



Sexto Semestre

Métodos de investigación cuantitativa

Unidad 3

**Muestreo,
procesamiento,
análisis e
interpretación de datos
cuantitativos**





Índice

Introducción	3
Competencia de la unidad	4
Logros de aprendizaje	4
3. Muestreo, procesamiento, análisis e interpretación de datos cuantitativos	5
3.1.1. Control de calidad de los datos.....	5
3.2.1 Tipos de muestreo	12
3.2.2 Planificación del muestreo	14
3.2.3 Tamaño muestral.....	15
3.2.4 El rigor en la investigación cuantitativa.....	16
3.2.5 La integración de métodos.....	17
3.3 Procesamiento, análisis e interpretación de datos.....	18
3.3.1 Categorización de la información.....	19
3.3.2 Codificación de una encuesta.....	22
3.3.3 La estadística en el análisis de datos cuantitativos	23
3.3.4 Estadística descriptiva	24
3.3.5 Representación gráfica de los datos.....	25
3.3.6 Estadísticas de tendencia central	27
3.3.7 Estadísticas de dispersión	34
3.3.8 Estadística inferencial.....	37
Cierre de la unidad	45
Fuentes de consulta	46



Introducción

En la unidad anterior pudiste observar la importancia que tienen las técnicas e instrumentos en la investigación cuantitativa. Un buen diseño metodológico es fundamental para el éxito de cualquier investigación científica. No obstante, después de que el instrumento está listo, ahora el investigador se enfrenta a otras interrogantes que debe resolver antes y después de la aplicación en el campo de estudio. En esta etapa de la investigación surgen preguntas como: ¿El instrumento diseñado realmente logra medir y recoger la información para la cual fue elaborado?, ¿Cómo realizó el levantamiento de la información en campo?, ¿Qué hago con toda la información levantada?, ¿Cómo realizó el procesamiento y análisis de la información?

En esta última unidad, trataremos de responder a las preguntas anteriormente planteadas, para ello, en la primera parte de la unidad revisaremos la importancia que tienen en el diseño de instrumento de investigación la confiabilidad y la validez, sobre todo porque son condiciones que debe asegurar el investigador antes de comenzar la aplicación del mismo en el campo; un instrumento que cumple con estas dos condiciones asegura buenos resultados de investigación.

En la segunda parte de la unidad, revisaremos un proceso fundamental para las investigaciones cuantitativas, particularmente para aquellas que optan por utilizar la encuesta como la técnica para lograr la comprobación de las hipótesis planteadas, nos referimos al muestreo. Normalmente el investigador se enfrenta al grave de problema de estudiar poblaciones de gran tamaño, las cuales resulta imposible analizar en su totalidad considerando las limitaciones a las que se enfrentan los investigadores de forma constante.

En este contexto el muestreo permite analizar porciones de la población seleccionada, otorgando el grado de científicidad requerido, pero sobre todo permite realizar descripciones, proyecciones y análisis válidos para el conjunto de la población seleccionada.

En un tercer momento, se revisará cómo se realiza el procesamiento, análisis e interpretación de los datos de una investigación cuantitativa. Después del levantamiento de la información el investigador se enfrenta a un proceso muy complejo, ¿qué hago con toda la información recuperada?, responder a esta interrogante puede ser muy complejo sobre todo cuando no existe claridad en cómo llevar a cabo este proceso. Es por eso, que en esta última fase se revisará una propuesta metodológica de cómo llevar a cabo este proceso; para ello se observará el procedimiento para realizar la codificación de una encuesta, se revisará las principales herramientas que te brinda la estadística para poder procesar, analizar e interpretar los resultados obtenidos a partir de una encuesta. Para lograr los propósitos anteriormente planteados, se han planteado las siguientes competencias y logros de aprendizaje:



Competencia de la unidad



Analiza la información cuantitativa haciendo uso de los fundamentos teóricos, metodológicos y prácticos para incidir en la prevención de enfermedad, y sugerir acciones de promoción de salud.

Logros de aprendizaje

- Definir los términos y describir las características de universo y muestra.
- Describir los tipos de muestra y los pasos que deben seguirse en su selección en estudios cuantitativos.
- Ejecutar los pasos del procesamiento, análisis e interpretación de datos.



3. Muestreo, procesamiento, análisis e interpretación de datos cuantitativos

Una vez realizados el levantamiento de la de información, le siguen el proceso de organización de los datos previo proceso de conteo (que hoy se ha facilitado gracias a los programas estadísticos) y otros que nos han originado cuadros y datos diversos. El proceso de análisis e interpretación es posible si se ha organizado la información en base a las variables propuestas, dándole así un “sentido” y orden, es decir, se convierten en conocimiento para el investigador. El interpretar hace referencia a la explicación que el investigador hace del significado de las cosas, hechos, sucesos, o en nuestro caso de los datos obtenidos.

A partir de ello podemos deducir que la interpretación y el análisis están ligados y depende uno del otro. No se puede hacer interpretación de la nada, deben existir datos que representen hechos para hacerlo.

3.1.1. Control de calidad de los datos

En la unidad anterior pudimos observar que la elaboración de instrumentos para la recolección de datos exige analizar la forma como dicho instrumento de medición cumple con la función para la cual ha sido diseñado. Un instrumento bien diseñado debe reunir dos cualidades importantes: confiabilidad y validez.

Como se estableció en la unidad anterior la confiabilidad se refiere a la capacidad del instrumento para arrojar datos o mediciones que correspondan a la realidad que se pretende conocer (Zorrilla: 2000¹), o sea, la exactitud de la medición, así como a la consistencia o estabilidad de la medición en diferentes momentos. A mayor confiabilidad de un instrumento, menor cantidad de error presente en los puntajes obtenidos. La estabilidad se relaciona con el grado en que el instrumento permite los mismos resultados en aplicaciones repetidas. Se dice que un instrumento es confiable si se obtienen medidas o datos que representen el valor real de la variable que se está midiendo. La confiabilidad se puede aumentar:

¹ Zorrilla S., Torres M., Luiz A., Alcino P (2000). Metodología de la Investigación. Mc Graw Hill, México,



Aplicando las reglas generales de elaboración de instrumentos, de tal forma que se eliminen los errores de medición, como preguntas ambiguas.

Aumentando el número de preguntas sobre determinado tema o mediante la inclusión de preguntas control.

