



Archivos Venezolanos de Farmacología y
Terapéutica
ISSN: 0798-0264
revista.avft@gmail.com
Sociedad Venezolana de Farmacología Clínica y
Terapéutica
Venezuela

Nociones generales de muestreo aplicadas a las ciencias de la salud

Espinosa-Castro, Jhon-Franklin; Hernández-Lalinde, Juan; Bermúdez-Pirela, Valmore; Rodríguez, Johel E; Peñaloza-Tarazona, Mariana-Elena; Chacón, Gerardo; Toloza-Sierra, Cristian A.; Gómez-Vahos, Jovany

Nociones generales de muestreo aplicadas a las ciencias de la salud

Archivos Venezolanos de Farmacología y Terapéutica, vol. 37, núm. 5, 2018

Sociedad Venezolana de Farmacología Clínica y Terapéutica, Venezuela

Disponible en: <https://www.redalyc.org/articulo.oa?id=55963207003>

copyright © Sociedad Venezolana de Farmacología y de Farmacología Clínica y Terapéutica. Derechos reservados. Queda prohibida la reproducción total o parcial de todo el material contenido en la revista sin el consentimiento por escrito



Esta obra está bajo una Licencia Creative Commons Atribución-NoComercial-SinDerivar 4.0 Internacional.

Nociones generales de muestreo aplicadas a las ciencias de la salud

General notions of sampling applied to health sciences

Jhon-Franklin Espinosa-Castro
Universidad Simón Bolívar, Departamento de Ciencias
Sociales y Humanas, Cúcuta, Colombia., Colombia
 jhonfec1983@gmail.com

Redalyc: <https://www.redalyc.org/articulo.oa?id=55963207003>

 <http://orcid.org/0000-0003-2186-3000>

Juan Hernández-Lalinde
Universidad Simón Bolívar, Departamento de Ciencias
Sociales y Humanas, Cúcuta, Colombia., Colombia

 <http://orcid.org/0000-0001-6768-1873>

Valmore Bermúdez-Pirela
Universidad Simón Bolívar, Facultad de Ciencias de la
Salud, Cúcuta, Colombia., Colombia

 <http://orcid.org/0000-0003-1880-8887>

Johel E Rodríguez
Universidad Simón Bolívar, Facultad de Ingenierías,
Cúcuta, Colombia., Colombia

 <http://orcid.org/0000-0002-8353-2736>

Mariana-Elena Peñaloza-Tarazona
Universidad Simón Bolívar, Departamento de Ciencias
Sociales y Humanas, Cúcuta, Colombia., Colombia

 <http://orcid.org/0000-0002-3863-0580>

Gerardo Chacón
Universidad Simón Bolívar, Facultad de Ingenierías,
Cúcuta, Colombia., Colombia

 <http://orcid.org/0000-0003-3615-578>

Cristian A. Toloza-Sierra
Universidad Simón Bolívar, Departamento de Sistemas,
Cúcuta, Colombia., Colombia

 <http://orcid.org/0000-0002-9908-0076>

Jovany Gómez-Vahos
Universidad Simón Bolívar, Departamento de Ciencias
Sociales y Humanas, Cúcuta, Colombia., Colombia

 <http://orcid.org/0000-0003-1958-7107>

NOTAS DE AUTOR

jhonfec1983@gmail.com

RESUMEN:

El artículo tiene por objetivo describir nociones generales de muestreo aplicadas a las ciencias en el área de la salud, analizando definiciones, fórmulas, procedimientos y datos. En estudios probabilísticos, la población objeto de estudio está determinada por la muestra a través de un plan de muestreo. En un proceso estadístico son importantes las unidades primarias, secundarias y terciarias de muestreo. Igualmente, el tipo de muestreo puede ser aleatorio simple, sistemático estratificado o por conglomerados. Los siguientes criterios requieren claridad y fundamentación: tamaño de la población, nivel de confianza, error relativo, efecto del diseño, efecto del diseño ajustado, parámetro de interés, varianza estimada y tasa de no respuesta, permitiendo disminuir costos y tiempo, además de inducir resultados y hallazgos desde la muestra en función de la población. Para analizar un estadístico o parámetro desde un resultado inferencial, se requiere de un experto en matemáticas o estadística, con conocimiento en técnicas para recolectar, analizar e interpretar la información, razones que permiten al investigador una mayor validez y confiabilidad de los resultados obtenidos de la investigación.

PALABRAS CLAVE: población, muestreo probabilístico, dato, diseño, método, ciencias de la salud.

ABSTRACT:

The purpose of the paper is to describe sampling general notions applied to health sciences, analyzing definitions, formulas, procedures and data. In probabilistic studies, the sample determine the target population through a sampling plan. In all statistical process, the primary, secondary and tertiary sampling units are important. Likewise, the sampling type may involve techniques as simple random sampling, systematic sampling, stratified random sampling or cluster sampling. In order to develop a good sampling plan, the following criteria has to be clearly set: population size, confidence level, relative error, design effect, adjusted design effect, parameters, estimated variance and non-response rate. The achievement of these goals may reduce costs and time, besides to allow the inference of results and findings from the sample to population. The calculation of estimators and the inference made on parameters, require an expert in statistics or mathematics, with knowledge and abilities in recollecting, processing, analyzing and interpreting the information gathered. This features allow the researcher to gain a greater reliability and validity of the investigation results.

KEYWORDS: population, random sampling, data, design, method, health sciences.

INTRODUCCIÓN

La consecución de los objetivos que se persiguen al momento de aplicar una encuesta depende en gran medida del conocimiento que tengan los involucrados acerca del tema. Nociones como las de parámetro, estadístico, población objetivo, población muestreada, marco muestral, errores de muestreo, errores de no muestreo, sesgos de selección, estratos, conglomerados, entre otras tantas, son fundamentales y deben manejarse al menos en un nivel básico¹. En tal sentido, proponemos en primer lugar la realización de una inducción general in situ, en la que se discutan estos conceptos de manera exhaustiva y se aclaren los puntos de interés desde el comienzo.

Una vez esclarecidos los asuntos anteriores, estaríamos en capacidad para llevar a cabo reuniones de trabajo con cada equipo de investigación y definir los objetivos principales vinculados a cada disciplina. Para esto, consideramos vital la atención a las siguientes cuatro preguntas²:

1. ¿Cuál es la población objetivo sobre la que cada equipo de investigación desea hacer inferencias?
2. ¿Cuál sería la población muestreada sobre la que cada equipo de investigación podrá hacer inferencias?
3. ¿Cuáles son las variables que cada equipo de investigación pretende estudiar en dicha población?
4. ¿Cuáles son los parámetros que cada equipo de investigación desea estimar en cada una de las variables?

El aporte que cada grupo haga acerca de cuál es la población objetivo de su investigación será crucial para poder delimitar —a posteriori— la población objetivo de la encuesta. Recordemos que la intención es la de aprovechar el esfuerzo que implica el muestreo para recabar la mayor cantidad de información posible, información que luego será utilizada por las diferentes disciplinas en pro de materializar publicaciones.

