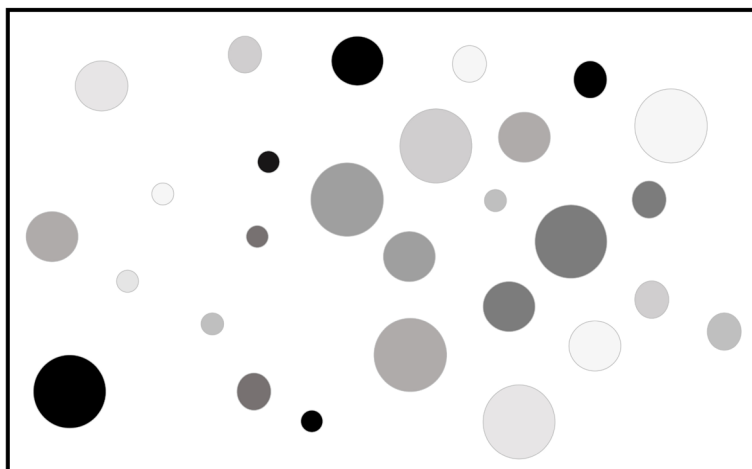
Con el término *población*, o algunas veces *universo*, se hace referencia a un agregado perfectamente definido en todos sus determinantes que es motivo de interés científico, tecnológico o administrativo, el cual está conformado por unidades homogéneas en su esencia, pero disímiles en atributos, que son objeto de examen dentro de la investigación y se abstraen formando un grupo con el fin de determinar los atributos propios y rasgos relevantes que la determinan como una entidad, o que se constituyen en el medio o recurso para inferir propiedades sobre métodos o instrumentos que contribuyan a su caracterización o beneficio.

La adherencia a un tratamiento farmacológico de un grupo de pacientes hipertensos mayores de cuarenta años, que asiste a un programa en una institución de salud, puede ser un tema de interés administrativo para un profesional de la enfermería. En este ejemplo sencillo se puede deducir que la población es justamente el grupo de pacientes que asiste al programa. La esencia de la población está en que todos los pacientes son hipertensos, mayores de cuarenta años y asisten al programa. Con seguridad, unos son hombres y otros mujeres; algunos tendrán una patología asociada particular y otros no la tendrán; unos poseerán una condición económica mejor que el resto y presentarán condiciones laborales diferentes, solo por señalar algunos atributos que hacen de los pacientes personas distintas, aunque compartan plenamente unos criterios que las identifican como integrantes de la población de interés.

La funcionalidad de la familia de la madre lactante puede ser otro ejemplo de un tema de interés. En este caso, la población está constituida por familias y no por personas, aunque estas conformen aquellas. Nótese que cada uno de los interlocutores interrogados es un conjunto formado por una pluralidad de personas interrelacionadas de acuerdo con alguna definición de esta estructura social.

Para sintetizar, los rasgos fundamentales del concepto de *población* se pueden esquematizar en la figura 3. El diagrama que muestra describe un conjunto de unidades homogéneas por su forma circular, pero disímiles en su tamaño, interior y contorno.

Figura 3. Esquema del concepto de población



Las unidades que reúne la población que constituyen dicho conjunto reciben un nombre especial dependiendo del tema científico, protocolo de investigación o área del conocimiento. Por ejemplo, la unidad puede denominarse paciente, sujeto, hogar, cuidador, materna, familia, puesto de trabajo, institución prestadora de servicios de salud, entre muchas. Independientemente de cómo se le llame, se refiere a cualquiera de los interlocutores interrogados. Unidad es la forma genérica de referirse a los elementos que integran la población.

Regresando a la metáfora central de este capítulo —la especie de diálogo que se da entre el investigador y la naturaleza—, cabe averiguar ahora sobre la viabilidad de llevar a cabo esos diálogos con todas y cada una de las unidades de la población. Si la totalidad de diálogos es realizable de manera exhaustiva, garantizando niveles altos y confiables de la calidad de las respuestas, sería el camino metodológico ideal para abordar un estudio en particular. Pero existen limitaciones físicas, económicas o metodológicas que hacen imperativo llevar a cabo la investigación utilizando un subconjunto de unidades de la población elegidas de manera aleatoria, denominado muestra. Entre otras, se toman como limitaciones las siguientes:

- La población es tan numerosa o conceptualmente considerada infinita, que materialmente es imposible acopiar la información requerida de todas y cada una de las unidades.
- Los procedimientos de medición son muy complejos o económicamente muy costosos.
- La población es relativamente homogénea.

- La información y los resultados que se obtendrán del estudio son apremiantes para algún propósito decisivo.
- La información y los resultados que se obtendrán del estudio tendrán validez momentánea, serán vigentes por un corto período de tiempo.
- Es más viable garantizar alta calidad de la información de una fracción, que de toda de la población.

La selección de la muestra debe estar regida por un proceso ajeno a la voluntad del grupo investigador, es decir, no condicionado a la intervención humana, para contribuir con otras acciones al control de los llamados sesgos en la investigación, proceso dotado de un carácter imparcial como exigencia ética, pero metodológicamente ineludible para garantizar inferencias estadísticas válidas relativas a propiedades colectivas de la población.