Elaborando instrucciones claras que orienten el llenado o utilización de los instrumentos.

Aplicando los instrumentos o realizar las mediciones en condiciones semejantes.

Por otro lado, la validez se refiere al grado en que un instrumento mide lo que se pretende medir. La forma de garantizar la validez de un instrumento es construirlo una vez que las variables han sido claramente especificadas y definidas, para que estas sean las que se aborden y no otras; también se puede recurrir a la ayuda de personas expertas en el tema para que revisen el instrumento, a fin de determinar si cumple con la finalidad establecida.

Todo instrumento debe ser probado en una situación real antes de su aplicación definitiva con el fin de identificar errores o evaluar el tiempo necesario para *aplicarlo* (Joberg: 2012).

²Esta prueba se realiza en un grupo de población semejante a la que participará en el estudio en un número que puede oscilar entre 10 y 50 individuos. Es posible que sea necesario hacer más de una prueba si la primera prueba da lugar a un número importante de modificaciones en el cuestionario, con el fin de asegurar que los cambios introducidos sean satisfactorios. El objetivo principal de la prueba es verificar el grado de comprensión de las preguntas por parte de los interrogados, así como su capacidad de responder acertadamente. La prueba deberá estar seguida de una discusión sobre las ambigüedades o dificultades presentadas en el transcurso de la entrevista y no simplemente limitarse a entrevistar a un grupo de personas.

3.1.2. Factores que pueden afectar la confiabilidad y la validez de los datos



recogidos

Durante la planeación y ejecución de una investigación es menester revisar las posibles fuentes de error controlables, las cuales puede provenir de (Hernández, 2010)³:

- a) Fallas en la selección del método y diseño.
- b) Errores por muestreo.
- c) Sesgo por el diseño del instrumento. Errores debido a las no respuestas.
- d) Sesgo en la codificación de las respuestas abiertas.
- e) Falta de validez y confiabilidad del instrumento, error sistemático.
- f) Error en el proceso de análisis estadístico, error aleatorio.
- g) Errores en la interpretación de los datos.

Estos errores se pueden clasificar en dos grandes grupos: error sistemático, relacionado con el sesgo o validez de la información y el error aleatorio, debido al azar o sea asociados con la precisión de los datos.

El error aleatorio es esencialmente atribuible a variación en la prueba, la extensión del cual puede depender de aspectos de diseño del estudio (ejemplo: consideraciones del tamaño de la muestra) y características estadísticas del estimador (ejemplo: varianza). Este error puede ser: tipo alfa y tipo beta.

El error sistemático por otro lado ocurre cuando hay diferencia entre el dato que es actualmente estimado y el verdadero efecto de la medición; este tipo de error es atribuible a aspectos metodológicos de diseño o análisis del estudio, particularmente la selección de sujetos, calidad de información obtenida, y variables de importancia más que variación de la prueba.

Los errores sistemáticos o sesgos se clasifican en (Castro: 1989) ⁴:

Sesgos de selección: se refiere a la distorsión en la estimación del efecto, debido a:

- La manera como los sujetos son seleccionados para conformar la población de estudio (muestra).
- La elección de grupos de comparación en los diferentes estudios, al elegir grupos que no son comparables.
- Pérdidas por no respuesta o por no continuar en el estudio, o por sobrevida selectiva (estudios de casos y controles y Cross seccional).

³ Hernández Sampieri, Roberto; Carlos Fernández Collado y Pilar Baptista Lucio. (2010) Fundamentos de la Metodología de la Investigación—México: McGraw-Hill / Interamericana de España, S.A.U. 613 p.



Sesgos de información: se refiere a la distorsión en la estimación del efecto debido a errores de medición o mala clasificación de sujetos sobre una o más variables⁵. Los errores pueden ocurrir en la medición de las variables de exposición, ocurrencia de la enfermedad y variables que confunden, provenientes de los datos de diferentes fuentes de información como cuestionarios, registros de datos médicos (historia clínica, registro de consulta), registros de empleadores sobre exposición.

Los datos son menos ciertos, si están relacionados con la memoria de las personas, por ejemplo, averiguar sobre exposición a determinado factor causal de una patología que ha ocurrido mucho tiempo atrás.

Los errores de clasificación de estado de la enfermedad son muy frecuentes en los estudios epidemiológicos, de acuerdo al espectro de la enfermedad, no son claramente identificables los estados, en que se encuentra la enfermedad.

Fallas que se suelen cometer en los cuestionarios (Hernández: 2010)⁶:

- Carencia de incentivo para el informante.
- Preguntar por información que se puede obtener de otra fuente.
- Incluir preguntas que pueden ser consideradas ridículas o inapropiadas.
- Incluir preguntas que determinan una respuesta favorable.
- Usar equívocos o preguntas ambiguas.
- Usar para la respuesta espacios muy limitados que disminuyen la utilidad de la información.
- Desarrollar un formato muy largo o complicado.

Sesgos de Confusión: resulta cuando el efecto de los factores en estudio es mezclado, en los datos, con los efectos de variables extrañas (Craig: 1989)⁷.

Las variables de confusión se definen como aquellos factores conocidos asociados con la exposición de interés y con la enfermedad bajo estudio. Es importante el control de las variables de confusión, porque estas pueden llevar a asociaciones espurias o sesgadas entre el factor de riesgo y la enfermedad de interés. La edad, nivel educacional, estado socioeconómico, son variables comunes de confusión.

Existen dos buenas vías de control de confusión:

- Realizando el diseño pareado del estudio.
- Usando técnicas estadísticas tales como la estratificación o la regresión.

⁵ IBIDEM

⁶ Hernández Sampieri, Roberto; Carlos Fernández Collado y Pilar Baptista Lucio. (2010) Fundamentos de la Metodología de la Investigación.— México: McGraw-Hill / Interamericana de España, S.A.U. 613 p.

⁷ Craig, J.R. y Metze, L.P. (1982). Métodos de la investigación psicológica. México: Interamericana (Ed. Original 1979)



En otras palabras la confusión se debe a una severa distorsión al fallar el investigador en la rigurosidad del análisis o en el diseño pareado. En el análisis de los datos se debe tener en cuenta la evaluación de la validez interna y externa, que dependen de la población sobre la cual se hace la inferencia:

La validez interna se manifiesta con unos datos libres de sesgos (León:1993)⁸. Está relacionada con la característica del diseño del estudio que garantice que la población donde se realice el estudio (muestra) represente la población a la cual se desea hacer inferencia de los datos, es decir se garantice que no existe error sistemático.