Así pues, las respuestas a todas estas interrogantes no solo ofrecerán transparencia en torno al rumbo del estudio, sino que suministrarán los elementos necesarios para escoger los instrumentos de investigación y calcular el tamaño muestral del proyecto. Recuérdese que, en un estudio poblacional, un analista podría

interesarse en estimar el porcentaje de adultos jóvenes con ansiedad, mientras que otro podría estar interesado en conocer la presión arterial sistólica promedio de adultos mayores. Para cada una de estas situaciones habrá diversas escalas disponibles y en cada uno de estos contextos el análisis matemático es diferente, de modo que la combinación y correcta inclusión de estos intereses en el diseño del muestreo facilitará la obtención de resultados válidos para todas las partes. Es decir, la presencia del equipo de asesores metodológicos y estadísticos es imprescindible⁴.

Una vez que se han fijado los propósitos conceptuales del proyecto se procede a la puntualización de las condiciones técnicas necesarias para llevarlo a cabo. En este sentido, uno de los aspectos primordiales tiene que ver con los instrumentos de investigación que se usarán para medir las variables objetivo. Nada se gana con establecer una estructura de muestreo adecuada y rigurosa, si para la recolección de los datos se utilizan instrumentos cuya confiabilidad y validez no ha sido comprobada. Más aún, de nada sirve el entrenamiento que se le dé al personal que trabajará en el campo, si la herramienta de que disponen no tiene consistencia interna, validez de contenido, validez de constructo, entre otras^{3,4}.

En numerosas situaciones de investigación el analista emplea escalas cuyas propiedades han sido valoradas en el extranjero o en contextos distintos a los que se desean explorar. Aplicar el instrumento bajo estas circunstancias podría ocasionar errores de medición importantes, conduciendo a interpretaciones desviadas de la realidad. El muestreo piloto ofrece una solución idónea, no solo para este inconveniente, sino para aquel que se presenta cuando se desea calcular el tamaño muestral y se desconoce la varianza de la población. En este sentido, proponemos nuevamente realizar una inducción en la que se discutan estos tópicos y se sienten las bases necesarias para que la etapa de evaluación de los instrumentos pueda concretarse adecuadamente⁵.

Para la evaluación de los instrumentos de investigación persigue contestar a cuatro preguntas básicas⁶:

1. ¿Qué características tienen los instrumentos que pretenden utilizarse o diseñarse?
2. ¿Qué tipo de validez fue evaluada, qué técnica se empleó para esto y qué resultados se obtuvieron?
3. ¿Qué tipo de confiabilidad fue evaluada, qué técnica se empleó para esto y qué resultados se obtuvieron?
4. ¿La valoración de estas propiedades ha sido realizada en poblaciones similares a las que se pretenden estudiar?

Cuando nos referimos a las características del instrumento nos concentramos en aspectos como la longitud de la escala y el formato de los ítems, ya sea que se hayan planteado como preguntas abiertas o preguntas cerradas de opción simple o múltiple. Cuestiones tan sencillas como evitar las dobles negaciones pueden resultar vitales al momento de interactuar con el encuestado^{5,6}.

La concisión con que han sido esbozadas las afirmaciones, el hecho de que sea autoadministrado o no, la formulación del consentimiento informado y más, son asuntos que deben ser analizados en detalle por los equipos de investigación y los asesores estadísticos. Asimismo, la valoración de las propiedades psicométricas debe ser objeto de revisión profunda^{5,6}. Consideramos que la Escuela de Psicología puede brindar un apoyo invaluable en esta fase, aportando sugerencias sobre qué escalas utilizar que tengan reconocimiento internacional y local en términos de validez y fiabilidad. Ahondando en esto, creemos que no basta con saber cuál es la consistencia interna del test mediante el coeficiente alfa de Cronbach, por ejemplo; consideramos que es igualmente importante saber si se ha evaluado la consistencia temporal o si ha habido necesidad de emplear otros índices como los de KR-20, Theta u Omega. Por último, pero no menos importante, los reportes de validez deben ser examinados con el mismo rigor con que ha sido inspeccionada la fiabilidad^{7,8}.

DISEÑO DEL PLAN DE MUESTREO

La recopilación de información, es la fase tal vez la más importante en un estudio poblacional, el plan de muestreo que se elija podría dirigir el proyecto hacia un éxito categórico o hacia un rotundo fracaso, de manera que la información que el equipo de asesores utilice para diseñar la estrategia tendrá relevancia capital para

el logro de los objetivos. Al respecto, señalamos lo crucial que puede ser el investigar en las bases de datos de organismos públicos o privados, pero también en ir más allá y recopilar en las intendencias municipales o regionales documentos como el listado de calles o avenidas, el registro de comunas, mapas, entre otros. Disponer de todo esto facilitará el diseño del plan, llegado el momento en el que tengan que definirse las unidades primarias, secundarias y terciarias de muestreo^{9,10}.

En la mayoría de las situaciones reales los elementos de la población no están claramente definidos o no son fácilmente alcanzables, ya sea porque su distribución geográfica así lo determina o porque el marco muestral no existe¹¹. Considérese, por ejemplo, el caso en el que se investiga el bienestar psicológico de pacientes que asisten a consulta de cardiología luego de haber sufrido un infarto al miocardio. Disponer de una lista de dichos sujetos es prácticamente imposible, no así el hecho de contar con un registro de los centros de atención médica que pudieran servir como lugar para encontrar a tales individuos. En esta situación, se puede emplear dicho listado para elegir al azar varios de estos centros y luego encuestar y entrevistar a todos de los pacientes que cumplan con las condiciones de la investigación. A esta estrategia se le conoce como muestreo por conglomerados unietápico o monoetápico, ya que en la segunda etapa en la que se llega al elemento de interés no se realiza un muestreo sino un censo de tales unidades. Si, por el contrario, se hubiesen seleccionado aleatoriamente algunos de los pacientes de cada uno de los sanatorios previamente escogidos, el procedimiento se habría cumplido en ambas fases y estaríamos en presencia de un muestreo por conglomerados bietápico. Obviamente este proceso podría extenderse tantas veces como lo requieran las circunstancias del estudio, dando origen a técnicas polietápicas o multietápicas^{12,13-16}.

