

Recolección y análisis de datos

Carlos Saúl Juárez Lugo¹

En el proceso de la investigación cuantitativa, el vínculo entre la teoría y el trabajo de campo es la etapa de la recolección de datos. Esta fase del proceso de indagación científica requiere del diseño de un plan que indique las acciones y procedimientos que conduzcan al investigador a reunir datos suficientes para dar respuesta a las preguntas de investigación. Dicha planificación del proceso de investigación requiere, entre otras acciones, identificar las fuentes de donde se obtendrán los datos (personas, observaciones, documentos, sucesos, etcétera), el lugar en donde se localizan tales fuentes, definir su muestra, establecer a través de qué medio o método el investigador va a recolectar los datos, así como la forma en que serán procesados y analizados (Bisquerra, 2009).

A la recolección de datos se encuentra vinculada la acción de medir o contar. Hablar de medición implica el empleo de un instrumento calibrado y una unidad de medida establecida para identificar la magnitud de la variable. Así también, el investigador puede contar los atributos de la variable de interés. En ambos casos, contar o medir, significa

¹ Centro Universitario UAEM Ecatepec, Universidad Autónoma del Estado de México, México.

asignar valores a las propiedades de los objetos, de los eventos o de las personas; el registro de estas propiedades por medio de los instrumentos de recolección de datos tiene que ser de la manera más fehaciente y cumplir con los criterios de confiabilidad y validez. El investigador que utiliza un instrumento para recolectar información de las variables de su interés está obligado a cumplir con dos requisitos: el principio de la confiabilidad y el principio de validez.

La **confiabilidad** de un instrumento de recolección de datos puede definirse como la ausencia relativa de errores de medición en un instrumento, se refiere al grado en que sus mediciones son consistentes (Kerlinger y Lee, 2002). Por ejemplo, si el investigador está interesado en medir la glucemia de un grupo de estudiantes utilizando para ello un glucómetro y este indicará en una primera medición que Laura obtiene un valor de 80 mg/dL, Agustín 93 mg/dL y Luis un valor de 72 mg/dL; 30 minutos más tarde el investigador repite la medición, los valores de Laura y Agustín no han cambiado, pero el de Luis se modificó a un valor de 100 mg/dL. En este ejemplo es posible afirmar que el instrumento es confiable en dos sentidos: por una parte, los resultados que produce el instrumento son constantes, es el caso de Laura y Agustín, por otra, el instrumento es sensible a las modificaciones de la variable medida y las podemos explicar; es el caso de Luis, pues ingirió unas galletas de chocolate inmediatamente después de la primera medición.

La confiabilidad de un instrumento de medición se calcula a través de los llamados coeficientes de confiabilidad que son fórmulas estadísticas cuyo resultado puede oscilar entre cero y uno, donde un coeficiente de cero significa que el instrumento no es confiable y un valor de uno representa una confiabilidad perfecta. Cuanto más se acerque el coeficiente a uno, menor error habrá en la medición y mayor confiabilidad tendrá el instrumento.

Los procedimientos utilizados con mayor frecuencia para establecer la confiabilidad de un instrumento de recolección de datos son el coeficiente de correlación r de Pearson y el coeficiente alfa de Cronbach. El primer coeficiente tiene tres modalidades que son: 1) confiabilidad test-retest, en el cual el instrumento es aplicado a los mismos participantes en dos ocasiones diferentes, con un intervalo de tiempo suficiente para disminuir la influencia del recuerdo sobre las respuestas. 2) El

método de formas equivalentes o paralelas, consiste en aplicar dos formas alternativas de la misma prueba a los participantes. 3) El método de división por mitades consiste en fraccionar en dos partes equivalentes la prueba de interés, la sumatoria de las respuestas de los reactivos nones – la sumatoria de los reactivos pares. En los tres métodos, cada uno de los participantes tendría un par de puntuaciones ($x - y$), los valores serían utilizados en la fórmula de r de Pearson.

El coeficiente alfa de Cronbach es uno de los métodos más potentes para medir la consistencia interna del instrumento de recolección de datos. Su cálculo comprende la suma de las varianzas de cada reactivo y la varianza total del instrumento. Este método es ampliamente utilizado por distintos investigadores para medir la confiabilidad o consistencia interna de un instrumento debido a que su cálculo asume que los ítems miden un mismo constructo y están altamente relacionados entre sí. El cálculo del coeficiente alfa de Cronbach se puede realizar de manera ágil con el Paquete Estadístico para las Ciencias Sociales mejor conocido como SPSS, por sus siglas en inglés.