En conclusión, toda investigación está expuesta a sesgos y errores que amenazan la confiabilidad y validez de los datos, los cuales pueden provenir de la selección del diseño o estrategia inadecuada para la solución del problema, el tamaño de la muestra, el diseño de los instrumentos o el plan de análisis. El investigador debe estar atento a las dificultades que puedan surgir, tratar de controlar o reducir al mínimo las limitaciones que puedan surgir; aunque algunas de ellas subsistirán (Argibay: 2006)⁹.

Cuando el investigador no logra controlar tales amenazas a la validez y confiabilidad de sus datos por lo menos debe darlas a conocer en sus informes y de esta manera reconocer que el diseño del estudio tiene limitaciones y que los resultados deben considerarse provisionales, lo cual de todos modos no invalida el estudio y en cambio da cuenta de la ética investigativa y de honradez científica (Kerlinger: 1988) ¹⁰.

3.2. El muestreo en estudios cuantitativos

Después de tener definido el tipo de estudio se debe contestar un interrogante: ¿quiénes van a ser medidos? Se define si va a ser una población completa o una muestra de ella. Para contestar estos interrogantes se debe tener claro la unidad de análisis, que puede

⁸ León, O.G. y Montero, I. (1993). Diseño de investigaciones. Madrid: McGraw-Hill.

⁹ Argibay, J.C. (2006), "Técnicas psicométricas. Cuestiones de validez y confiabilidad". En: Subjetividad y Procesos Cognitivos, 8, 15-33.

¹⁰ Kerlinger, F.N. (1988). Investigación del comportamiento. México: McGraw-Hill



estar conformada por personas, organizaciones, objetos y estará muy de acuerdo con los objetivos y el problema a investigar.

Cuando se tiene definida la unidad de análisis y la unidad de muestreo (por ejemplo. Unidad de análisis: persona; unidad de muestreo la vivienda) se procede a delimitar la población, plantear sus características (tiempo y lugar) y seleccionar la muestra en caso necesario. (Aramburú:2011)¹¹

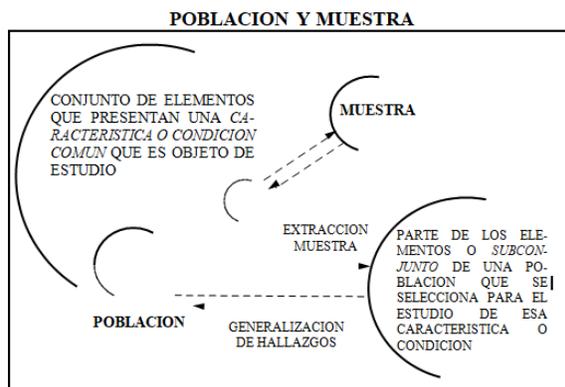
La población en investigación no se refiere siempre a la definición de tipo demográfico: conjunto de personas ubicadas en un área geográfica determinada, claramente definida. Se pueden formar poblaciones estructurales de diferentes maneras: los nacidos vivos durante un determinado tiempo en un país o ciudad; los niños que cursen 4° y 5° de primaria en escuelas públicas de la ciudad de Neiva en un determinado año.

¹¹ Aramburú, Carlos Eduardo (2001). Métodos y técnicas de investigación social. Gerencia social. Diseño, monitoreo y evaluación de proyectos sociales. Lima-Perú: Universidad del Pacífico. ISBN 9972-603-32-6..



DEFINIR LA POBLACIÓN, TAMAÑO Y ELEMENTOS QUE LA COMPONEN
Determinar la unidad de observación, la unidad muestral y sus características
Determinar aquella información necesaria para hacer la selección de la muestra
Definir el tamaño de la muestra
Definir el método de selección de la muestra
Definir los procedimientos que deben seguirse para la selección de la muestra

Una población también puede estar constituida por elementos como historias clínicas, registros médicos, certificados de defunción. Los eventos biológicos generalmente no se estudian sobre la totalidad de los individuos en un área, sino en una fracción de esta población denominada muestra (Aramburú: 2001).¹²



Fuente: Blalock, H. (1986). Población y Muestra (Figura 1): 'Estadística Social'. Fondo de Cultura Económica. México. 1986. Retomado de <https://es.scribd.com/document/356083141/15MIS-Blalock-1-Unidad-4>

La muestra se define como un conjunto de objetos y sujetos procedentes de una población (Kohan, 1994)¹³; es decir un subgrupo de la población, cuando esta es definida como un conjunto de elementos que cumplen con unas determinadas especificaciones. De una población se pueden seleccionar diferentes muestras

¹² "Introducción al Diseño de una Encuesta". Disponible en : <http://www.estadistica.mat.uson.mx/Material/etapas.pdf>

¹³ Cortada de Kohan, N. (1994), Diseño estadístico, Buenos Aires, Editorial Universitaria de Buenos Aires.



Las muestras pueden ser:

Representativas: cuando reflejan o representan las características de la población de donde provienen, por lo tanto los resultados son aplicables o inferibles a dicha población (Cordero: 2015)¹⁴.

Las muestras representativas deben ser cualitativamente buenas, es decir:

- Homogéneas, compuesta por elementos de una sola población.
- Adecuadas, incluir todas las posibles variaciones en las características de los elementos de la población.
- No viciadas, no presentar sesgo de selección, porque puede invalidar el estudio, no importa cuán sofisticado sea.
- No representativas: son aquellas que no reflejan las características de la población, por lo tanto los resultados no son inferibles a la población. Estos resultados pueden ser válidos cuando no existen fallas en la metodología de recolección de datos.

3.2.1 Tipos de muestreo

Además de esta clasificación las muestras pueden ser probabilísticas y no probabilísticas. En las primeras los elementos tienen la misma probabilidad de ser escogidos, se hace una selección al azar o aleatoria de los elementos o unidades de muestreo.

En las muestras no probabilísticas la selección no depende del azar, los elementos se escogen de acuerdo a unas características definidas por el investigador o la investigación, es decir depende de decisiones de personas por lo tanto suelen estar sesgadas (Argibay:2009)¹⁵.

Elegir entre una muestra representativa o no, o una muestra probabilística o no, depende de los objetivos del estudio, el tipo de investigación y de la contribución que se piensa realizar con dicho estudio.