Por consiguiente, un sorteo resulta ser la manera más neutral de selección de la muestra, contraria a una selección intencional. Suele hablarse de *muestra al azar* o *muestra aleatoria*, como términos acuñados para efectos de este proceso de selección y que se pueden adjetivar adicionalmente como simples, probabilísticas, sistemáticas, de conglomeradas, estratificadas, multietápicas, entre otras, dependiendo de la estructura de la población, de los propósitos y de la metodología de la investigación.

El sorteo para seleccionar una muestra, y el soporte de los argumentos de las inferencias a partir de ella, se fundamentan directamente en el concepto que origina la teoría de probabilidad: el *experimento aleatorio*. Antes de precisar su definición es menester hacer un comentario etimológico relacionado con los vocablos *al azar* y *aleatorio*. El término *aleatorio* proviene del latín, *aleatorius*, y hace su aparición en España en el siglo XIX; su significado es “propio del juego de dados”, de *alea*: dado. El término *azar*, del árabe *az-zahr*, el dado para jugar, vocablo que apareció en España en el siglo XIII, derivado de *zahr*, la flor del naranjo, pues esta estaba representada en una de las caras de los dados usados por los árabes. Proveniente del griego, menos usado pero equivalente a los anteriores términos, es el adjetivo estocástico [1].

En la ciencia del cuidado hay múltiples manifestaciones del azar que trascienden la definición etimológica, pues el azar es un componente de la vida; por tal razón, aunque la palabra se refiera al dado, esta se relaciona con las situaciones sobre las que no se tiene control o poder de decisión. La estadística enfoca el azar no solo al sorteo de unidades para constituir una muestra, sino a circunstancias de algunos fenómenos espontáneos que ella modela y a la estructura de algunos métodos estadísticos.

En síntesis, si todos los elementos de la población son interlocutores asequibles, el diálogo se desarrollaría inexorablemente con todos ellos; en caso contrario, es el azar quien juega el rol seleccionador de los interlocutores.

Las variables: los códigos de comunicación

Como medir implica la correspondencia entre el objeto y el símbolo, la relación matemática entre el conjunto de interlocutores y el conjunto de símbolos da paso al concepto de *variable*, su clasificación y nivel de medición.

Dado un atributo de interés de las unidades de la población y el recurso para su medición, a la correspondencia unívoca entre la población y el conjunto de símbolos previstos para la medición se le denomina *variable*. Es decir, a cada unidad de la población le corresponde un único símbolo. Generalmente recibe el nombre con el cual se le designa al atributo, y se le representa mnemotécnicamente por una letra, usualmente mayúscula, o por la agregación de ellas y otros símbolos. Se acompaña su nombre de las unidades utilizadas, cuando las haya.

La determinación de la adaptación del neonato a la vida extrauterina se mide a través del Apgar. Los símbolos que se aprestan para este efecto originan el conjunto $\{0, 1, \dots, 10\}$. La variable concretada por este conjunto puede etiquetarse de manera simple y nemotécnica, entre muchas otras, Apgar, AP, X, Z W... Si se precisa incluir el momento de la medición, nemotécnicamente pueden usarse etiquetas para ella como APG1, APG5, APG10, para indicar la puntuación al minuto, a los cinco minutos y a los diez minutos de nacer.

El examen del entorno de la práctica de enfermería puede medirse mediante la escala Environment Scale of the Nursing Work Index de 31 ítems. $ITEM_j$, $j = 1, 2, \dots, 31$, puede ser el nombre que recibe la variable genérica referente a la respuesta al ítem j y está formalizada por el conjunto $\{1, 2, 3, 4\}$ que representa sus niveles de respuesta. El total, como suma de los valores convencionales en cada ítem, puede llamarse total, entorno, suma o cualquiera otra denominación y está determinada por el conjunto $\{31, 32, \dots, 124\}$ [2].

La cantidad de minerales en el hueso de una paciente puede ser un atributo relevante dentro de una investigación. Esta se puede medir cuantificando la densidad mineral ósea mediante rayos X, ultrasonidos o isótopos radiactivos. Cualquiera sea el recurso, presenta imágenes que hacen posible calcular la cantidad mineral ósea por superficie. Los símbolos para esta medición (densitometría ósea) serán un conjunto de símbolos que representan un subconjunto de números reales. No solo se refieren a los valores normales

establecidos entre 0,97 y 1,28, sino a inferiores a 0,97 y mayores de 1,28. Las unidades en este caso se representan por g/cm^2 [3].

Al atributo de interés en la medición de un objeto casi siempre se le nombra de igual manera que la variable asociada a él. Sin embargo, es importante aclarar que no se miden variables; se miden atributos del objeto. No se mide el Apgar, se mide la adaptación del neonato a la vida extrauterina.