Por otra parte, el término de **validez** de un instrumento de recolección de datos se sintetiza en la pregunta ¿el investigador está midiendo lo que en verdad quiere medir? La respuesta a la pregunta en apariencia es sencilla y simple, si conocemos la variable podemos hacer preguntas acerca de esa variable que conocemos (Kerlinger y Lee, 2002). Sin embargo, en la práctica no es tan sencillo asegurar que estamos midiendo la variable, lo que queremos medir, sobre todo cuando el investigador construye el instrumento de recolección de datos.

El investigador debe reunir evidencia relacionada con al menos tres tipos de validez. La primera evidencia tiene que ver con el contenido y se denomina así, *validez de contenido*. Se refiere al grado en que los reactivos del instrumento son representativos de cada reactivo del universo teórico posible. La validez de contenido implica juzgar la pertinencia del reactivo como parte del contenido definitivo del instrumento. Esto se logra, por una parte, con la literatura teórica y empírica que explica y describe las variables que el investigador pretende medir. Por otra parte, el juicio de expertos en el tema de estudio indicará al investigador la pertinencia de cada reactivo para medir el constructo. Un instrumento que pretenda medir la habilidad para resolver problemas de aritmética

básica debe incluir preguntas relacionadas con operaciones de suma, resta, multiplicación y división.

La segunda evidencia es la *validez de criterio*, se estudia al comparar las puntuaciones de una prueba con una o más variables externas, o criterio, que se sabe está relacionada con el atributo que se estudia. Se divide en dos: 1) validez predictiva, involucra el uso de criterios futuros para establecer la validez del instrumento. Por ejemplo, cuando se predice el fracaso o éxito escolar de los estudiantes a partir de sus medidas de aptitud académica ¿Qué tan bien predice la prueba EXANI II el promedio final de la licenciatura? 2) La validez concurrente implica el uso de criterios presentes, por ejemplo, si el investigador desarrolla una prueba de inteligencia nueva y correlaciona esos puntajes con los datos de la prueba de inteligencia ya existente, se espera una correlación positiva entre ambas puntuaciones, es decir, ambas miden el mismo constructo.

Por último, la *validez de constructo* reúne información acerca de qué tan bien las propiedades del instrumento explican el modelo teórico empírico que sustenta a la variable de interés del investigador. Esta validez tiene un alto grado de complejidad tanto para su cálculo que se realiza con procedimientos de análisis factorial como para su pretendida explicación pues requiere de un marco teórico sólido que sustente la relación entre las variables que pretende medir el instrumento (Zamora *et al.*, 2010).

Instrumentos de recolección de datos

Una vez identificados los requisitos que deben cumplir los instrumentos para la recolección de datos podemos conocer algunas de estas herramientas. Los métodos de recolección de datos son variados en su forma, estructura y medio de aplicación. El instrumento más utilizado por el novel investigador es el cuestionario, seguramente por su aparente facilidad en la elaboración, aplicación y procesamiento de la información. Un cuestionario es un instrumento de recopilación de datos que operacionaliza determinados problemas que son objeto de investigación. Esta operacionalización delimita estrictamente las variables en factores medibles, se realiza mediante la formulación escrita de una serie de enunciados interrogativos que, respondidos por las personas muestra,

permiten examinar el objeto de estudio de la investigación (Cohen y Swerdlik, 2007).

El contenido de las preguntas de un cuestionario puede ser de naturaleza abierta o cerrada. Las preguntas abiertas son aquellas cuya respuesta no se ha establecido *a priori* y el respondiente la elabora en el momento. Tiene la ventaja de que el informante puede redactar la respuesta con libertad. Su inconveniente principal es la categorización de todas las respuestas posibles. La clasificación y registro de la respuesta abierta depende, por una parte, de la claridad y extensión con la que fue redactada la idea, por otra parte, de la habilidad y experiencia del investigador para agrupar en número suficiente sin perder información relevante o tergiversarla.

Las preguntas cerradas se caracterizan por que su posible respuesta ya fue establecida por el investigador. La alternativa para contestar puede ir de un Sí o No hasta varias posibles respuestas en las que se puede elegir solo una o más de una opción de respuesta de manera simultánea. Las preguntas cerradas son más fáciles de codificar y analizar, sin embargo, en ocasiones, para incluir todas las posibles respuestas es necesario una larga lista de alternativas. Un ejemplo de una pregunta cerrada es la siguiente:

¿Cuáles fueron las causas que influyeron en tu cambio de residencia?