Cuando la muestra es probabilística, se debe calcular el tamaño muestral previamente, luego se procede a utilizar cualquiera de los siguientes métodos para la selección de las unidades de muestreo.

- **Muestra aleatoria simple:** es la más sencilla solo el azar decide. Se utilizan los métodos de lotería, o los números aleatorios para seleccionar los elementos; las conclusiones pueden ser para toda la población si la muestra es representativa. Requiere tener un listado de los elementos de la población o un mapa del área, cuando la unidad de muestreo es la vivienda¹⁶.

¹⁴ Cordero R. "El mundo de las encuestas. Reflexiones sobre su desarrollo e importancia" (2005), Universidad Diego Portales.

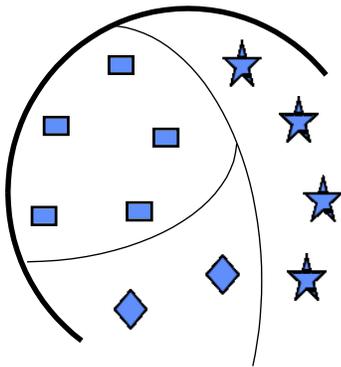
¹⁵ Argibay, Juan Carlos. (2009). Muestra en investigación cuantitativa. Subjetividad y procesos cognitivos, 13(1), 13-29. Recuperado en 07 de febrero de 2018, de http://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S1852-73102009000100001&lng=es&tlng=es.

¹⁶ IBIDEM



Muestra sistemática (Gardner, 2003)¹⁷: en este tipo de muestreo, teniendo los datos del total de población (N) y el tamaño muestral (n) se obtiene el salto muestral que consiste en la comparación de estos dos valores (N/n). También requiere un listado de elementos de la población, con número aleatorio de la calculadora o el computador se selecciona el primer elemento de la muestra, a este número se le suma el salto muestral y da el número del segundo elemento: $n_1 + (N/n) = n_2 + (N/n) = \dots n$. Esta muestra se debe evitar cuándo se sabe o se sospecha que las características que se estudian, presentan ciclos en la población. Se puede inferir a la población si es una muestra representativa.

Muestra estratificada¹⁸: este tipo de muestra se utiliza generalmente para control de variables de confusión (sesgos), el procedimiento es el siguiente: Se definen los estratos de la variable que se quiere controlar (edad, estado socioeconómico, escolaridad).



Se debe conocer la proporción (%) de cada estrato con respecto a la población.

Aplicando esta proporción al tamaño muestral se obtiene el tamaño de cada submuestra en cada estrato.

Utilizando el método aleatorio simple se seleccionan los elementos de cada submuestra.

Se combinan las submuestras.

Las probabilidades son diferentes para cada estrato. Son “homogéneos” dentro de sí y heterogéneos entre sí.

También se puede obtener una muestra estratificada no proporcional, cuando se asigna igual proporción a todos los estratos.

Muestra por conglomerados (Kohan, 1994)¹⁹: es útil en investigaciones que abarcan extensas zonas geográficas.

- Utilizando un mapa, se divide el total de la población en conglomerados homogéneos entre sí, heterogéneos dentro de sí.
- Se toma al azar una muestra de esos conglomerados.

¹⁷ Gardner, R.C. (2003), *Estadística para Psicología usando SPSS para Windows*, México, Pearson Educación de México.

¹⁸ IBIDEM

¹⁹ Cortada de Kohan, N. (1994), *Diseño estadístico*, Buenos Aires, Editorial Universitaria de Buenos Aires.



- De cada uno de los conglomerados escogidos, se toman al azar los elementos de la muestra.

Las muestras no probabilísticas o también llamadas dirigidas suponen un procedimiento de selección informal y un poco arbitrario (Kohan: 1994), son utilizadas en muchas investigaciones, sobre todo las que requieren la selección de sujetos con una determinada característica, especificadas en el planteamiento del problema. En estas se incluyen:

Muestra de sujetos voluntarias (Gardner, 2003)²⁰: Se trata de muestras fortuitas, utilizadas con frecuencia en Medicina, por ejemplo sujetos que acceden voluntariamente a participar en un estudio que monitorea los efectos de un medicamento. Con este tipo de muestra no se puede inferir, ya que las características de los sujetos de la muestra pueden ser diferentes al total de la población.

Muestra por conveniencia: también se trata de una muestra fortuita, se selecciona de acuerdo a la intención del investigador por ejemplo estudios en pacientes hospitalizados, siempre que el hospital no atienda al total de la población²¹.

Muestras de pacientes por cuotas: la proporción de participantes en las encuestas lo decide el investigador de acuerdo, al comportamiento de ciertas variables demográficas en la población (Polit: 1997)²²: Se le dice a un entrevistador que en la calle entreviste a 200 personas,(50%) mujeres y (50%)hombres, proporciones iguales por grupo étnico. La decisión de quién participa es del entrevistador.