Formalmente, las variables se clasifican en discretas y continuas. Si la variable ha dispuesto un conjunto finito o infinito enumerable de símbolos —esto es, que ellos puedan ubicarse a manera de una lista, el primero, el segundo, el tercero, etc.—, se dice que la variable es discreta. Cuando dicho conjunto de símbolos configura un conjunto infinito no enumerable, la variable se clasifica como continua. Cabe señalar que la sensibilidad del recurso no define por sí sola el tipo de variable, sino que en la medida en que tenga mayor resolución, producirá mayor número de señales correspondiendo a un mayor número de símbolos.

Si el atributo que se desea medir es el peso del neonato en kilogramos, una báscula puede permitir lecturas correspondientes a números enteros únicamente; otra puede ofrecer lecturas con un decimal, otra con cuatro decimales. Con cada una se pueden asignar símbolos, pero la variable peso del recién nacido a término, que subyace, es independiente del instrumento que genera sus símbolos. La primera báscula producirá, por ejemplo, el símbolo 3, la segunda 3,2 y la tercera, para la misma situación producirá el símbolo 3,2652. Tal vez en un caso concreto no sea útil ni necesaria la tercera balanza, la de mayor resolución, pero el suponer aumento de la sensibilidad de un instrumento motiva a pasar a un plano conceptual superior y mirar el tipo de variable desde las propiedades de los símbolos escogidos.

El número de gestaciones previas de una materna es una variable discreta puesto que se dispone del conjunto $\{0, 1, 2, \dots\}$. El primer elemento, visto como una lista es el 0; el segundo el 1, el tercero el 2, y así sucesivamente. Entre el 0 y el 1 no hay otro símbolo de la variable que modifique esta lista ni entre el 1 y el 2 ni tampoco entre el 2 y el 3, como ejemplos.

Por su parte, la edad de un paciente adulto joven, identificada como X , es una variable conformada por los símbolos numéricos entre 20 y 40, que formalmente se pueden expresar por medio del conjunto $\{x \mid 20 \leq x \leq 40\}$. Si 20 es el primer símbolo de la lista, ¿cuál es el segundo? El símbolo 21 no puede ser, porque 20.5 es un símbolo que podría también serlo. Pero igualmente lo serían 20.1, 20.15 o 20.1567. Como se deduce, no hay un segundo símbolo ni un tercero; no es posible constituir una lista. Es una variable continua.

Algunas variables incluidas en un estudio de crecimiento fetal son continuas, por ejemplo, la edad gestacional, el diámetro biparietal, el grosor placentario y la cantidad de líquido amniótico; no lo es el número de embriones o fetos, porque esta es una variable discreta. En un estudio de entorno de la práctica de enfermería son variables discretas el número de pacientes asignados, el número de pausas en el turno, pero no la intensidad del ruido acústico de un puesto de trabajo.

Se había mencionado que los mensajes de la naturaleza que al interlocutor indagador le interesa recibir deben leerse a la luz de un repertorio de signos y reglas semióticos; es por ello que en muchas circunstancias se busca asistir ese propósito utilizando numerales como símbolos de la medición y así aprovechar de alguna manera las propiedades que poseen los números como entidades abstractas representadas por los numerales. Una propiedad de los números cuyos numerales han sido elegidos como símbolos en una medición será transferida a estos cuando a la luz de la esencia real del atributo del objeto en cuestión esta tenga sentido; por ejemplo, el número 45 es mayor que el número 20. Esta relación puede transferirse al caso de la sobrecarga del cuidador, puesto que tiene 45 puntos de sobrecarga en la Entrevista Zarit, tiene mayor carga que aquel que tiene 20 puntos en la misma escala.

Por otra parte, hay atributos a los cuales no se les pueden transferir las propiedades que tienen los números; por ejemplo, si 1 significa cuidador familiar principal, 2 cuidador a distancia y 3 cuidador supervisor, aunque 3 es mayor que 1, esta relación no es transferible al atributo tipo de cuidador, para el cual se eligieron los símbolos numéricos {1,2,3}.

Se denomina nivel de medición de una variable a cada uno de los grados dentro de una jerarquía establecida por la posibilidad de utilizar entre sus caracteres los símbolos aritméticos: =, $\neq$, <, >, +, -, x, o $\div$. Los niveles son: nominal, ordinal, de intervalo y de razón, y se describen a continuación.

Nivel nominal. Una variable tiene nivel de medición nominal si entre sus símbolos solo es posible aplicar los signos de relación: = y $\neq$. Esto es, si a dos objetos se les asigna el mismo símbolo es debido a que los objetos son iguales en el atributo en consideración. Por supuesto, si ellos difieren, se les asignarán símbolos diferentes. Una variable con este nivel permite hacer distinción entre objetos. El tipo de diabetes, la presentación fetal y el tipo de familia son variables con nivel de medición nominal.

La medición de la presentación cefálica fetal puede producir una variable cuyos símbolos 1, 2, 3 y 4 reemplazan respectivamente las posiciones de vértice, de sincipucio, de frente y de cara. El símbolo 2 no es menor que 4

ni 3 es mayor que 1; aunque alguna de ellas sea la posición más deseable, estos símbolos solo sustituyen los nombres de las posiciones.