- a) Cuestiones familiares
- b) Cuestiones económicas/laborales
- c) Cuestiones de inseguridad/violencia
- d) Cuestiones educativas y profesionales
- e) Otro (especificar): _____

En la recolección y análisis de datos es importante considerar la categorización que se hace de las respuestas posibles de los participantes debido a que esos valores, dependiendo del nivel de medición del ítem, le corresponde un análisis estadístico particular, así como su interpretación pues difieren según una encuesta o una escala de actitud, por ejemplo.

Escala para medir la actitud

Otro método utilizado con frecuencia por los investigadores para recolectar datos por medio de enunciados son las llamadas escala de actitud.

Existen distintos métodos para obtener información relativa de la predisposición que una persona dice tener hacia algo o alguien. El método más común para su medición es la escala de actitud tipo Likert, usada con frecuencia para la exploración de la percepción pública en áreas de la comunicación, la política, la salud, la mercadotecnia y la psicología.

Aiken (2003) señala que una actitud es una predisposición aprendida para responder positiva o negativamente ante un objeto, una situación, institución o persona en particular. Las actitudes son un indicador de la conducta de un individuo, pero no la conducta en sí. Las actitudes tienen tres componentes: el afecto hacia el objeto de actitud, el conocimiento que se tiene de él y la posible respuesta conductual que genere. Así también las actitudes tienen una dirección (positiva o negativa) y una intensidad (alta o baja).

La escala de actitud tipo Likert, llamada así por su creador Rensis Likert, consiste en una serie de juicios en forma de enunciados ante los cuales se solicita la reacción del participante. El enunciado que se muestra al sujeto y su respuesta, solicitada en términos de grado de acuerdo o desacuerdo, representa la propiedad que el investigador está interesado en medir. La sumatoria del valor de las respuestas indica la intensidad de la actitud de la persona o personas que participaron en la encuesta. Ejemplo de un reactivo tipo Likert:

Mi rendimiento académico depende de mi capacidad.

1. Muy en desacuerdo, 2. En desacuerdo, 3. Indeciso, 4. De acuerdo, 5. Muy de acuerdo

Pruebas estandarizadas

Otra forma común para recolectar datos son las llamadas pruebas estandarizadas. Son instrumentos diseñados específicamente para medir un tipo de variable particular, cuentan en la mayor parte de los casos con evidencia de validez y confiabilidad respecto a una población específica. Estos instrumentos estandarizados han sido utilizados por varios investigadores en distintos países, sus reportes científicos son publicados en revistas especializadas, es allí donde se puede encontrar información pormenorizada de esos instrumentos.

Las pruebas estandarizadas tienen la ventaja de que ya están elaboradas y la comunidad científica ha probado su eficacia en poblaciones

específicas, muestran evidencia de su validez y confiabilidad, así como de los resultados obtenidos con el instrumento y su relación con otras variables. Esta información fortalece el trabajo del investigador al compararla con sus propios resultados. La desventaja de estos instrumentos radica en que muy probablemente fueron creados en países, culturas y poblaciones diferentes a las que el investigador está interesado en aplicarlas. El investigador requiere buscar el instrumento de entre varios posibles y elegir aquel que mejor mida las variables que desea medir.

La codificación. Entre el instrumento y el análisis

Los datos recolectados con un instrumento de medición deben codificarse y registrarse de manera ordenada en una hoja especial de un programa de computadora como es Excel o el Paquete Estadístico para la Ciencias Sociales (SPSS). La captura de los datos se da por medio de códigos asignados a cada respuesta de todas las preguntas del instrumento. Una vez capturada la información el investigador está en la posibilidad de realizar el análisis estadístico pertinente. Aquí la importancia de comprender el proceso de codificación y captura de los datos de una investigación.

En primer lugar, es primordial identificar el nivel de medición del reactivo pues al investigador le dará pistas del análisis estadístico que puede realizar con esa información. El nivel de medición, explica Ritchey (2008), identifica las propiedades de medición de la variable (cualitativa-cuantitativa) y prescribe el tipo de operaciones matemáticas que pueden usarse adecuadamente con dicho nivel, así como el análisis estadístico pertinente para probar las hipótesis.