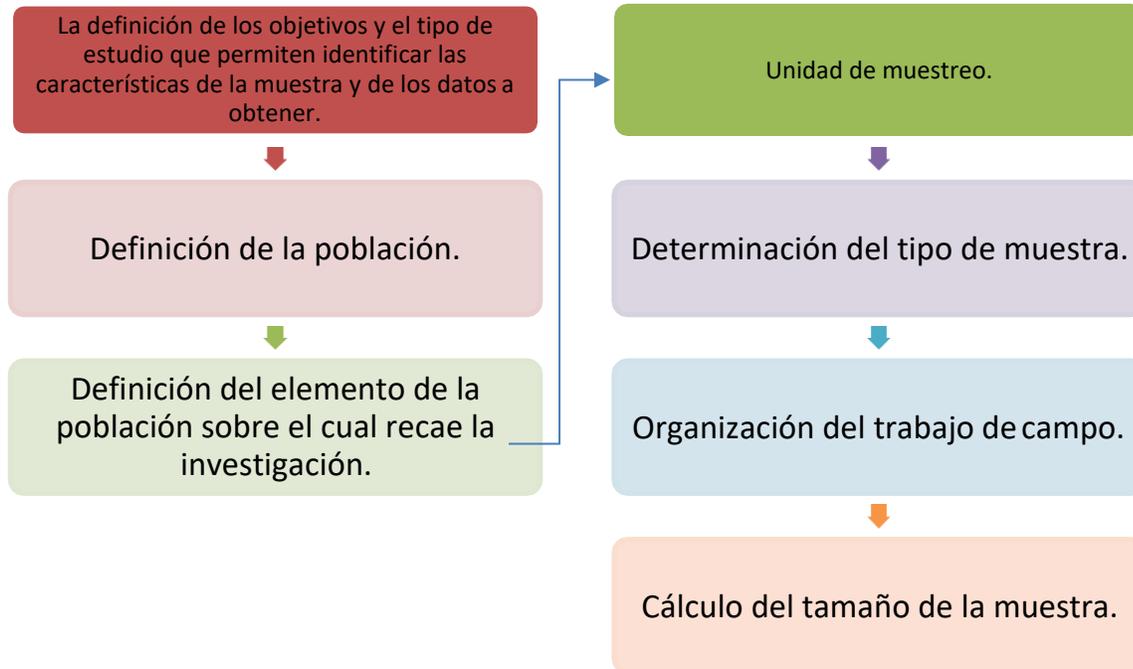
3.2.2 Planificación del muestreo

El procedimiento para la planificación del muestreo incluye los siguientes pasos:

²⁰ Gardner, R.C. (2003), Estadística para Psicología usando SPSS para Windows, México, Pearson Educación de México.

²¹ IBIDEM

²² Polit, D.F. y Hungler, B.P. (1997), *Investigación científica. En ciencias de la salud*, México, McGraw-Hill Interamericana.



3.2.3 Tamaño muestral

El tamaño muestral va a depender de (Arnau:1982)²³:

El tamaño de la población. Definir si es **finita** cuando se conoce el total de los elementos de la población o **infinita** cuando no se conocen.

La variabilidad de los elementos de esta población con relación a las variables en estudio (varianza). La naturaleza y frecuencia del evento estudiado (si se conoce o no). La precisión establecida, error de muestreo, para garantizar los resultados.

Los recursos administrativos, financieros, humanos y equipos.

Utilizando la siguiente fórmula se puede calcular el tamaño muestral, cuando se conocen datos como promedios en poblaciones infinitas:

$$n = Z^2 \beta^2 / e^2$$

Z: equivale al nivel de confiabilidad de los datos está muy de acuerdo con el error de los datos que se acepte o sea el alfa o el beta.

²³ Arnau Gras, J. (1982), Psicología experimental, México, Trillas.



β : significa la varianza de los datos.

e: significa el error de muestreo, que tan representativa se desea la muestra. $n: Z^2 pq / \{(e^2 + (Z^2 pq / N)\}$

Esta fórmula se utiliza para calcular tamaño muestral cuando se conocen datos como tasas, o proporciones de un evento en la población finita, por esto la formula incluye en su denominador el factor de corrección.

Para calcular el tamaño muestral en estudios analíticos, se necesita conocer los siguientes datos:

- Nivel de α dispuesto a tolerar.
- Nivel de β dispuesto a tolerar.
- Mínimo Riesgo Relativo considerado clínicamente útil de detectar.
- Estimativo de la incidencia o de la prevalencia de la enfermedad y de la exposición.

Con estos datos utilizando un paquete estadístico como Epi-info se puede calcular el tamaño muestral, para estudios descriptivos o analíticos (casos y controles o cohortes). Después de tener calculado el tamaño muestral se procede a seleccionar las unidades de muestreo utilizando los tipos de muestreo revisados previamente.

3.2.4 El rigor en la investigación cuantitativa

A lo largo de la segunda fase de la investigación, en la que se incluye la recogida de datos en el campo, el investigador cuantitativo habrá de seguir tomando una serie de decisiones, modificando, cambiando, alterando o rediseñando su trabajo (León: 2003)²⁴. La duración del proceso de aplicación de la encuesta, el tiempo de sistematización de la información, pero sobre todo, cómo analizar y publicar los resultados obtenidos a partir de la aplicación de los instrumentos de recolección de información.

Cuando el investigador se siente a gusto y relajado y se centra en lo que está sucediendo, y los participantes comienzan a entender qué es lo que se está estudiando y reconocen el interés especial del investigador, entonces pueden facilitar mucha más información para la indagación. En ese momento este momento puede comenzar la recogida productiva de datos. Esta es la etapa más interesante del proceso de investigación; la luz, el orden y la comprensión van emergiendo.

²⁴ León, O.G. y Montero, I. (2003), Métodos de investigación en psicología y educación. Madrid, McGraw-Hill/Interamericana de España



Pero ello sucede gracias al esfuerzo que supone la indagación realizada (Cordero:2005).²⁵ El proceso de análisis de datos ya comienza en esta etapa, y se inicia un proceso de recogida de aquellos datos que realmente interesan al desarrollo de la investigación, de tal forma que se evite la recogida de informaciones innecesarias.

La utilización de métodos de manejo de datos es imprescindible. La sistematización de la información y notas de campo deben ser referenciadas, unidas a sus fuentes pero, a la vez, separadas de las mismas, y por supuesto organizadas eficazmente. Durante esta etapa es preciso asegurar el rigor de la investigación. Para ello debemos tener en cuenta los criterios de suficiencia y adecuación de los datos. La suficiencia se refiere a la cantidad de datos recogidos, antes que al número de sujetos. (Larios: 2001)²⁶ La suficiencia se consigue cuando se llega a un estado de saturación informativa” y la nueva información no aporta nada nuevo. La adecuación se refiere a la selección de la información de acuerdo con las necesidades teóricas del estudio y el modelo emergente.

Otra forma de asegurar el rigor es a través del proceso de triangulación utilizando diferentes métodos, datos, teorías o disciplinas. Progresivamente el investigador se va integrando en el lugar de estudio, hasta que llega un momento en que es parte del mismo (Larios: 2001). En esta fase, el investigador no puede llevar a cabo una recogida productiva de datos, ya que, por un lado, pierde sensibilidad ante las actividades cotidianas. Por otro, pierde perspectiva hacia el lugar y los miembros del grupo. La señal inconfundible de que ha llegado el momento de abandonar el campo es cuando el investigador llega a ser considerado parte integrante del contexto en el que se encuentra, cuando se considera un “nativo”.