En el contexto anterior, los hospitales forman conjuntos naturales denominados conglomerados, situación que podría repetirse con escuelas, oficinas, salones de clase, barrios, comunas, entre otros. Un trabajador social podría interesarse en conocer las relaciones familiares de niños en edad escolar. Para esto, obtiene el listado de todas las escuelas localizadas en la región de interés clasificándolas como conglomerados de primer orden. En el argot estadístico, a estos grupos se les conoce como unidades primarias de muestreo (UPM). Supongamos ahora que el trabajador social elige cinco escuelas, utilizando para esto cualquiera de las estrategias aleatorias de selección y posteriormente se dirige a los centros que resultaron escogidos para obtener la información que se requerirá en la siguiente etapa. El investigador podría registrar el número de cursos de cada escuela y la cantidad de alumnos por curso, construyendo de esa manera el marco muestral de la segunda fase. Nótese que las unidades secundarias de muestreo (USM) están conformadas por los cursos, de modo que la escogencia al azar de una muestra de estos sería necesaria para acceder finalmente a los estudiantes. Para concluir, imaginémonos que el investigador selecciona de nuevo cinco cursos al azar de cada uno de los colegios y que, una vez en estos, selecciona aleatoriamente cinco alumnos de cada salón. En esta última fase, las unidades terciarias de muestreo (UTM) son en sí mismas los elementos de la población, representadas evidentemente por los niños en edad escolar. La muestra final estaría integrada por $5 \times 5 \times 5 = 125$ sujetos y se clasificaría como aleatoria polietápica por conglomerados debido a: (i) la selección de las unidades se ha realizado al azar en cada etapa, (ii) ha habido más de dos etapas de muestreo y (iii) se han empleado conglomerados para consumir la selección^{17,18}.

Los argumentos previos se han planteado con la finalidad de exponer situaciones semejantes a las que se afrontarán en cualquier contexto. A pesar de que a estas alturas del proyecto es irresponsable dar una versión final y definitiva del diseño muestral, nos atrevemos a indicar que —con una alta probabilidad— se necesitará un muestreo aleatorio, estratificado, polietápico y por conglomerados. La variable de estratificación debería fijarse según cuestiones geográficas, aunque podría plantearse a partir de la diferenciación de zonas urbanas y rurales. Los conglomerados podrían estar representados por comunas, bloques, calles u hogares, esto dependerá de aspectos logísticos, estadísticos y económicos. El tamaño de la muestra se calculará con base en los siguientes criterios: (i) tamaño de la población objetivo, (ii) coeficiente de confianza deseado, (iii) error máximo admisible expresado como una fracción del parámetro de interés, (iv) efecto del diseño (*deff*),

(.) efecto del diseño ajustado para correlación intraclase, (vi) parámetro de interés, (vii) varianza estimada de la población y (viii) tasa de no respuesta (TNR)¹⁸⁻²⁰.

Para ilustrar esto, consideremos el siguiente escenario hipotético: se desea estimar la proporción de votantes de un municipio que elegirían al alcalde como gobernador. Asumamos que se exige un nivel de confianza de 95% y un error máximo admisible de 4% (error relativo). Puesto que se ha utilizado una estrategia basada en muestras complejas, es necesario introducir el *deff* en la operación de cálculo. Aceptemos por ahora que este factor puede oscilar desde 1.5 hasta 3.0 si no se ajusta según el número de hogares por sector y la correlación intrahogares. La TNR ante esta interrogante puede ser elevada, fijémosla en 25% y utilicemos el criterio de máxima indeterminación ($pq=0.25$) debido a que se desconoce la varianza de la población y el posible valor del parámetro p . Por último, la población adapta para votar del municipio es de 260125 electores.

Teniendo en cuenta los anteriores datos, El tamaño de la muestra vendría dado por la ecuación inicial para poblaciones infinitas, tal que^{11,13,16}:

$$n_0 = \left(\frac{z_{\alpha/2}}{rp} \right)^2 pq.$$

(Ec. 1)

$$n_0 = \left(\frac{1.96}{0.04 \times 0.50} \right)^2 0.25 = 2401.$$

Haciendo los ajustes correspondientes:

$$n_0 = 2401 \times deff \times (1 + TNR) = 2401 \times 2 \times 1.25 = 6003.$$

Puesto que $N=260125$, podría aplicarse la corrección de población finita aunque de antemano sepamos que tendrá poco efecto debido a que la fracción de muestreo ni siquiera alcanza el 3%. Sin embargo:

$$n = \frac{n_0}{1 + \frac{n_0}{N}}.$$

(Ec. 2)

$$n = \frac{6003}{1 + 0.023} = 5868.$$

Reemplazando en (Ec. 2):

Pensemos ahora que el grupo de asesores quiere tener una estimación más realista del *deff*. Se puede emplear el número de hogares promedio que serán visitados por sector y una aproximación del coeficiente de correlación intraclase para obtenerla. Así:

$$deff \cong 1 + (n-1)r_c. \quad (\text{Ec. 3})$$

Asumiendo que se visiten 30 hogares por sector y una correlación dentro del hogar de 0.10 (baja), se obtiene: $deff \cong 1 + (30-1)0.10 = 3.90$.

Si ahora se sustituye este valor en (Ec. 1) y luego este resultado se introduce en (Ec. 2), el tamaño muestral pasaría a ser de:

$$n = \frac{11\,705}{1 + 0.045} = 11\,201.$$

Como puede apreciarse, el diseño del plan de muestreo es una actividad compleja y dinámica que involucra procesos de iteración tendentes a equilibrar las necesidades estadísticas y presupuestarias considerando además la viabilidad técnica del proyecto. Desarrollar esto a distancia, manteniendo los niveles de calidad exigidos para una propuesta como esta, es prácticamente imposible.

Estimación de los recursos

Si el diseño de la estrategia de muestreo y el cálculo del tamaño de la muestra son puntos esenciales del proyecto, el realizar una estimación concienzuda y afinada de los recursos necesarios para emprenderlo. No es poco frecuente que, llegados a esta etapa, la propuesta deba ser aplazada o suspendida por insuficiencia de fondos, material, personal o tiempo. La experiencia indica que quienes solicitan este tipo de trabajos subestiman de manera considerable el esfuerzo humano y monetario, es tarea del asesor hacérselos ver, delimitando las expectativas del solicitante desde el comienzo, no con la intención de desanimarlo sino por el contrario, con el propósito de orientarlo y ofrecerle la opción que minimice los costos y maximice las ganancias que pretende obtener, sean o no de orden material²¹.

Consideraciones TNR

La pregunta entonces sería, ¿qué se puede hacer para evitar la no respuesta? Si esta se pretende eliminar por completo, lo única opción es no realizar la encuesta. Siempre habrá una cantidad de personas que no contestarán, una fracción de hogares a los que no se podrá llegar o un porcentaje de casas en las cuales no habrá individuos al momento del abordaje. Sin embargo, sí puede haber una reducción notable de este factor, en especial cuando el grupo de investigación se prepara para disminuir estas circunstancias²².

Esta estrategia debería ir acompañada de campañas de publicidad, por lo menos por un lapso de tres semanas previas a la ejecución del muestreo final. El hecho de que los encargados de conducir el sondeo aparezcan en programas de radio y televisión, que además se publiquen artículos en medios impresos o virtuales, hará que la sociedad sea mucho más permeable al proyecto y que la participación sea elevada. Pueden ofrecerse charlas divulgativas a las que asistan diferentes representantes y en las que se expongan las bondades de la investigación. Estos, a su vez, servirán como factor multiplicador entre sus conocidos y ayudarán a expandir la red de influencia del estudio. A modo de cierre, nótese que la estimación de los recursos implica la consideración exhaustiva de diversos aspectos técnicos, para lo cual es necesario que exista una comunicación estrecha entre las autoridades de la organización y los asesores del proyecto²².