El nivel de medición nominal clasifica la variable de interés al menos en dos categorías, mismas que no tienen un orden jerárquico. Indica diferencias respecto a una o más características. Las operaciones matemáticas se limitan al conteo de casos de cada categoría de la variable y la comparación entre ellas. Los números utilizados para identificar las categorías tienen una función de clasificación y no se pueden manipular aritméticamente. Un ejemplo de este nivel de medición es la asignación de números a la variable sexo: Hombre = 1, Mujer = 2. Es posible obtener

la frecuencia de cuántos hombres y cuántas mujeres hay en un grupo en particular. Si calculamos el promedio de la variable sexo el resultado sería 1.5 [$1 + 2 = 3/2 = 1.5$]. Aritméticamente es correcta la operación, pero no existe un sexo clasificado como 1.5.

El nivel de medición ordinal clasifica a la variable en varias categorías. Se distingue de la nominal porque tiene la propiedad de clasificar las categorías de mayor a menor, de la mejor a la peor o de la primera a la última. Las categorías establecidas indican jerarquía, orden, pero no se conoce con precisión qué tanto es la diferencia entre una y otra, por ello no se pueden realizar operaciones aritméticas. Un ejemplo del nivel de medición ordinal es la clasificación de clase social: $3 = alta$, $2 = media$, $1 = baja$. Las escalas de actitud tipo Likert pertenecen en estricto sentido a este grupo: *muy de acuerdo*, *de acuerdo*, *indeciso*, *en desacuerdo*, *muy en desacuerdo*.

El nivel de medición nominal y ordinal forman el grupo de variables cualitativas y el investigador las utiliza para establecer grupos de comparación. El otro grupo está formado por el nivel de medición de intervalo y de razón, incluyen a las variables llamadas cuantitativas.

El nivel de medición de intervalo incluye las cualidades nominales y ordinales, además la distancia entre las puntuaciones (intervalos) es la misma pues su origen es una unidad de medida establecida. El cero es arbitrario, no es real. Es posible realizar operaciones aritméticas y sus derivaciones. Como ejemplo tenemos la medición de la temperatura con un termómetro.

El último nivel de medición es el de razón. Comprende las cualidades de los tres niveles anteriores y se distingue por que el valor cero es real, indica la ausencia del atributo. De igual manera es posible realizar operaciones aritméticas y estadísticas. Como ejemplo tenemos el número de horas en un día que un estudiante utiliza las redes sociales.

Cuando ideamos un instrumento de recolección de datos, un análisis de contenido, un registro de observación, o bien utilizamos algún instrumento que ya existe como lo es una prueba psicométrica, debemos pensar en cómo se van a codificar las respuestas, la estructura de la matriz y el análisis estadístico que corresponde a esos mismos datos. Una vez que el investigador identificó el nivel de medición del reactivo (nominal, ordinal, intervalos, de razón), puede asignar los códigos de las categorías a las respuestas del instrumento de recolección de datos.

Cada uno de los reactivos deberá tener su propio código y en conjunto formarán el llamado libro de códigos que será utilizado por la persona que efectúe la captura en un archivo de programa de computadora. Por ejemplo, las siguientes preguntas se codifican y se clasifican según el nivel de medición (Tabla 14.1).

Tabla 14.1. Ejemplo de la manera en que se codifican
y se clasifican las preguntas

Pregunta	Opción de respuesta	Se codifica
Edad	[____]	El que indica en años cumplidos (nivel de medición = razón)
Sexo	Hombre [_] Mujer [_]	Hombre = 2, Mujer = 1 Nivel de medición nominal
Máximo grado de estudios	Primaria [_]	1 = Primaria
	Secundaria [_]	2 = Secundaria
	Preparatoria [_]	3 = Preparatoria
	Licenciatura [_]	4 = Licenciatura
	Posgrado [_]	5 = Posgrado (nivel de medición = ordinal)

Fuente: elaboración propia.

La matriz de datos

Una vez que se obtuvo la información y se codificó, se procede a su captura en una matriz en la que sus columnas harán referencia a las variables, los renglones o filas a los casos (sujetos) y en la intersección la clave o el valor asignado al fenómeno observado (conducta) como se puede apreciar en la figura 14.1.

Las principales recomendaciones para diseñar una matriz para la captura de datos como se muestra en la figura son las siguientes:

1. Es indispensable que los cuestionarios tengan un número de folio con el que se va a identificar a cada sujeto al vaciar sus respuestas en la matriz.

2. La primera columna y fila son para indicar los casos y los tratamientos respectivamente. En ocasiones es necesario utilizar más de una columna y más de una fila por el tipo de casos que se va a manejar.
3. Indicar el rótulo para cada columna. Este rótulo, también conocido como encabezado de columna, deber ser corto y claro pero que informe acerca del contenido, por ejemplo, *Preg01*, *Suma*, *Prom*. Microsoft Excel reconoce los encabezados en varias funciones estadísticas.
4. Por la naturaleza del cuestionario, una pregunta se puede dividir a su vez en varias columnas que de igual manera se tienen que distinguir una de otra.