El delirium, como déficit en las funciones cognitivas, puede ser medido según si es debido a enfermedad médica, si es inducido por sustancias, si es por abstinencia de sustancias, si es atribuido a múltiples etiologías o si es no especificado. Aunque no práctico, se podrían utilizar para el efecto los símbolos 130, 140, 150, 160 y 170; el símbolo 170 no significa que sea el mayor de todos los símbolos. Por consiguiente, la presentación cefálica y la causa del delirium son, en efecto, variables que tienen nivel de medición nominal.

Nivel ordinal. Una variable tiene nivel de medición ordinal, si entre sus caracteres solo es posible aplicar los signos de relación: $=$, $\neq$, $<$ y $>$. Es decir, si a dos objetos se les asignan dos símbolos distintos, digamos $a \neq b$, entonces $a < b$ o $b < a$, según la convención usada en la medición. Una variable con este nivel permite hacer distinción entre objetos y establecer un orden entre ellos. El estadio de la úlcera por presión es una variable con nivel de medición ordinal.

El DSM-IV, en su sección de trastornos de inicio en la infancia, la niñez o la adolescencia, para referirse al retraso mental utiliza F70.9 para denotar al retraso mental leve, F71.9 para el retraso mental moderado, F72.9 para el retraso mental grave y F73.9 para denotar al retraso mental profundo. Usando los símbolos $<$ y $>$, $F71.9 < F73.9$ (un retraso mental moderado es menos serio que un retraso profundo) y $F72.9 > F70.9$ (un retraso mental grave es más serio que un retraso leve). La variable, en consecuencia, tiene nivel de medición ordinal.

Nivel de intervalo. Una variable alcanza el nivel de medición de intervalo si entre sus caracteres elegidos como numerales solo es posible aplicar los signos de relación y operación: $=$, $\neq$, $<$, $>$, $+$ y $-$. Entonces, si a dos objetos se les asignan dos símbolos distintos, digamos $a < b$, un objeto supera a otro en el atributo en consideración en la cantidad $(b-a)$. Una variable con este nivel permite hacer distinción entre objetos, establecer un orden entre ellos, determinar diferencias cuantitativas, pero no proporciones entre los mismos.

La variable Apgar tiene nivel de medición de intervalo. Al comparar el símbolo 4 con el 8, por ejemplo, el 8 representa un mejor estado general del neonato que el 4. La diferencia $8-4 = 4$ denota el número de puntos de la escala en los cuales un neonato tiene un mejor estado que el otro, pero no tiene sentido decir que uno tiene el doble del otro. La temperatura del paciente es una variable que alcanza este nivel de intervalo.

Nivel de razón. Una variable alcanza el nivel de medición de razón si entre sus caracteres elegidos como numerales es posible aplicar los signos de relación y operación: $=$, $\neq$, $<$, $>$, $+$, $-$, $\times$ y $\div$. Si a dos objetos se les asignan dos símbolos distintos, digamos $a < b$, además de que difieren en el atributo en consideración en la cantidad $(b-a)$, un objeto es $\frac{b}{a}$ veces mayor que el otro en el atributo en consideración. Una variable con este nivel permite hacer distinción entre objetos, establecer un orden entre ellos, determinar diferencias cuantitativas y proporciones entre sí. El peso de la gestante, la edad gestacional y el tiempo de cicatrización de una úlcera por presión son variables con nivel de medición de razón.

El peso al nacer es una variable con nivel de medición de razón. Si el niño *A* pesó 4300 gramos y el niño *B* 2150 gramos, los bebés nacieron con un peso diferente; el bebé *A* pesó más que el *B*; la diferencia de peso fue de 2150 gramos y, además, el niño *B* pesa la mitad de lo que pesa el niño *A*.

Conservamos las categorías de la clasificación de Stevens [4], pero las explicamos de manera diferente. No incluimos el cero como elemento diferenciador entre los niveles de intervalo y de razón, para evitar confusiones. Casi todos los autores señalan al cero como el punto de diferencia entre el nivel de intervalo y el de razón. Se dice que el cero en el nivel de razón es un cero natural, un cero absoluto que indica ausencia o carencia del atributo en consideración, mientras que el cero en el nivel de intervalo es arbitrario.

En el anterior sentido, el número de embarazos previos es una variable con nivel de medición de razón. El símbolo 0 debe asignársele a la materna cuyo embarazo actual es el primero; corresponde al cero natural que significa justamente que la materna no ha tenido embarazos anteriores. Las demás condiciones del nivel pueden verificarse para esta variable.

Para concluir la presentación de esta sección, debemos advertir que diferentes autores refieren a las variables con nivel nominal u ordinal como cualitativas o categóricas, y que aquellas con niveles superiores de intervalo o razón las distinguen como cuantitativas. A la variable con nivel ordinal se le conoce igualmente como cuasicuantitativa. Las variables cualitativas son llamadas binarias o dicotómicas, si poseen solo dos símbolos, o politómicas, en caso de poseer más de dos.