Reclutamiento y adiestramiento

El perfil de la persona que aspira a convertirse en encuestador debe ser definido con cierta rigurosidad, además de velar porque el individuo posea las aptitudes y actitudes necesarias, no olvidemos que ellos son el último eslabón de la cadena que une a la institución con la persona seleccionada en el muestreo, de modo que de la interacción que haya entre ambos, depende mucho la calidad de los resultados. La participación de los asesores será básica, sobre todo en las labores de adiestramiento en tópicos metodológicos y estadísticos. Esto no solo servirá como entrenamiento real para los encuestadores, sino que permitirá obtener estimaciones de los parámetros de interés y de la TNR, valoraciones que podrían ser usadas para ajustar el tamaño muestral calculado^{22,23}.

Muestreo final

Esta es la etapa operativa, la que mayor exigencia tiene en términos logísticos. La vigilancia constante de las actividades será decisiva para el logro de las metas planteadas al inicio, de manera que la comunicación entre los líderes de la investigación, supervisores de campo y encuestadores será permanente mientras dure la experiencia. Cualquier inconveniente deberá ser atendido con prontitud; cualquier duda, contestada con celeridad. Es por ello que el entrenamiento y sensibilización allanarán el camino para minimizar los imprevistos que habrán de suscitarse durante el muestreo. Ningún equipo de encuestadores y supervisores podrá comenzar su labor sin antes tener claro lo que tiene que hacer y cómo hacerlo. Deberá establecerse un comando que monitoree el trabajo y las operaciones²⁴.

Por otra parte, sugerimos que el muestreo se vaya acometiendo por tandas, las cuales pueden ejecutarse en plazos de una semana e ir haciendo revisiones con la finalidad de ajustar la estrategia. se espera que el tiempo de interacción sea mucho mayor, en especial si se incorpora la investigación relacionada con las ciencias de salud, en la cual habrá que auscultar al sujeto y tomar muestras de sangre, además de medir variables como la presión arterial, frecuencia cardíaca, circunferencia abdominal, talla, actividad física, hábitos de tabaquismo, consumo de alcohol^{25,26}, entre otras. Situaciones similares pueden presentarse al aplicar instrumentos psicológicos. En este caso, la experiencia nos señala que la utilización de escalas específicas suele demandar desde 20 hasta 30 minutos, mientras que la administración de test generalizados puede tomar varias horas²⁴.

Las deliberaciones anteriores nos permiten concluir que el número de encuestadores dependerá de lo siguiente: (i) del tiempo de administración del instrumento más los suplementos de rigor, (ii) del número de días en que se desea terminar la encuesta, (iii) de las limitaciones presupuestarias y (iv) de la complejidad de la interacción exigida por la investigación. A tal efecto, es necesario obtener una estimación del tiempo estándar para cada una de las disciplinas que participarán. Nuevamente cobra relevancia el que se realice el muestreo piloto además de simulacros, actividades de las cuales se podrán extraer las cifras necesarias para la valoración²⁷⁻²⁹.

Por último —y retomando lo de las tandas de muestreo— resultará vital que las jornadas iniciales se dediquen a explorar únicamente los aspectos sociales, demográficos y económicos del hogar que resultase seleccionado. Una de las razones de tal aseveración estriba en que este tipo de encuesta es poco invasiva; constituye una forma fresca, amena y rápida de abordar a la familia en las primeras instancias del estudio y permite que el encuestador se gane al núcleo familiar. Otro de los motivos para hacer tal recomendación, es que en este tipo de indagaciones se recolectan datos que luego pueden ser cotejados con las estadísticas proyectadas del censo y así verificar qué tan similares son, constituyéndose en una herramienta de control para el muestreo²⁷⁻²⁹.

MANEJO DE LA NO RESPUESTA

El fenómeno de la no respuesta es, en la misma medida, complejo e importante, en especial si se pretende que las estimaciones obtenidas de la encuesta verdaderamente representen al universo estudiado. Para comprenderlo mejor, concibamos a la población como un conjunto formado por dos estratos: (1) los que fueron seleccionados en la muestra y contestarán y (2) los que fueron seleccionados y no contestarán. En una situación real se desconocerán las proporciones de cada grupo hasta tanto no finalicen las tandas de muestreo, por ello resulta vital que los encuestadores registren la no respuesta minuciosamente, indicando la cantidad de unidades seleccionadas que no aportaron información y especificando las razones de esto^{22,30}.

Para entenderlo mejor, se plantea el siguiente ejemplo: todas las mujeres que habitan en la Comuna 6 de Cúcuta, que estén o hayan estado involucradas en relaciones de pareja heterosexuales y con edades comprendidas desde los 18 años en adelante. Supongamos además que el investigador conduce una encuesta basada en entrevistas personales realizadas en los hogares seleccionados mediante muestreo complejo y que obtuvo una TNR de 43% repartida entre las mujeres que no pudieron localizarse, las que estaban incapacitadas para contestar o simplemente aquellas que se rehusaron a participar. En consecuencia, la población muestreada está constituida por todas las mujeres de la Comuna 6 de Cúcuta, que tuvieran o hubieran tenido relaciones de pareja heterosexuales al momento de la encuesta, con edades comprendidas desde los 18 años en adelante y que hayan estado dispuestas a participar en el sondeo. Lo último es esencial para entender la importancia de la no respuesta: con un porcentaje tan elevado como 43%, difícilmente se pueden hacer extrapolaciones consistentes a la población objetivo y las inferencias válidas deberían limitarse únicamente a la población muestreada. Hay razones suficientes como para sospechar que la percepción de

las mujeres que quisieron contestar es significativamente diferente de quienes se negaron, de modo que si las estimaciones se hacen sin realizar ajustes, bien podría sobreestimarse o subestimarse el parámetro de interés.

Análisis: la población objetivo está constituida por todas las mujeres de la Comuna 6 de Cúcuta, involucradas en relaciones de pareja heterosexuales y con edades comprendidas desde los 18 años en adelante. Supongamos que se realizó un muestreo aleatorio simple sin reposición, que el tamaño de la muestra fue de 2203 mujeres, necesario para estimar la puntuación promedio de abuso percibido (Y) de una escala psicométrica estandarizada, con un error relativo del 1% bajo el supuesto de que $Y \sim N(50, 156.25)$, un nivel de confianza de 95% y ajuste para una población finita de aproximadamente $N=27400$ individuos. Por último, admitamos que la TNR fue la que se reportó previamente y que alcanzó cifras de 43%. La tabla 1, exhibe los resultados de esta investigación simulada.