3.2.5 La integración de métodos

Es menester señalar que no se puede decir que un paradigma es mejor que el otro, es decir, no se trata de ver que el cualitativo es el mejor y que el cuantitativo está superado o muy criticado (Zorrilla: 2000)²⁷ Ambos permiten un acercamiento para conocer aspectos de la realidad distintos, y conocerlos de diferentes maneras. Dependiendo del interés en qué es lo que se quiere conocer, por qué se quiere conocer, es que puede decidirse la utilización de uno u otro método.

²⁵ Cordero R. (2005) “El mundo de las encuestas. Reflexiones sobre su desarrollo e importancia” (2005), Universidad Diego Portales.

²⁶ Larios Osorio, Vicente (2001) “¿Cómo hacer una encuesta?”. U.A.Q. México

²⁷ Zorrilla S., Torres M., Luiz A., Alcino P. Metodología de la Investigación. Mc Graw Hill, México, 2000



Las dos metodologías ofrecen elementos importantes, tienen límites y tienen posibilidades. La tarea del investigador, en todo caso, es conocer los potenciales de cada paradigma, estar muy claro en sus preguntas de investigación y saber en cuál de ellos ubicarse para generar el conocimiento que quiere. No es válido establecer una separación tajante o dicotomía entre los dos enfoques metodológicos. Estos se complementan en el conocimiento, explicación y comprensión de la realidad social (Joberg: 2002)²⁸. En consecuencia debe considerarse la investigación desde un sentido de totalidad que elude la polaridad entre categorías metodológicas no enfrentadas. La realidad social es total y los distintos abordajes del conocimiento no solo no se pueden integrar sino que necesariamente deben complementarse. De todas maneras las alternativas metodológicas cuantitativa y cualitativa cumplen cada una su papel y su selección depende de la pregunta de investigación y del interés que anime al investigador.

3.3 Procesamiento, análisis e interpretación de datos

Aunque situamos esta fase tras el trabajo de campo, en modo alguno queremos significar que el proceso de análisis de la información recogida se inicia tras el abandono del escenario. Antes al contrario, la necesidad de contar con una investigación con datos suficientes y adecuados exige que las tareas de análisis se inicien durante el trabajo de campo. No obstante, por motivos didácticos la situamos como una fase posterior. El análisis de datos cuantitativo va a ser considerado aquí como un proceso realizado con un alto grado de sistematización que, a veces, permanece implícita en las actuaciones emprendidas por el investigador.

En este sentido, resulta difícil hablar de una estrategia o procedimiento general de análisis de datos cuantitativos, con la salvedad de lo que pueda inferirse a partir de las acciones identificadas en un análisis ya realizado (Hernández: 2010)²⁹. No obstante, tomando como base estas inferencias, es posible establecer una serie de tareas u operaciones que constituyen el proceso analítico básico, común a la mayoría de los estudios en que se trabaja con datos cuantitativos. Estas tareas serían: a) reducción de datos; b) disposición y transformación de datos; y c) obtención de resultados y verificación de conclusiones.

En cada una de estas tareas es posible distinguir, asimismo, una serie de actividades y operaciones concretas que son realizadas durante el análisis de datos, aunque no necesariamente todas ellas estén presentes en el trabajo de cada analista (Castro: 1997)³⁰. En ocasiones, determinadas actividades pueden extenderse hasta constituir por sí mismas el proceso de análisis o, por el contrario, pueden no ser tenidas en cuenta en el tratamiento de los datos, de acuerdo con los objetivos del trabajo, el enfoque de investigación, las habilidades del investigador, etc.

²⁸ Joberg S., Metodología de la Investigación social. Editorial Trillas, México. 2002.

²⁹ Hernández Sampieri, Roberto; Carlos Fernández Collado y Pilar Baptista Lucio. (2010) Fundamentos de la Metodología de la Investigación.— México: McGraw-Hill / Interamericana de España, S.A.U. 613 p

³⁰ Bonilla Castro, Elsy y Rodríguez Sehk, Penélope. Más allá del dilema de los métodos. La investigación en ciencias sociales. 3ª Ed. Santafé de Bogotá, Ediciones Uniandes, 1997. Cap. 4. El proceso de la investigación cualitativa. Pág. 119-145.



Entre ellas no siempre se establece una sucesión en el tiempo, y pueden ocurrir de forma simultánea, o incluso estar presentes varias de ellas dentro de un mismo tipo de tarea.

3.3.1 Categorización de la información

Significa que se busca reducir los datos de nuestra investigación con el fin de expresarlos y describirlos de alguna manera (conceptual, numérica o gráficamente), de tal manera que respondan a una estructura sistemática, inteligible para otras personas, y por lo tanto significativa. (Rodríguez: 1996)³¹ La reducción de datos es una clase de operación que se realiza a lo largo de todo el proceso de investigación y puede hacerse de distintas formas.

Entre las tareas de reducción de datos cualitativos, posiblemente las más representativas y al mismo tiempo las más habituales sean la categorización y la codificación. Incluso a veces se ha considerado que el análisis de datos cualitativos se caracteriza precisamente por apoyarse en este tipo de tareas. Las categorías son las clasificaciones más básicas de conceptualización, y se refieren a clases de objetos de los que puede decirse algo específicamente. (Thiebaut: 1999)³². Del griego: *Kategoría*. Desde Aristóteles hasta Kant, las categorías han tenido un importante papel en la historia de la Filosofía. Son las últimas –en el sentido de las más básicas, elementales– estructuras lógicas desde las que clasificamos nuestros juicios.

En la metodología cuantitativa, los datos recogidos necesitan ser traducidos en categorías con el fin de poder realizar comparaciones y posibles contrastes, de manera que se pueda organizar conceptualmente los datos y presentar la información siguiendo algún tipo de patrón o regularidad emergente. La categorización (es decir, cerrar o establecer las categorías) facilita la clasificación de los datos registrados, y por consiguiente, propicia una importante simplificación.

La categorización consiste en la segmentación en elementos singulares, o unidades, que resultan relevantes y significativas desde el punto de vista de nuestro interés investigativo. La investigación cuantitativa se mueve en un volumen de datos muy grande, por lo que se hace necesario categorizarlos para facilitar su análisis y poder responder a los objetivos que pueden ser cambiantes a medida que se va obteniendo la información (Hernández: 2010)³³.

Las categorías pueden constituirse utilizando una palabra de una idea que sea similar en otras ideas, o creando un nombre en base a un criterio unificador, logrando que al final del proceso todas las ideas estén incluidas en alguna categoría. Al construir las categorías no se deben hacer interpretaciones previas y siempre respetar la información obtenida. Cuando se han incluido muchas ideas en una categoría se debe analizar la posibilidad de dividirla en subcategorías para facilitar el análisis posterior.