El experimento aleatorio: facilitador en la selección de interlocutores y gestor de los elementos primarios de inferencia inductiva

Un experimento aleatorio es una acción repetible indefinidamente bajo las mismas condiciones, caracterizada por la imposibilidad de predecir el resultado en cualquiera de sus repeticiones y por la factibilidad de describir la totalidad de sus resultados posibles.

Por la etimología del término *aleatorio*, el ejemplo connatural de este concepto es el lanzamiento de un dado y el registro del puntaje obtenido. Es una acción repetible indefinidamente; no es posible pronosticar el puntaje que muestre el dado en cualquiera de sus lanzamientos, pero es perfectamente determinable el conjunto $\{1,2,3,4,5,6\}$ como la totalidad de sus resultados posibles. Otro ejemplo puede consistir en la acción de lanzar dos veces un dado. Con las mismas razones anteriores se puede decir que corresponde a un experimento aleatorio, solo que en este caso este cuenta con 36 resultados perfectamente descritos por el conjunto $\{(i, j) \mid i, j=1,2,3,4,5,6\}$, 36 parejas que representan cada uno de los resultados de la acción.

Sortear una de diez incubadoras de una unidad de cuidado intensivo neonatal e identificar al bebé ubicado en ella, o sortear entre los profesionales de enfermería un(a) enfermero(a) y determinar su nivel de *burnout*, son acciones que pueden idealizarse como experimentos aleatorios. El primero tendrá como resultados posibles el conjunto $\{1,2,\dots,10\}$ y el segundo, el conjunto de puntajes posibles que asigne un instrumento de medición para tal fin.

El conjunto de todos los posibles resultados de un experimento aleatorio se denomina espacio muestral, y cualquier subconjunto suyo, evento aleatorio. De una manera intuitiva, una correspondencia unívoca entre el conjunto de resultado de un experimento aleatorio y el conjunto de números reales se define como variable aleatoria. Es decir, es una asignación de un número real a cada uno de los resultados de un experimento aleatorio. También se le suele representar por una letra, usualmente mayúscula, o agregación de ellas y otros símbolos.

La suma de puntos en los lanzamientos de un dado dos veces es una variable aleatoria, porque al resultado (i, j) le corresponde el número $i+j$. El hacerle corresponder al bebé ubicado en la incubadora seleccionada al azar su Apgar o su peso al nacer hace que el Apgar y el peso del recién nacido sean abstraídos como dos variables aleatorias.

La definición puede presentarse de manera más amplia, si la correspondencia se establece con un conjunto de símbolos. Vistas de esta forma, las variables aleatorias tendrán los mismos niveles de medición mencionados y, a su vez, se deduce que la diferencia con las variables a la luz de la definición de variable del apartado anterior es, en esencia, la selección aleatoria del objeto en la medición. En otras palabras, como la selección de la incubadora fue aleatoria, los atributos del bebé que la ocupa pueden medirse, y las variables derivadas de ese proceso representan variables aleatorias.

Aunando la selección aleatoria de unidades de la población con los símbolos asignados a ellas relativos a una variable de interés, se formula el siguiente experimento aleatorio particular:

Experimento: Sortear una unidad de la población, observarla, examinarla o intervenirla, y medir uno de sus atributos de interés. El resultado de la medición corresponde a cualquier símbolo de la variable en consideración, etiquetada como X , y concebida, por supuesto, como una variable aleatoria.

Primera repetición: Sortear una primera unidad de la población, observarla, examinarla o intervenirla, y medir el atributo de interés. A cualquiera de los resultados posibles registrado en esta repetición se le etiqueta por medio de la variable aleatoria X_1 .

Segunda repetición: Sortear una segunda unidad de la población, observarla, examinarla o intervenirla, y medir el atributo de interés. A cualquiera de los resultados posibles registrado en esta repetición se le etiqueta por medio de la variable aleatoria X_2 .

Tercera repetición: Sortear una tercera unidad de la población, observarla, examinarla o intervenirla, y medir el atributo de interés. A cualquiera de los resultados posibles registrado en esta repetición se le etiqueta por medio de la variable aleatoria X_3 .

Y así sucesivamente, hasta concluir con la última repetición: Sortear una *última* unidad de la población, observarla, examinarla o intervenirla, y medir el atributo de interés. A cualquiera de los resultados posibles registrado en esta repetición se le etiqueta por medio de la variable aleatoria X_n .

Dotada de otras propiedades estadísticas, que no se mencionan en esta oportunidad, a la sucesión de variables aleatorias $X_1, X_2, \dots, X_n$, se le denomina muestra aleatoria. El subíndice n indica el tamaño de la muestra, que corresponde al número de unidades de la población seleccionadas aleatoriamente.

Nótese que la expresión *muestra aleatoria* tiene dos acepciones: la primera, para referirse a un conjunto de objetos elegidos al azar, y la

segunda, como lista de etiquetas. El contexto de un escrito o de una exposición permitirá distinguir si la referencia indica un subconjunto de unidades de la población sorteado para fines inductivos o se está señalado a esta sucesión de etiquetas. Esta acepción es indispensable para construir una definición esencial, el término *estadístico*. En la siguiente sección se expone su significado.