	A	B	C	D	E	F	G	H	I	AD	AE	AF	AF
1	SUJETO	Edad	Sexo	Calf	PREG 1	PREG 2	PREG 3	PREG 4	PREG 5	ANSIEDAD	CONFIANZA	MOTIVACIÓN	TOTAL
2	1	19	1	79	5	4	3	3	3	29	13	11	53
3	2	19	1	79	4	4	4	3	5	39	14	13	66
4	3	20	1	80	4	3	4	2	5	31	12	13	56
5	4	18	1	66	4	4	3	2	4	30	12	12	54
6	5	22	2	73	5	4	4	3	4	34	12	12	48
7	6	19	1	74	4	4	4	3	5	32	12	13	57
8	7	19	2	69	4	4	3	3	5	32	12	14	58
9	8	19	1	86	5	4	4	4	4	36	14	13	63
10	9	19	1	72	5	4	3	2	4	27	14	11	52
11	10	19	1	97	4	3	3	3	3	26	14	7	47
12	11	24	1	86	5	4	4	3	4	37	14	12	63
13	12	20	1	90	4	3	3	3	4	32	13	11	56
14	13	19	2	90	5	4	2	3	4	28	13	10	51
15	14	20	1	98	3	3	3	3	4	25	15	9	49

Figura 14.1. Ejemplo de una matriz para la captura de los reactivos de un instrumento de recolección de datos en una hoja de Excel.

Fuente: elaboración propia.

En la actualidad, se ha extendido el uso de recursos tecnológicos para el diseño, elaboración y aplicación de cuestionarios por medio de formularios electrónicos como los que incluye Teams y Google Drive. Estos programas gratuitos de administración de encuestas agilizan la obtención de datos pues se aplican de manera simultánea a un gran número de personas y devuelven una base de datos rudimentaria.

Análisis de datos

Uno de los objetivos de la investigación científica es el de generalizar los resultados que se encontraron en la muestra a la población a la que pertenecen. Esta extrapolación es posible gracias a la herramienta llamada estadística inferencial que incluye las técnicas que emplean los datos obtenidos en la muestra para, a partir de ellos, probar hipótesis y hacer inferencia sobre sus respectivas poblaciones (Pagano, 2006), si es alta la probabilidad de que la media de la muestra esté cerca de la media de la distribución normal podrá hacer generalizaciones (Torres, 2019).

Para probar las hipótesis inferenciales se pueden realizar dos tipos de análisis estadísticos: los análisis paramétricos (Aron, 2001) y los no paramétricos (Siegel y Castellán, 2003). Cada tipo de análisis estadístico exige ciertas características y presuposiciones que lo sustentan y que de ellas depende la elección de qué clase de análisis efectuar; ambos tipos de análisis no son excluyentes en una investigación, es posible complementar con ambas tomando en cuenta el nivel de medición y las variables que conforman las hipótesis.

En las últimas décadas el desarrollo y perfeccionamiento de las pruebas estadísticas para el análisis de datos ha sido enorme. Esta diversificación en la estadística inferencial ha llevado a establecer un conjunto de criterios que se tienen que tomar en cuenta para elegir la mejor prueba estadística válida cuando se quiere probar una hipótesis en un diseño de investigación específico. Además, la posibilidad de acceder a programas de cómputo que agilizan el análisis de datos, como son Excel y SPSS, ahorran el tiempo de cálculo de manera manual, sin embargo, el usuario debe tener claro qué es lo que está haciendo con los datos y cómo interpretarlos.

Siegel y Castellán (2003) establecen una lista de cinco criterios necesarios a observar para elegir la prueba estadística más conveniente a la naturaleza de la investigación que se realiza. Los criterios son: modelo estadístico, potencia, eficacia, medición y pruebas estadísticas.

El modelo estadístico

Este criterio comprende las condiciones, generales y particulares, bajo las cuales tenemos que decidir si utilizaremos los modelos estadísticos

paramétricos o los no paramétricos. Desde un punto de vista general, la naturaleza de la población y la forma de muestreo que fue delimitado en nuestro diseño de investigación, establecen un modelo estadístico. Por ejemplo, si la población se distribuye normalmente, tiene varianzas homogéneas, la muestra de individuos es representativa y la puntuación de sus cualidades se realiza con una escala de intervalos, tomaremos una dirección hacia el modelo paramétrico. Ahora bien, el modelo estadístico, como criterio para elegir una prueba específica, indica que cada una de las pruebas establece un conjunto de requisitos que el investigador está obligado a observar y cumplir, estos son: los requisitos de tamaño de la muestra, número de grupos, nivel de medición de las variables y distribución de la población bajo la curva.