En el caso de la investigación, la categorización puede durante la fase de la operacionalización, pues de aquí surgen las categorías e indicadores para el diseño de los instrumentos de

³¹ Gregorio Rodríguez Gómez, Javier Gil Flores, Eduardo Garcés Jiménez, Metodología de Investigación Cualitativa, Editorial Algibe, 1999, Pág. 205

³² Thiebaut, Carlos (1999) Conceptos fundamentales de la Filosofía, Alianza Editorial, Madrid.

³³ Hernández Sampieri, Roberto; Carlos Fernández Collado y Pilar Baptista Lucio. (2010) Fundamentos de la Metodología de la Investigación.— México: McGraw-Hill / Interamericana de España, S.A.U. 613 p



investigación o después de haber hecho el levantamiento de la información. Es decir, la categorización puede estar predefinida por el analista (lo que usualmente se hace en el método de la encuesta), o por el contrario, puede surgir a medida que se analizan los datos ya recogidos (Craig: 1982)³⁴:

1. El primer caso (antes) consiste en establecer un conjunto de categorías (o clases de fenómenos o hechos) a partir de las teorías que estudian ese fenómeno o hecho (método etc).
2. En el segundo caso (después), tanto si se ha aplicado la encuesta, como si solo hay observación en terreno, se establecen categorías de análisis.

A partir de los antecedentes recogidos en el marco teórico, se establecen las ideas o tópicos más sobresalientes que deberían ser consultados / investigados / recopilados en terreno, haciendo una lista con ellos (León: 1993)³⁵. Por ejemplo, si un investigador tiene que averiguar sobre el consumo de alimentos en determinado lugar, puede terminar confeccionando la lista de categorías que sigue (categorías cuya importancia habría quedado muy clara en su Marco Teórico, que es lo que las sugiere):

Área de investigación: ALIMENTACIÓN

Categorías:

1. Satisfacción y control del hambre.
2. Dieta.
3. Uso de condimentos.
4. Comidas.
5. Servicios comerciales de comidas.

Definición operacional. Se procede a definir cada una de las categorías, indicando cómo se las encuentra en la realidad, según lo presentan las teorías recogidas en el Marco Teórico o en textos especializados. No se recomienda definir de acuerdo a los 26543588Diccionarios, porque estas definiciones son demasiado generalizadas y a menudo no reflejas la realidad que se está estudiando. Se definen las categorías y las subcategorías.

Área de investigación, Alimentación: Información referente a la comida de la familia (o escuela, o lugar).

³⁴ Craig, J.R. y Metze, L.P. (1982). Métodos de la investigación psicológica. México: Interamericana (Ed. Original 1979)

³⁵ León, O.G. y Montero, I. (1993). Diseño de investigaciones. Madrid: McGraw-Hill.

**CATEGORÍAS**

1. CATEGORÍA: Satisfacción y control del hambre. Conceptos sobre la comida y la ingestión de alimentos que tienen los entrevistados (Murdock): 1997)³⁶.

SUBCATEGORÍAS

- 1.1. Exceso en la comida: (Descripción o concepto de lo que se entiende por este concepto, según la realidad a estudiar: ...)
- 1.2. Actitud ante el derroche de alimentos: (concepto: ...)
- 1.3. Adaptación especial a la escasez y el hambre: (concepto: ...)
- 1.4. Dieta, en términos de tipo y cantidad de alimentos ingeridos: (etc. ...)
- 1.5. Elaboraciones culturales del apetito, etc.
2. CATEGORÍA Dieta: Tipos de alimentos ingeridos, según la clasificación XXX; horas del día y cambios en la ingesta según las estaciones del año.

SUBCATEGORÍAS

- 2.1. Abundancia o escasez de comida: (descripción del concepto).
- 2.2. Cambios de las dietas según los alimentos disponibles (concepto).
- 2.3. Alimentos fundamentales y otros que se consumen.
- 2.4. Proporción de los diversos alimentos en la dieta.
- 2.5. Diferencias de grupos en la dieta (v.g.: por edad, sexo, clase, etc.).
- 2.6. Diferencia entre alimentos comestibles y dañosos.
- 2.2.7. Preferencias, prohibiciones y tabúes en los alimentos, etc.
3. CATEGORÍA: Uso de condimentos: Yerbas aromáticas y especias usadas en la preparación y consumo de los alimentos.

SUBCATEGORÍAS:

- 3.1. Clases de condimentos usados.
- 3.2. Momento en que son usados, o incorporados los condimentos a los alimentos.
- 3.3. Preparación o elaboración de las yerbas aromáticas o especias.
- 3.4. Importancia social de los condimentos.
4. CATEGORÍA: Comidas: Alimentos ingeridos durante el día normal o en momentos especiales.

SUBCATEGORÍAS:

- 4.1. Regularidad de las comidas.
- 4.2. Frecuencia de comidas irregulares (v.g.: horas, número al día, participantes, composición, etc.)
- 4.3. Comidas regulares (horas, número al día, participantes, etc.)

³⁶ Murdock, George P. (1994), Guía para la clasificación de los datos culturales, Editado por la Universidad Autónoma Metropolitana, México, 2ª edición.



5. CATEGORÍA: Servicios comerciales de comidas: Empresas o personas ofreciendo la venta de comidas por una utilidad.

SUBCATEGORÍAS:

- 5.1. Tamaño del servicio.
- 5.2. Número de comidas que se ofrecen.
- 5.3. Etc., etc.

3.3.2 Codificación de una encuesta

Es momento de recuperar la codificación realizada al inicio del diseño del instrumento de investigación, pues es parte fundamental del proceso de análisis, el cual parte de este primer ejercicio de atomización de conceptos, indicadores y variables.