Los datos: los mensajes decodificados

Dato, del latín *datum*, “lo que se da”, se refiere al símbolo de la variable que fue asignado a un objeto como resultado de su medición. Un Apgar de 9 puntos a los cinco minutos de nacer es un símbolo disponible en la determinación de la adaptación del neonato a la vida extrauterina, que se convierte en un dato cuando haya sido asignado al estado general de un neonato en particular: “fue lo dado”, lo obtenido en esa medición particular. El símbolo $1,23 \text{ mg/cm}^2$ es un dato producido por la densitometría ósea de una paciente específica.

Una vez haya concluido el proceso de medición con la última unidad seleccionada, la lista o sucesión de variables aleatorias $X_1, X_2, \dots, X_n$, toman forma en una lista o sucesión de datos referentes a la variable representada por X . Cada una de las variables se convierte en un dato específico, es decir, “lo dado” en cada medición individual. Para distinguir las dos listas, esta última puede denotarse con letras minúsculas como: $x_1, x_2, \dots, x_n$, que dentro de la metáfora central de este capítulo representan a los mensajes decodificados y finalmente recibidos de acuerdo con la naturaleza del diálogo llevado a cabo. Los datos son entonces la transcripción de las respuestas, y, a su vez, el insumo para diferentes análisis que finalmente darán cuenta de las motivaciones del diálogo tomado como analogía de la investigación.

Como la decodificación del mensaje recibido del objeto se materializa en la asignación de símbolos, entendidos como datos, ya sea que la investigación haya tenido una cobertura total o que por medio de una muestra se infiera el comportamiento estadístico de la población, se dispone de un conjunto de datos acopiados referentes a cada atributo estudiado en los objetos incluidos en el estudio. A cada atributo le corresponde una variable que, para efectos de la organización, custodia, guarda y manejo de sus datos en algún medio o documento estructurado en filas y columnas, como lo es una hoja de cálculo, se requiere etiquetarlos explícitamente o mnemotécnicamente por una letra, agregación de ellas y otros símbolos. Un esquema aconsejable del registro de los datos se presenta en la tabla 1.

Varios programas de análisis estadístico, incluido el SPSS [5], si-
guen este patrón. Una fila corresponde a los datos de un objeto relativos
a todas y cada una de las q variables consideradas en la investigación,
mientras que una columna incluye los datos de una variable en particular
relativos a los n objetos incorporados en el estudio. Así, los datos del ob-
jeto $i, i=1, 2,..., n$, ubicados horizontalmente en una de las filas, estarán
representados por $v_{1i}, v_{2i}, v_{3i},... v_{qi}$, y los datos de la variable $j, j=1,2,...,q$,
por $v_{j1}, v_{j2}, v_{j3},..., v_{jn}$, ubicados verticalmente en una de las columnas.

Tabla 1. Esquema sugerido de registro y guarda de datos

Identificación del objeto	Variable 1	Variable 2	Variable 3	.	.	.	Variable q
1	v_{11}	v_{21}	v_{31}				v_{q1}
2	v_{12}	v_{22}	v_{32}	.	.	.	v_{q2}
3	v_{13}	v_{23}	v_{33}				v_{q3}
.	.	.	.		.		.
.	.	.	.		.		.
.	.	.	.		.		.
n	v_{1n}	v_{2n}	v_{3n}	.	.	.	v_{qn}

La figura 4 ejemplifica el esquema sugerido de registro y guarda
de datos en el caso particular de un estudio de sobrecarga y competencia
para el cuidado en el hogar de cuidadores de personas con enfermedad
crónica, usando el programa SPSS.

Figura 4. Caso particular del esquema sugerido de registro y guarda de datos

The screenshot shows the IBM SPSS Statistics Processor interface. The menu bar includes Archivo, Edición, Ver, Datos, Transformar, Analizar, Marketing directo, Gráficos, Utilidades, Ventana, and Ayuda. The toolbar contains icons for file operations, data manipulation, and analysis. The main window displays a data entry form with the following columns: CUIDADOR, ZTOTAL, IMPACTO, COMPETENCIASY EXPECTATIVAS, INTERPERSONAL, COMPETENCIATOTAL, CONOCIMIENTO, UNICIDAD, INSTRUMENTAL, DISFRUTAR, ANTICIPACIÓN, and RELACIÓNINTERACCIÓN. The data is organized into rows, with the first row labeled '1' and the last row labeled '22'. The 'ZTOTAL' column has a value of 59.0. The 'IMPACTO' column has a value of 34.0. The 'COMPETENCIASY EXPECTATIVAS' column has a value of 6.0. The 'INTERPERSONAL' column has a value of 8.0. The 'COMPETENCIATOTAL' column has a value of 170.0. The 'CONOCIMIENTO' column has a value of 23.0. The 'UNICIDAD' column has a value of 28.0. The 'INSTRUMENTAL' column has a value of 24.0. The 'DISFRUTAR' column has a value of 35.0. The 'ANTICIPACIÓN' column has a value of 18.0. The 'RELACIÓNINTERACCIÓN' column has a value of 42.0. The status bar at the bottom indicates 'IBM SPSS Statistics Processor está listo'.