Potencia

De acuerdo con Siegel y Castellán (2003), la potencia de una prueba se define como la probabilidad de rechazar la hipótesis nula cuando de hecho es falsa (potencia = $1 - \text{probabilidad de error tipo II} = 1 - \text{Beta}$). Sucede en ocasiones que al tomar decisiones con respecto a la hipótesis nula podemos cometer errores de dos clases:

- Error tipo I, consiste en la decisión de rechazar la hipótesis nula cuando esta es verdadera.
- Error de tipo II, consiste en la decisión de no rechazar la hipótesis nula cuando esta es falsa.

Siegel y Castellán (2003) destacan que la potencia de una prueba, $1 - \text{Beta}$, mide la probabilidad de rechazar acertadamente la hipótesis de nulidad, es decir, cuando es falsa. Así también, la potencia está relacionada con la dirección de la hipótesis (H_1), una prueba de una cola es más poderosa que una de dos. Estos autores afirman que la potencia de una prueba estadística se incrementa al aumentar el tamaño de la muestra, de aquí la importancia de comprender las condiciones adecuadas para utilizar una prueba estadística paramétrica o no paramétrica específica, es decir, su eficacia.

Eficacia

Este criterio resalta las condiciones de una prueba en la cual opera con mayor eficacia que otra prueba similar. Es necesario conocer las cualidades y requisitos bajo los cuales dichas pruebas (paramétricas y no paramétricas) alcanzan el máximo nivel posible de potencia – eficacia, como puede ser el nivel de medición, el número ideal de individuos en la muestra, las colas de distribución correspondiente a cada prueba.

Medición

La medición de un sujeto específico de la muestra en una variable es la puntuación del sujeto para esa variable. Por lo que al contextualizar esa puntuación en un nivel de medición ayuda a identificar el tipo de operaciones matemáticas (suma, resta, raíz cuadrada, etc.) que puede usarse apropiadamente con dicho nivel, así como las fórmulas estadísticas que utiliza para probar las hipótesis teóricas (Ritchey, 2008). Si las puntuaciones satisfacen los criterios del nivel de medición nominal u ordinal podemos aspirar a utilizar los modelos no paramétricos. Si las puntuaciones satisfacen los criterios de los niveles de medición de intervalo o razón utilizaremos la estadística paramétrica.

Pruebas estadísticas

Para tomar la decisión de utilizar el análisis paramétrico debe partirse de los siguientes supuestos (Torres, 2019):

1. La distribución poblacional de la variable dependiente es normal.
2. El nivel de medición de la variable dependiente es por intervalos o razón.
3. Cuando dos o más poblaciones son estudiadas, tienen una varianza homogénea.

Para utilizar el análisis no paramétrico, debe partirse de los siguientes supuestos:

1. La distribución de la población se acepta como no normal.
2. El nivel de medición de las variables es nominal u ordinal.

Existe una amplia gama de pruebas estadísticas inferenciales. Cada una de ellas se basa en un modelo específico y establece requisitos particulares (Gómez y Bowater, 2015). Veamos algunos ejemplos de las pruebas estadísticas inferenciales paramétricas y no paramétricas. Las pruebas que a continuación se presentan cumplen con la mayor parte de los criterios mencionados y se pueden calcular con el programa SPSS.

***r* de Pearson**

Es una prueba estadística paramétrica que se utiliza para analizar el grado de relación que existe entre dos variables medidas en un nivel por intervalos o razón (Figura 14.2). La hipótesis por probar es correlacional del tipo “*A mayor X, mayor Y*”, “*A menor X, menor Y*”, “*A mayor X, menor Y*”, “*A menor X, mayor Y*”.

$$r = \frac{n(\sum XY) - (\sum X)(\sum Y)}{\sqrt{\left[n(\sum X^2) - (\sum X)^2 \right] \left[n(\sum Y^2) - (\sum Y)^2 \right]}}$$

Figura 14.2. Fórmula de *r* de Pearson.

Fuente: elaboración propia.

La correlación puede variar de -1.00 a +1.00, cuanto más se acerque el valor del coeficiente *r* a 1 mayor será la relación, entre más se acerque a cero la relación será nula. Esta prueba estadística es sensible a valores extremos motivo por el cual es necesario depurar la muestra para evitar errores en la interpretación del resultado.