Resultados del ejemplo hipotético

Resultados del ejemplo hipotético

Estratos	n	Y
1: Mujeres de la población objetivo que fueron seleccionadas y contestaron (57%)	1256	48,57
2: Mujeres de la población objetivo que fueron seleccionadas y no contestaron (43%)	947	Desconocido
Total	2203	Desconocido

Los resultados anteriores evidencian el tremendo impacto que puede tener el ignorar la no respuesta y estimar el parámetro asumiendo que los que contestaron tiene el mismo comportamiento que los que no lo hicieron. Ciertamente, la no respuesta tendría un efecto despreciable solo si su proporción es baja o si la diferencia entre los grupos también es baja (o ambas). El problema radica en que la diferencia de medias es desconocida, incluso luego de haber terminado el muestreo, de manera que lo que se puede hacer es estimar la TNR y ejercer algún tipo de acción preventiva desde el inicio. Sustentemos esto con un poco de matemática: la media global del abuso experimentado por las mujeres sería igual al promedio ponderado de la media de quienes contestaron y de quienes no lo hicieron. La ponderación viene dada por el peso de cada estrato, es decir, por las tasas de respuesta y no respuesta.

Así:

$$Y = w_1 y_1 + w_2 y_2.$$

(Ec. 4)

Supongamos que la opinión en ambas subpoblaciones es la misma, esto es, . Además, las subpoblaciones son complementarias, de modo que $w_1 + w_2 = 1$

En consecuencia:

$$Y = y_{12}(w_1 + w_2) = y_{12}.$$

(Ec. 5)

Lo que se demuestra en (Ec. 5) es que si el nivel de la variable de interés es el mismo en cada estrato, la TNR pierde importancia y podría ignorarse. Lamentablemente esto no suele suceder, de hecho, numerosas investigaciones señalan que los grupos que suministran información tienden a ser significativamente distintos de los que no. Veamos esto desde el siguiente enfoque: supongamos que las mujeres a las que no se pudo acceder puntúan dos veces más alto en términos del abuso que perciben de sus parejas, es decir,

Desarrollando (Ec. 4):

$$\begin{aligned}
 Y &= w_1 y_1 + 2w_2 y_1 = y_1(w_1 + 2w_2), \\
 Y &= w_1 y_1 + 2w_2 y_1 = y_1(1 - w_2 + 2w_2), \\
 Y &= w_1 y_1 + 2w_2 y_1 = y_1(1 + w_2).
 \end{aligned}$$

(Ec. 6)

Obsérvese en (Ec. 6) que si todas las mujeres seleccionadas en el muestreo fuesen entrevistadas efectivamente, $\lim_{y_2 \rightarrow 0} Y = y_1$. En el caso extremo, si nadie contestase, $\lim_{y_2 \rightarrow 1} Y = 2y_1$. Esto implicaría que el puntaje promedio de abuso estaría subestimándose y se reportarían cifras cercanas a la mitad del valor real. Si la situación fuera la esbozada en ejemplo hipotético, la media estará condicionada a una TNR=0.43 y a una relación de $y_2 = 2y_1$, de modo que su valor sería de aproximadamente: $Y = (48.57)(0.57) + (48.57 \times 2)(0.43) = 69.46$ puntos, y no de 48.57 como se refleja en la tabla 1.

Con base en los comentarios previos sugerimos que la no respuesta final sea abordada mediante dos maniobras principales: (i) insistir en que se concrete la entrevista, visitando en repetidas ocasiones los hogares seleccionados que no aportaron información y (ii) construir modelos probabilísticos que permitan predecir el comportamiento de las variables de interés en el estrato desconocido. En cuanto a la primera estrategia, diversos estudios indican que pueden llevarse a cabo hasta cuatro intentos, mientras que otros señalan cantidades mayores, que suelen oscilar desde 10 hasta 20 solicitudes. Creemos que esto debe manejarse con cuidado, teniendo la previsión de no saturar al sujeto escogido y de no generar emociones de rechazo hacia el encuestador.

La cantidad de intentos en cuestión deberá fijarse meticulosamente, tomando en cuenta la idiosincrasia, costumbres y afectos de la población objetivo, así como las referencias de entes especializados como el DANE, INE, IADB, ONU, entre otros. En lo que respecta a los modelos de predicción, la teoría estadística del muestreo pone a disposición varias opciones, entre las que destacan los métodos de imputación de datos, regresión, muestreo en dos fases y métodos de fijación de pesos para la no respuesta.

LA ORGANIZACIÓN Y TABULACIÓN

Llegados a este punto nos enfrentamos con la ardua labor de ordenar todos los papeles que se hayan utilizado en la encuesta y que hayan servido para vaciar la información captada. Hablamos de una tarea sencilla en términos de dificultad, pero laboriosa y primordial. El volumen de datos que se maneja en este tipo de actividades es considerable y los errores de transcripción suelen ser frecuentes debido a factores como el cansancio, rutina y desatención. De manera pues que la vigilancia y seguimiento que se le dé a esta fase también resultará decisiva para el logro de los objetivos. En tal sentido, aconsejamos que los trabajos de tabulación sean precedidos por una faena meticulosa de organización, en la que se revise la integridad física de todos los documentos y se separen en grupos cuyas cantidades sean fáciles de manejar. Dichas actividades podrían ser ejecutadas por los mismos encuestadores y verificadas por los supervisores de campo, de manera que no sean necesarias otras contrataciones de personal. Posteriormente, el vaciado de los datos debería repartirse entre los equipos de tabuladores que se conformen, los cuales a su vez deberían tener asignado un verificador por grupo. Si actuamos con cautela desde el inicio y esparcimos la auditoría en varias etapas, se minimizará la probabilidad de encontrar errores al momento del análisis estadístico. Enfatizamos de nuevo: valdría la pena tantear la posibilidad de que el mismo personal que trabajó en el campo sea el encargado de preparar los papeles y copiar los resultados a los programas respectivos. Tal vez esto ayude a disminuir costos, aunque

evidentemente tendrá que ser escrutado en función de la confidencialidad que quiera dársele a la información ^{31,32}.

Luego de culminada esta etapa se daría inicio a la auditoría y al análisis exploratorio de datos mediante técnicas estadísticas. Este paso es tan importante como el anterior e involucrará la utilización de numerosos procedimientos, así como de muchas horas de trabajo frente al computador. Para ello, habremos de apoyarnos en el uso de programas como SPSS, Stata, SAS, R y EpiDat, empleando técnicas básicas de estadística descriptiva, pero también metodologías avanzadas como pruebas de bondad de ajuste, pruebas de independencia, contrastes de homocedasticidad, pruebas de normalidad univariadas y multivariadas, análisis factorial, análisis de conglomerados, análisis discriminante, entre otros. Además, en este punto del proyecto deberían implementarse los modelos para el ajuste de la no respuesta que ya se señalaron ³³⁻³⁵.

Análisis de los resultados

La estadística inferencial paramétrica se sustenta en varias conjeturas, una de las cuales consiste en asumir que la muestra se ha obtenido mediante un muestreo aleatorio simple (MAS). La teoría de esta rama de la estadística exige que las variables aleatorias que conforman dicho conjunto sean independientes entre sí e idénticamente distribuidas (*vaiid*), esto es, que cada variable tenga la misma distribución de probabilidad y que todas sean mutuamente independientes ¹⁷⁻¹⁹.