	CUIDADOR	ZTOTAL	IMPACTO	COMPETENCIASY EXPECTATIVAS	INTERPERSONAL	COMPETENCIATOTAL	CONOCIMIENTO	UNICIDAD	INSTRUMENTAL	DISFRUTAR	ANTICIPACIÓN	RELACIÓNINTERACCIÓN
1	1	48.0	34.0	6.0	8.0	170.0	23.0	28.0	24.0	35.0	18.0	42.0
2	2	43.0	25.0	8.0	10.0	159.0	26.0	32.0	22.0	30.0	18.0	31.0
3	3	41.0	27.0	6.0	8.0	158.0	26.0	29.0	20.0	29.0	18.0	34.0
4	4	51.0	29.0	14.0	8.0	166.0	25.0	31.0	21.0	35.0	18.0	34.0
5	5	35.0	19.0	8.0	8.0	195.0	31.0	38.0	29.0	36.0	19.0	42.0
6	6	31.0	21.0	6.0	4.0	173.0	30.0	33.0	21.0	36.0	18.0	38.0
7	7	59.0	32.0	16.0	11.0	153.0	26.0	30.0	20.0	27.0	18.0	32.0
8	8	53.0	35.0	12.0	6.0	159.0	24.0	30.0	20.0	34.0	18.0	33.0
9	9	46.0	28.0	10.0	8.0	148.0	20.0	25.0	19.0	35.0	18.0	33.0
10	10	40.0	26.0	8.0	6.0	150.0	25.0	28.0	19.0	31.0	18.0	29.0
11	11	56.0	33.0	15.0	8.0	156.0	23.0	32.0	19.0	34.0	17.0	33.0
12	12	34.0	20.0	6.0	8.0	162.0	26.0	32.0	21.0	36.0	18.0	29.0
13	13	41.0	27.0	6.0	8.0	143.0	22.0	29.0	21.0	29.0	12.0	34.0
14	14	37.0	24.0	9.0	4.0	160.0	26.0	32.0	18.0	34.0	18.0	32.0
15	15	34.0	21.0	7.0	6.0	144.0	23.0	31.0	18.0	32.0	12.0	29.0
16	16	36.0	24.0	6.0	6.0	139.0	23.0	27.0	18.0	29.0	12.0	34.0
17	17	52.0	40.0	6.0	6.0	139.0	20.0	23.0	19.0	30.0	18.0	29.0
18	18	35.0	25.0	6.0	4.0	153.0	27.0	27.0	20.0	29.0	18.0	32.0
19	19	46.0	29.0	11.0	6.0	150.0	23.0	28.0	18.0	33.0	18.0	34.0
20	20	44.0	34.0	6.0	4.0	156.0	24.0	30.0	22.0	33.0	17.0	34.0
21	21	37.0	21.0	9.0	7.0	163.0	25.0	35.0	22.0	34.0	18.0	29.0
22	22	54.0	33.0	13.0	8.0	152.0	28.0	27.0	20.0	29.0	18.0	34.0

Las ilustraciones referentes a SPSS [5] que se consignan en este libro no tienen como propósito explicar el manejo de este software estadístico paso a paso, sino mostrar que se trata de un recurso útil como cualquier otro.

El registro y guarda de los datos se puede hacer en una hoja electrónica como Excel de Microsoft [6] o directamente generando un archivo de datos de SPSS o de otro software estadístico.

Aparte de su registro y guarda adecuados, los datos como mensajes decodificados son el insumo para procesos consecutivos de su análisis en la investigación, cuya gestión idónea garantiza gran parte del éxito de la misma. En alguno de esos procesos, cada una de las mediciones individuales quedará englobada por medio de expresiones matemáticas adecuadas para efectos de la determinación de los atributos propios del grupo de unidades elegidas, o para la inferencia de estos a la población establecida como una entidad y razón de la investigación. En este sentido, se presenta a continuación la definición del objeto matemático que permite llevar a cabo el englobe de los datos.

Un estadístico es una variable aleatoria que depende única y funcionalmente de las aleatorias $X_1, X_2, \dots, X_n$, y de constantes conocidas. Es, decir, es la condensación de las n variables de la muestra aleatoria en una variable. Un valor particular del estadístico se obtiene a través de su expresión matemática utilizando los datos $x_1, x_2, \dots, x_n$. En el caso particular de haber podido llevar a cabo el proceso de medición en todas y cada una de las unidades de la población —es decir, se acopiaron todos los datos de la población—, una expresión matemática que los condense en un único valor se denomina parámetro.

La expresión

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i ,$$

que llamaremos posteriormente promedio de la muestra es un estadístico porque es la suma de las variables aleatorias que conforman la segunda acepción de muestra aleatoria dividida por el tamaño de la muestra, que es una constante determinada.