Coefficiente de correlación de rango de Spearman *r_s*

La versión no paramétrica para calcular la correlación entre variables es *r_s* Spearman. Se utiliza para saber si dos conjuntos de puntajes están relacionados o el grado de esa correlación. Esta medida de asociación requiere que ambas variables sean medidas por lo menos en una escala ordinal, de manera que los objetos o individuos en estudio puedan

colocarse en dos series ordenadas. La interpretación es similar a la r de Pearson (Figura 14.3).

$$r_s = 1 - \frac{6 \sum_{i=1}^N di^2}{N^3 - N}$$

Figura 14.3. Coeficiente de correlación de rango de Spearman r_s .

Dentro de la amplia gama de pruebas estadísticas paramétricas, existen dos que son las más utilizadas debido a su consistencia, sencillez y exigencia en el proceso. Una es la prueba t student y la otra es el Análisis de Varianza tipo I (ANOVA I), ambas pruebas se emplean para conocer si existe diferencia estadísticamente significativa entre las medias de los grupos.

t student

La prueba t student es un procedimiento práctico y muy poderoso que se usa ampliamente en las ciencias de la salud y en educación. Tiene dos modalidades, muestras dependientes y muestras independientes. Ambas se utilizan para evaluar si dos grupos de valores, medidos en intervalos o razón, difieren entre sí de manera significativa respecto a sus medias. Los valores pueden proceder de un mismo grupo en un diseño de investigación de preprueba–posprueba (Figura 14.4), o bien de dos grupos diferentes en un diseño de grupo experimental–grupo control (Figura 14.5).

$$t = \frac{\frac{\sum x}{n}}{\frac{S}{\sqrt{n}}}$$

Figura 14.4. Fórmula de prueba t student dependiente.

Fuente: elaboración propia.

$$t = \frac{X_2 - X_1}{\sqrt{\frac{S_2^2}{n_2} + \frac{S_1^2}{n_1}}}$$

Figura 14.5. Fórmula de prueba *t student* independiente.

Fuente: elaboración propia.

La prueba *t student* opera bien con tamaños de muestra de al menos 30 valores, incluso menos de 30, pero no es recomendable porque la probabilidad de que la muestra sea representativa de la población disminuye drásticamente; si supera los 120 valores la distribución *t* se aproxima a la distribución normal satisfactoriamente. La versión no paramétrica de la *t* es la *U* de Mann-Whitney.

Análisis de Varianza de tipo I

En los casos donde el investigador tiene más de dos grupos o variables es recomendable utilizar la prueba *F*, base del análisis de varianza de tipo I. El análisis de varianza de un factor (ANOVA I) es una prueba estadística para analizar si más de dos grupos difieren significativamente entre sí en cuanto a sus medias y varianzas. La comparación entre variables se realiza con una variable independiente de tipo categórica (nominal u ordinal) y una variable dependiente con un nivel de medición por intervalos o razón. Los cálculos se presentan en la tabla 14.2 del análisis de varianza.

El contenido de esta tabla concentra los valores en su cálculo manual y es la forma en que los programas de cómputo Excel y SPSS devuelven el resultado que se tiene que interpretar para tomar una decisión en la prueba de hipótesis. La versión no paramétrica es la prueba de ANOVA de Kruskal-Wallis.

Tabla 14.2. Tabla de concentrado de resultados del proceso de ANOVA 1

Fuentes de variación	Suma de cuadrados	Grados de libertad	Estimación (mean square)	F
Entre grupos	SCE	gl entre	SCE/gl entre	$\frac{\text{SCE/gl entre}}{\text{SCD/gl dentro}} = F$
Dentro de los grupos	SCD	gl dentro	SCD/gl dentro	
TOTAL				

Fuente: elaboración propia.

Prueba *chi* cuadrada X^2

Es una prueba estadística no paramétrica para evaluar hipótesis acerca de la relación entre dos o más variables categóricas. La hipótesis por probar es de tipo correlacional y analiza si las frecuencias observadas son diferentes de lo que pudiera esperarse en caso de ausencia de correlación. Esta cualidad de la *chi* cuadrada es una excelente alternativa para el investigador que busca establecer posibles relaciones entre variables que se dividen en varias categorías y, además, establecer cómo se distribuyen los sujetos en las posibles combinaciones de las celdas, ya que el contenido de cada celda es mutuamente excluyente, es decir, un mismo participante no puede estar en más de una casilla. El nivel de medición de la prueba es nominal u ordinal. Trabaja por medio de tablas de contingencia (Tabla 14.3).