Consideremos el siguiente ejemplo: un ingeniero industrial realiza una investigación en una línea de producción de cervezas para lo cual selecciona 50 botellas al azar. Supongamos que la variable de interés es el volumen de llenado, de manera que cada unidad seleccionada se somete a la medición de esta característica para luego ser desechada. Puesto que la embotelladora de cervezas trabaja de manera continua, la extracción de 50 unidades no tiene efecto en la probabilidad de encontrar una botella defectuosa, es decir; la probabilidad de éxito p permanece constante en cada ensayo y estos son independientes. Además, en vista de que cada envase es extraído de la misma línea, la cual a su vez tiene los mismos parámetros de operación, el volumen de llenado de cada unidad tendrá el mismo comportamiento probabilístico. En otras palabras, la muestra así seleccionada constituye un conjunto de *vaiid*. El ingeniero ahora se enfoca en estimar el volumen de llenado (Y) mediante un intervalo de confianza bilateral (ICB) del 95%, partiendo de la suposición de que la media y la desviación estándar fueron de $y = 354.13$ y $s = 10.58$ mL, respectivamente. Puesto que el tamaño muestral es grande, aplican las condiciones del teorema de límite central y la ecuación para construir el ICB sería:

$$y \pm z_{\alpha/2} s / \sqrt{n}.$$

(Ec. 7)

Reemplazando en (Ec. 7):

$$354.13 \pm (1.96)(10.58/\sqrt{50}),$$

$$354.13 \pm 2.93,$$

$$[351.20, 357.06].$$

El resultado anterior puede interpretarse desde el enfoque clásico y frecuentista. Desde la perspectiva clásica, podría afirmarse con un 95% de confianza que el volumen de llenado de cerveza oscila desde 351.20 hasta 357.06 mL. Por otro lado, desde el punto de vista frecuentista, si el ingeniero repitiese una y otra vez el experimento y construyese en cada repetición el intervalo de confianza correspondiente, cabría esperar que 95% de ellos contengan el verdadero valor del parámetro.

Para ilustrar la importancia de adecuar el modelo inferencial a la técnica de muestreo utilizada, consideremos ahora que el ingeniero seleccionó las 50 unidades sin reposición pero de un contenedor con 75 cervezas. Nótese que en esta ocasión la población no puede ser tratada como infinita y que, además, la escogencia de cada unidad sí afectará la probabilidad de que el siguiente envase tenga o no determinada condición, de modo que los ensayos no deben concebirse como independientes y estaríamos en presencia de un muestreo aleatorio simple sin reposición (MASSR).

La ecuación para la estimación por intervalos en este caso es la siguiente:

$$y \pm z_{\alpha/2} \left(\sqrt{1 - \frac{n}{N}} \right) \left(\frac{s}{\sqrt{n}} \right),$$

(Ec. 8)

dónde $\sqrt{1 - n/N}$ es conocido como factor de corrección de población finita (fpc). Sustituyendo en (Ec. 8):

$$\begin{aligned} & 354.13 \pm (1.96) \left(\sqrt{1 - 50/75} \right) (10.58/\sqrt{50}), \\ & 354.13 \pm 1.70, \\ & [352.43, 355.83]. \end{aligned}$$

Como puede apreciarse, el intervalo de confianza obtenido en esta oportunidad tiene un ancho de $355.83 - 352.43 = 3.40$ mL, en tanto que el anterior exhibe una amplitud de $357.06 - 351.20 = 5.86$ mL. Esta reducción de 2.46 unidades equivale a un aumento en precisión de 58%.

Para continuar exponiendo la trascendencia de usar la técnica correcta en el análisis de resultados, ahondemos un poco más e imaginémonos otro escenario. Digamos que la Escuela de Psicología de la Universidad Simón Bolívar desea conocer las capacidades cognitivas de sus estudiantes, para lo cual emplea un test que mide dicha variable a partir de cuatro dimensiones: orientación, atención, memoria y lenguaje. Los psicólogos que conducen la investigación tienen razones para creer que dichas habilidades pueden diferir de una carrera a otra. Probablemente los alumnos de ingeniería exhiban funciones cognitivas más elevadas en el área de la memoria, mientras que los de comunicación social manifiesten mayores destrezas en el plano del lenguaje. Si el asesor estadístico efectúa un MAS, podría suceder que, por azar, algunos estratos de la población no estén adecuadamente representados en la muestra y que, por ende, haya una estimación incorrecta del parámetro. Ante este contexto, lo recomendable sería efectuar un muestreo aleatorio estratificado (MAE), pudiendo definir como variable de estratificación la carrera que cursa el estudiante. Solo para efectos informativos, la ecuación para construir un ICB bajo este esquema sería¹⁷⁻¹⁹:

$$y \pm z_{\alpha/2} \sqrt{\sum_{h=1}^H \left(1 - \frac{n_h}{N_h} \right) \left(\frac{N_h}{N} \right)^2 \left(\frac{s_h^2}{n_h} \right)},$$

(Ec. 9)

dónde N_h es el tamaño del h-ésimo estrato, n_h es el tamaño de la muestra correspondiente al h-ésimo estrato, s_h^2 es la varianza del h-ésimo estrato y N es el tamaño de la población.

Todas las reflexiones esbozadas en este apartado sirven para destacar la importancia de utilizar la técnica estadística compatible con el diseño. Nos hemos permitido abusar de la matemática para aclarar este asunto

de manera profunda. Al momento de realizar el análisis, por supuesto, nos valdremos de los módulos especializados que paquetes como SPSS, SAS, Stata y R tienen para trabajar con datos extraídos de muestras complejas. Cobrará vital relevancia la inclusión de los pesos, factores de expansión, probabilidades de inclusión y cualquier otro elemento de ajuste que sea necesario para lograr estimaciones insesgadas³³⁻³⁵.

Como la investigación es un conjunto de procesos sistemáticos, críticos y empíricos que se aplican al estudio de un fenómeno o problema³⁶. Si iniciamos de que el método científico es uno, existen diversas formas de identificar su práctica o aplicación, de modo que la investigación se puede clasificar de diversas maneras. Enfoques positivistas promueven la investigación empírica con un alto grado de objetividad suponiendo que, si alguna cosa existe, existe en alguna cantidad y su existe en alguna cantidad se puede medir³⁷.

Para Zorrilla (1993, p. 43) citado por Grajales, uno de estos tipos de investigación es la aplicada, porque guarda íntima relación con la básica, pues depende de los descubrimientos y avances de la investigación básica y se enriquece con ellos, pero se caracteriza por su interés en la aplicación, utilización y consecuencias prácticas de los conocimientos, la investigación aplicada busca el conocer para hacer, para actuar, para construir, para modificar³⁷.