La expresión

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 ,$$

que llamaremos varianza de la muestra es un estadístico; depende de las variables de la muestra aleatoria, del promedio y tamaño de la muestra.

La expresión

$$T_c = \frac{(\bar{X} - \mu_0)\sqrt{n}}{S} ,$$

también es un estadístico, teniendo en cuenta que depende del promedio de la muestra, de la raíz cuadrada de la varianza de la muestra y de dos constantes que deben ser conocidas. Por supuesto, un valor particular de cada una de ellas será calculado con los datos obtenidos.

En capítulos posteriores será frecuente la presencia de estas expresiones matemáticas, pues los diferentes métodos de inferencia cuentan con ellas para sus diversas finalidades.

Ejercicios del capítulo 1

Propósito: Consolidar la apropiación de los conceptos de *medición* y *variable*, y particularmente, la identificación del nivel de medición, por medio del examen de los caracteres explícitos o implícitos de variables y de la pertinencia de la utilización de los símbolos aritméticos entre los caracteres, como también iniciar el acercamiento al manejo de un archivo de datos.

a) Mc Collum et al., 2005, en su estudio sobre diferencias de género en diabetes mellitus y su efecto en el autocuidado, definieron las variables que se listan a continuación. Identifique su nivel de medición.

Variable	Nivel de medición
Edad	
Raza (blanca, negra, hispanica, otra)	
Años de educación	
Nivel de ingresos (pobre, bajos ingresos, medios ingresos, altos ingresos)	
Índice de masa corporal kg/m2	
Número de comorbilidades	
Tiene limitaciones físicas - Sí/No	
Fuma - Sí/No	
Tiene depresión - Sí/No	
Tiene limitaciones cognitivas - Sí/No	

b) Whitlow *et al.*, 2015, en su estudio sobre estrategias para mejorar la cooperación de grupo en estudios de calidad de vida, definieron las variables que se presentan en la siguiente tabla. Identifique su nivel de medición.

Variable	Nivel de medición
Tipo de cáncer (sistema nervioso central, tumor sólido, leucemia, otro)	
Puntaje de calidad de vida (SF-36)	

c) Identifique el tipo de variable, discreta o continua, según sea el conjunto de sus símbolos y del aspecto del objeto que subyace a ella.

Variable	Tipo
Nivel de ácido úrico	
Tiempo de analgesia posoperatoria	
Índice de masa corporal kg/m ²	
Número de comorbilidades	
Edad gestacional	
Frecuencia cardíaca	
Número de hijos	

d) Guarde en una hoja de cálculo los datos de una muestra de 25 pacientes, referentes a las variables cuyos nombres encabezan las columnas de la tabla siguiente.

Paciente	Índice de masa corporal	Presión sistólica	Presión diastólica	Glicemia	Hemoglobina glicosilada	Colesterol
1	21,46	120	77	97	11	4,08
2	20,06	120	72	98	12	4,01
3	21,82	101	71	101	10	5,46
4	20,09	114	72	118	9	3,68
5	19,42	121	73	122	11	3,51
6	21,21	110	88	80	10	4,51
7	21,63	121	78	93	9	3,49
8	22,23	128	80	80	12	5,17
9	21,79	120	61	96	10	4,67
10	21,65	124	78	85	11	3
11	19,16	121	69	117	9	5,92
12	23,3	120	81	109	10	3,89

Continúa en la siguiente página.

Paciente	Índice de masa corporal	Presión sistólica	Presión diastólica	Glicemia	Hemoglobina glicosilada	Colesterol
13	21,89	120	78	92	12	3,7
14	21,88	114	78	130	10	3,74
15	24,17	120	86	93	12	4,42
16	24,95	124	62	114	10	4,52
17	19,95	128	78	70	12	4,27
18	20,42	111	74	96	10	2,98
19	20,29	103	73	121	12	3,72
20	21,63	123	73	129	9	4,79
21	24,16	119	83	87	9	3,93
22	21,8	120	78	101	9	4,5
23	22,64	116	81	108	12	3,23
24	21,73	100	70	105	11	4,5
25	21,79	128	89	125	9	3,8

e) Si se nombra a la glicemia como X , indique qué valores son, respectivamente, x_1, x_2, x_{13}, x_{24} y x_{25} . Ordene el archivo anterior, de menor a mayor, con respecto a la glicemia. Con $x_{(1)}$ se representa el valor de la glicemia menor; con $x_{(2)}$ el siguiente valor ordenado; con $x_{(3)}$ el valor que le sigue al segundo, y así sucesivamente, de tal manera que $x_{(25)}$ representa el mayor valor de la glicemia. Ahora indique qué valores son, respectivamente, $x_{(1)}, x_{(2)}, x_{(13)}, x_{(24)}, x_{(25)}$.

f) Un estadístico tiene como expresión matemática,

$$\hat{\delta}_{25} = \frac{(x_{(2)} - x_{(1)}) + (x_{(25)} - x_{(24)})}{25x_{(13)}}$$

¿Qué valor numérico tiene este estadístico en la muestra?