Tabla 14.3. Tabla de contingencia de la X^2

		Variable B	
		B1	B2
Variable A	A1	A1-B1	A1-B2
	A2	A2-B1	A2-B2

Fuente: elaboración propia.

Conclusiones

Es oportuno recordar que la estadística es una herramienta que nos permite conocer, comprender y hacer inferencias acerca del comportamiento humano y de la salud. Como herramienta es necesario conocer ampliamente sus características para hacer un uso adecuado y realizar una interpretación correcta de la información reunida con los instrumentos de recolección de datos que a su vez deben cumplir con los requisitos de ser válidos y confiables.

Glosario

Confiability. La ausencia relativa de errores de medición en un instrumento de medición.

Correlación. Relación entre dos variables.

Cuestionario. Instrumento de recopilación de datos que operacionaliza determinados problemas que son objeto de investigación, le caracteriza una serie de preguntas.

Datos. Números o mediciones obtenidas como resultado de observaciones.

Distribución normal estándar. Distribución normal con media 0, desviación estándar 1 y área total igual a 1.00.

Error de tipo II (error tipo β). Ocurre cuando se acepta H_0 siendo realmente falsa. La probabilidad de un error de tipo II está dada por β .

Error tipo I (error de tipo α). Consiste en rechazar H_0 cuando realmente es verdadera. La probabilidad de cometer un error de tipo I está dada por el nivel α .

Estadística descriptiva. Técnicas que se utilizan para describir o caracterizar los datos muestrales obtenidos.

Estadística inferencial. Técnicas que utilizan los datos muestrales obtenidos para hacer inferencias sobre poblaciones.

Homocedasticidad. Propiedad que se refiere a que varias distribuciones tienen igual variabilidad (la misma varianza).

Homogeneidad de varianzas. Condición que existe cuando dos o más varianzas muestrales provienen de poblaciones con varianzas iguales.

Medición. Asignación de números a objetos o sucesos de acuerdo con conjunto de reglas predeterminadas o arbitrarias. Utilizan una unidad de medida aceptada por la comunidad científica o elaboran una adecuada al caso.

Nivel alfa (α). Nivel de significación establecida por el investigador para inferir la intervención de factores no aleatorios. Se asocia a la posibilidad de generalizar los resultados del análisis estadístico de la muestra a la población.

Nivel de significación. Valor de probabilidad considerado tan raro en la distribución muestral específica bajo la hipótesis nula, que nos inclina a aceptar la intervención de factores no aleatorios. Los niveles de significación comúnmente aceptados son .05 y .01.

Validez. Grado en que un instrumento mide lo que pretende medir.

Variable. Una característica o fenómeno que puede tomar diferentes valores.

Variabes cualitativas. Variables que difieren en cuanto a su clase, como ocurre con el sexo. Se cuentan no se miden.

Variabes cuantitativas. Variables que difieren en el "cuánto", como el caso de los niveles de inteligencia. Utilizan alguna unidad de medida existente.

Varianza. Suma de los cuadrados de las desviaciones, dividida por N.

Referencias

- Aiken, L. R. (2003). *Test psicológicos y evaluación* (11ª ed.). Pearson Educación.
- Aron, A. (2001). *Estadística para psicología* (2ª ed.). Pearson Educación.
- Bisquerra, R. (2009). *Metodología de la investigación educativa*. La Muralla.
- Cohen, R. J. y Swerdlik, M. E. (2007). *Pruebas y evaluación psicológicas. Introducción a las pruebas y a la medición*. McGraw Hill.
- Gómez, D. y Bowater, R. J. (2015). *Estadística y ciencia. Investigación cuantitativa en diversas disciplinas*. Fontarama.
- Kerlinger, F. y Lee, H. (2002). *Investigación del comportamiento. Métodos de investigación en ciencias sociales*. McGraw Hill.
- Pagano, R. (2006). *Estadística para las ciencias del comportamiento* (7ª. ed.). Thomson.
- Ritchev, F. (2008). *Estadística para las ciencias sociales* (2ª. ed.). McGraw Hill.
- Siegel, S. y Castellán, N. (2003). *Estadística no paramétrica aplicada a las ciencias de la conducta* (4ª. ed.). Trillas.
- Torres, J. (2019). *Estadística aplicada a ciencias de la salud*. Dextra.
- Zamora, S., Monroy, L. y Chávez, C. (2010). *Análisis factorial: una técnica para evaluar la dimensionalidad de las pruebas*. Centro Nacional de Evaluación para la Educación Superior.