Es decir, la investigación aplicada o tecnológica, está orientada a resolver los problemas que se presentan en los procesos de producción, distribución, circulación, consumo de bienes y servicios de cualquier actividad humana. Se denomina aplicadas, porque en base a investigación básica, pura o fundamental en las ciencias fácticas o formales se formulan problemas o hipótesis de trabajo para resolver los problemas de la vida productiva de la sociedad. Se llama también tecnológico, porque su producto no es un conocimiento puro, sino tecnológico. Este tipo de investigaciones están orientadas a mejorar, perfeccionar u optimizar el funcionamiento de los sistemas, los procedimientos, normas, reglas tecnológicas actuales a la luz de los avances de la ciencia y la tecnología; por tanto, este tipo de investigación no se presta a la calificación de verdadero, falso o probable sino a la de eficiente, deficiente, ineficiente, eficaz o ineficaz, según Ñaupas, 2013, citado por Teodoro y Nieto, 2018^{38,39}.

La investigación se fundamenta en un paradigma tipo positivista, es decir, percibe la uniformidad de los fenómenos, aplica la concepción hipotética-deductiva como una forma de acotación y predica que la materialización del dato es el resultado de procesos derivados de la experiencia. Esta concepción se organiza sobre la base de procesos de operacionalización que permiten descomponer el todo en sus partes e integrar éstas para lograr el todo⁴⁰.

Teniendo en cuenta los siguientes argumentos, se utiliza un enfoque cuantitativo, utiliza la recolección de datos para probar hipótesis con base en la medición numérica y el análisis estadístico, con el fin de establecer pautas de comportamiento y probar teorías. Permitiendo trabajar desde un alcance exploratorio, descriptivo, correccional y explicativo^{40,43-48}.

El diseño de la investigación puede ser experimental (Preexperimentos, Cuasiexperimentos y Experimentos “puros”) igualmente o no experimental (Longitudinales o evolutivos y Transeccionales o transversales)⁴³⁻⁴⁸, porque el diseño es un plan o estrategia que se desarrolla para obtener la información que se requiere en una investigación y responder al planteamiento³⁶.

Para la presentación del escrito, se realizó una revisión documental que consistió en detectar, consultar, obtener la bibliografía y otros materiales útiles para los propósitos del estudio, de los cuales se extrae, recopila información relevante y necesaria para la investigación (Hernández, Fernández y Batista., 2014, p. 61). Es decir, una revisión documental de cualquier clase,..., se concreta exclusivamente en la recopilación de información en diversas fuentes. Indaga sobre un tema en documentos -escritos u orales-⁴¹.

Dependiendo de la población seleccionada, debe aplicarse un muestro probabilístico, seleccionado un subgrupo específico con característica comunes en función de la población, llamado muestra. Es decir, todos

los elementos tienen la misma posibilidad de ser elegidos. (Hernández, Fernández y Batista., 2014, p. 175). Fundamentado en un muestreo aleatorio simple, estratificado, sistematizado y conglomerado¹².

La técnica para recolectar información, puede aplicarse la encuesta, técnica que utiliza un conjunto de procedimientos estandarizados de investigación mediante los cuales se recoge y analiza una serie de datos de una muestra de casos representativa de una población o universo más amplio, del que se pretende explorar, describir, predecir y/o explicar una serie de características, según García, 1993 citado por Anguita, 2003⁴².

El análisis e interpretación de información, puede utilizar Excel y/o un software estadístico⁴³, para tabular, procesar y presentar a través de textos, cuadros, tablas de distribución de frecuencias, y gráficos, para su divulgación e interpretación, en estudios con alcance exploratorio, descriptivo, correccional y explicativo⁴⁴⁻⁴⁹, donde se ajusta el análisis de resultados a los contextos de salud⁵⁰ y otras técnicas de utilidad en este tipo de análisis⁵¹.

REFERENCIAS

1. Alcalá Velásquez OJ. *Tópicos de estadística aplicados a las ciencias sociales*. (1. Ed.). Caracas, Venezuela: Universidad Católica Andrés Bello; 2012.
2. Bulia Melo EE. *Estadísticas para las ciencias económicas, administrativas y contables*. (1. Ed.). Cúcuta, Colombia. Universidad Libre Seccional Cúcuta; 2014.
3. Quezada Lucio N. *Metodología de la investigación*. Estadística aplicada en la investigación. (1. Ed.). Lima, Perú: Empresa Editora Macro E.I.R.L; 2010.
4. González Raposo MS. *Métodos, diseños y técnicas de investigación social*. (1. Ed.). Mérida, Venezuela: Editorial Venezolana C.A; 2016.
5. Freund JE, Simon GA. *Estadística elemental*. (8. Ed.). México D.F.: Pearson Educación; 1992.
6. Mason RD, Lind DA, Marchal WG. *Estadística para administración y economía*. (11. Ed.). México D.F.: Alfaomega Grupo Editor S.A; 2004.
7. Webster AL. *Estadística aplicada a los negocios y la economía*. (3. Ed.). Bogotá, Colombia: McGraw Hill; 2000.
8. Anderson DR, Sweeney DJ, Williams TA. *Estadística para administración y economía*, (10. Ed.). México, D.F: Cengage Learning Editores, S.A; 2008.
9. Sierra Hernández JJ. *Inferencia Estadística. (Vol. 1)*. Bogotá. D.C, Colombia: Escuela de Ciencias Básicas, Tecnología e Ingeniería. Universidad Nacional Abierta y a Distancia; 2010.
10. Sierra Hernández JJ. *Inferencia Estadística. (Vol. 2)*. Bogotá. D.C., Colombia. Escuela de Ciencias Básicas, Tecnología e Ingeniería. Universidad Nacional Abierta y a Distancia; 2010.
11. Martínez Bencardino C. *Estadística y muestreo*. (13ª Ed.). Bogotá, Colombia: Ecoe Ediciones; 2010.
12. Casal J, Mateu E. *Tipos de muestreo*. Revista. Epidem. Med. Prev. 2003; .(1), pp. 3-7.
13. Azorín Poch F. *Curso de muestreo y aplicaciones*. (1. Ed.). Bilbao, España: Editorial Aguilar; 1969.
14. Milton JS. *Estadística para biología y ciencias de la salud*. (1. Ed.). Madrid, España: McGraw Hill — Interamericana; 1999.
15. Martínez Bencardino C. *Estadística Básica Aplicada*. (1ª Ed.). Bogotá, Colombia: Ecoe Ediciones; 2012.
16. Martínez Bencardino C. *Estadística y Muestreo*. (1ª Ed.). Bogotá, Colombia: Ecoe Ediciones; 2012.
17. Canavos GC. *Probabilidad y estadística: aplicaciones y métodos*. (1ª Ed.). Iztapalapa, México D.F.: McGraw Hill; 1988.
18. Walpole RE, Myers RH, Myers SL. *Probabilidad y estadística para ingenieros*. (6ª Ed.). México D.F.: Pearson Educación; 1999.
19. Devore JL. *Probabilidad y estadística para ingeniería y ciencias*. (7ª Ed.). Mexico. D.F.: Cengage Learning Editores; 2008.

20. Casas Sánchez J. *Inferencia Estadística para economía y administración de empresas*. (1ª Ed.). Madrid, España: Centro de Estudios Ramón Areces; 1997.
21. Bello Parias LD. *Estadística como apoyo a la investigación*. (1ª Ed.). Medellín, Colombia. Universidad de Antioquia; 2011.
22. Triola MF. *Probabilidad y estadística*. Novena edición. México. Pearson Educación; 2004.
23. Portus Govinden L. *Introducción a la Estadística*. (1ª Ed.). Bogotá D.F. Colombia. McGraw Hill; 2001.
24. Velasco Sotomayor G, Piotr Wisniewski M. *Probabilidad y estadística para ingeniería y ciencias*. (1ª Ed.). México: Thomson Learning; 2001.
25. Garcia, P., Molerio, O. & Pedraza, L. (2014). La prevención del tabaquismo y el alcoholismo en adolescentes y jóvenes desde las instituciones educativas. En *Psicogente*, 17(31), 93-106.
26. Muñoz, M. & Uribe, A. (2013). Estilos de vida en estudiantes universitarios. En *Psicogente*, 16(30), 356-367.
27. Montgomery D. *Estadística y probabilidad*. (2ª Ed.). México D.F: Limusa; 2002.
28. Ruiz Maya L, Martín Pliego FJ. *Estadística II: Inferencia*. (2ª Ed.). Madrid, España: Alfa Centauro; 2001.
29. Peña D. *Fundamentos de Estadística*. (1ª Ed.). Madrid, España: Alianza Editorial; 2001.
30. Droesbeke JJ, Lavalée P. "No respuesta" en las encuestas. *Revista electrónica. La sociología en sus escenarios*. 2009; 1(5): 1-10.
31. Morales Robayo A. *Probabilidad*. Bogotá. D.C., Colombia: Escuela de Ciencias Básicas, Tecnología e Ingeniería. Universidad Nacional Abierta y a Distancia; 2010.
32. Ortigón Pava MF. *Estadística descriptiva. (Segunda versión)*. Bogotá. D.C., Colombia: Escuela de Ciencias Básicas, Tecnología e Ingeniería. Universidad Nacional Abierta y a Distancia; 2010.
33. Contreras F. *Estadística descriptiva y análisis descriptivo con SPSS*. San Cristóbal, Venezuela: Aplicaciones de la Curva Normal. Fondo editorial UNET; 2007. 271-321.
34. Pérez López C. *Estadística aplicada a través de Excel*. Madrid, España: Pearson Educación, S.A; 2002.
35. Pulido HG, De La Vara Salazar R, González PG, Martínez CT, Pérez MDCT. *Análisis y diseño de experimentos*. México. D.F: McGraw-Hill; 2008.
36. Hernández Sampieri R, Fernández Collado C, Baptista Lucio MP. *Metodología de la investigación*. (6. ed.). México: Quirón Ediciones. McGraw-Hill; 2014.
37. Grajales GT. (2000). Tipos de investigación. Disponible de: <http://tgrajales.net/investipos.pdf>
38. Ñaupas HM. *Metodología de la investigación científica y elaboración de tesis*. Lima: Universidad Nacional Mayor de San Marcos; 2013.
39. Teodoro N, Nieto E. (2018). Tipos de investigación. Disponible de: <http://repositorio.unisdg.edu.pe/bitstream/USDG/34/1/Tipos-de-Investigacion.pdf>
40. Palella Stracuzzi S, Martins Pestana F. *Metodología de la investigación cuantitativa*. (3. ed.). Caracas, Venezuela: FEDUPEL; 2010. p.88, 90.
41. Espinosa Castro JF, Mariño Castro LM. *Estrategias para la permanencia estudiantil universitaria*. Barranquilla, Colombia: Ediciones Universidad Simón Bolívar; 2017.
42. Anguita JC, Labrador JR, Campos JD, Casas Anguita J, Repullo Labrador J, Donado Campos J. *La encuesta como técnica de investigación. Elaboración de cuestionarios y tratamiento estadístico de los datos*. Atención primaria. 2003; 31(8): pp. 527-538.
43. Díaz A, Oñate G, Riquena N. Uso de paquetes informáticos para el análisis de datos en ciencias biomédicas y sociales. *Ciencia e Innovación en Salud*. 2015; 3(2):43-49. DOI 10.17081/innosa.3.2.93
44. Kuzmar I, Consuegra JR, Cortés Castell E, Rizo Baeza M, Almanza C, Antonio K, Moreno S. *Hábitos y estado nutricional relacionados con las diferentes carreras universitarias*. AVFT. Archivos Venezolanos de Farmacología y Terapéutica. 2018; 37(3): 275-282.

45. Ortiz R, Bermúdez V, Torres M, Guzmán Lozada, J A, Valdiviezo Romero A J, Castillo Cueva OL, Chimbo Oyaque TA. *La actividad física de ocio como factor protector para la obesidad en la población adulta del área rural de Quingeo*. AVFT. Archivos Venezolanos de Farmacología y Terapéutica. 2018; 37(1): 42-46
46. Montenegro Rivera CM, Martínez Merlo JA. *Caracterización sociodemográficas, hábitos alimentarios y actividad física de estudiantes de primaria de una institución educativa pública del sector urbano*. AVFT. Archivos Venezolanos de Farmacología y Terapéutica. 2018; 37(4): 406-410
47. Silva A, Martínez, MS, Parra G, Palmar J, Linares S, Bermúdez V. *Asociación entre óxido nítrico e hipotensión en pacientes sometidos a hemodiálisis crónica del Hospital Universitario de Maracaibo*. Revista Latinoamericana de Hipertensión. 2018; 13(1): 8-13
48. Ortiz R, Torres M, Peña Cordero S, Alcántara Lara V, Supliguicha Torres M, Vásquez Procel X, Bermúdez V. *Factores de riesgo asociados a hipertensión arterial en la población rural de Quingeo Ecuador*. Revista Latinoamericana de Hipertensión. 2017; 12(3): 95-103
49. Contreras J, Graterol M., Wilches S, Garicano C, Hernández JD, Martínez MS, Graterol, R. *The San José de Cúcuta Metabolic Syndrome Prevalence Study: Design and Scop*. Revista Latinoamericana de Hipertensión. 2015; 10(4): 72-78
50. Ceballos, F. & Betancur Villegas J. "Simulación Discreta Aplicada a los Modelos de Atención en Salud", Investigación e Innovación en Ingenierías, vol. 2, no. 2, 2014. DOI: <https://doi.org/10.17081/invinno.2.2.2045>
51. Cabeza, R. "Localización de Datos de Contactos Personales Utilizando Técnicas de Minería Web y Redes Sociales", Investigación e Innovación en Ingenierías, vol. 4, no. 1, 2016. DOI: <https://doi.org/10.17081/invinno.4.1.2020>