Después de trabajar preparando un instrumento de encuesta y de obtener el cuestionario posterior, la codificación de la encuesta prepara la estructura para introducir los resultados individuales (León: 1993)³⁷. No es una tarea difícil, pero necesita mucha atención y precisión. La codificación consiste en dos pasos. El primero es la asignación de las variables y el segundo es la determinación de los valores. Las variables incluyen preguntas de la encuesta y los valores son típicamente asignaciones numéricas para las elecciones que los encuestados seleccionan, ya sea que las preguntas sean para respuestas abiertas o cerradas. Una forma sencilla de realizar este procedimiento es:

- a) Utiliza una copia en blanco del cuestionario para simplificar el proceso de codificación.
- b) Utiliza un cuestionario guía para identificar cada pregunta por separado. Cada pregunta representa una variable e incluso dentro de una pregunta puede existir más de una variable. Una pregunta que tiene una respuesta cerrada, pero pide una respuesta abierta para aclarar la respuesta tiene dos variables, que consiste en la primera pregunta, seguida de la respuesta por escrito, que es la segunda variable. Si el cuestionario tiene preguntas que permiten a los encuestados elegir múltiples respuestas, cada una de las respuestas se convierte en una variable.
- c) Numera las variables en el formulario de encuesta en blanco. La forma más fácil de enumerarlas es seguir el sistema de numeración de la encuesta y usar decimales o letras para diferenciar una pregunta de la siguiente. Si la pregunta 10 tiene tres variables, podrías enumerarlas 10,1, 10,2 y 10,3 o 10.a, 10.b y 10.c. Cuando la pregunta tiene varias respuestas, cuenta las opciones y asigna a cada una un número, como 7.1 a 7.5 si existen cinco opciones para esa pregunta.

³⁷ León, O.G. y Montero, I. (1993). Diseño de investigaciones. Madrid: McGraw-Hill.



- d) Asigna los valores para cada variable. El software de procesamiento u hoja de cálculo de una encuesta normalmente utiliza números para ordenar y calcular la distribución de las respuestas. Si una variable tiene cinco opciones diferentes para que elijan los encuestados, escala las respuestas de 1 a 5. Si hay 8, escálalas, de 1 a 8.
- e) Navega por algunas de las encuestas si contiene preguntas abiertas y busca patrones de respuesta. Da números a las respuestas comunes para hacer más fácil y más rápida la entrada de datos y la interpretación. Asigna un número para las respuestas poco comunes al elegir una categoría otros. Es probable que necesites una hoja de papel para la clave de la codificación de las respuestas abiertas.
- f) Haz una copia de la encuesta que utilizaste para codificar las variables y los valores en caso de que no puedas encontrar la original o si la envías hacia fuera para la entrada de datos, para que tengas una copia a mano por si surgen preguntas.
- g) Numera cada encuesta si no están ya numeradas. Esto te ayuda a encontrar los errores de entrada de datos que se puedan producir al introducir los datos.

La codificación de una encuesta mediante un software como Microsoft Office o Statistical Package for the Social Sciences (SPSS), puede ser más sencilla, sobre todo al momento del procesamiento y análisis de la información, no obstante en ningún caso se puede pasar por alto el proceso de categorización de la información.

Para aprender más...

Sí te interesa conocer más sobre el procesamiento de una encuesta utilizando Excel te invitamos a revisar la siguiente liga:

<https://estadisticainformatica.wordpress.com/2011/09/12/codificacion-y-tabulacion-de-encuestas-en-excel-paso-a-paso/>

3.3.3 La estadística en el análisis de datos cuantitativos

La estadística permite recolectar, analizar, interpretar y presentar la información que se obtiene en el desarrollo de una determinada investigación; el paso siguiente a la elaboración del Plan de investigación estadístico es la recolección definitiva de los datos (Ritchey: 2002)³⁸ Esta recolección consiste en los procedimientos de observación y anotación o registro de los hechos en los formularios que se han diseñado previamente. De esta recolección depende en gran parte la

³⁸ Ritchey F. (2002), Estadística para las ciencias sociales. Mc Graw Hill.



calidad del análisis que se realice, ya que pueden existir interpretaciones falsas y análisis erróneos de las situaciones, cuando existen fallas en la recolección de la información.

Por esto todo dato o grupo de datos obtenido, antes de ser totalizado y utilizado requiere un examen crítico, sobre aspectos de exactitud, precisión y representatividad, lo que se denomina la crítica del dato; después en caso que sea necesario, se procede a su codificación. En la actualidad el procesamiento de la información que permite el cálculo de medidas estadísticas para el análisis se hace a través de programas de computador, por lo que se debe poner más atención al significado de los resultados que a la utilización de las fórmulas.

Una vez se terminan de recoger los datos, se deben organizar y resumir para obtener información significativa, es decir, analizar los datos utilizándose para esto: 1) la estadística descriptiva 2) Distribución normal (puntuaciones Z), 3) Razones y tasas, 4)

3.3.4 Estadística descriptiva

El concepto básico de la descripción estadística es la distribución de frecuencias, método para organizar y resumir datos, que son ordenados indicándose el número de veces que se repite cada valor (Bello: 2005). ³⁹Esta distribución puede realizarse con las variables medidas desde el nivel nominal hasta el de razón.

Los datos recolectados y tabulados se disponen sistemáticamente de acuerdo a su complejidad, y se presentan de varias formas, ya sea en forma combinada o individual así:

- a) Textual
- b) Cuadros o tablas
- c) Gráficas

Cuadros o tablas

Los cuadros o tablas se refieren a un arreglo sistemático de la información dispuesto en filas y columnas con fines comparativos. Los datos deben estar ordenados para que ofrezcan algún tipo de información; son un buen complemento del texto en los informes. En el esquema siguiente se presentan los elementos constitutivos de una tabla⁴⁰.

Los cuadros o tablas pueden ser:

- Unidimensional cuando se refieren a una sola variable.

³⁹ Bello, León Darío. (2000). Estadística como apoyo a la Investigación. Editorial L.Vieco e Hijos Ltda

⁴⁰ IBIDEM



- Bidimensional cuando se incluyen dos variables.
- Pluridimensional o multidimensional cuando se incluyen tres o más variables.

DISEÑO DE TABLAS

Título | **TABLA 1.** Personas incapacitadas en un periodo de dos semanas e ingreso económico familiar anual. Colombia, 1994-1995

Nota de encabezado | Tasa por 1.000 Habitantes

Zonas	Total	Ingreso económico *			
		120.000 o menos	120.001 250.000	250.001 500.000	500.001 y más
Urbana	104	142	112	100	73
Rural	113	120	111	119	75
TOTAL	109	127	112	115	73

Nota de pie | * En pesos

Fuente | Fuente: MINISTERIO DE SALUD PÚBLICA. Estudio de recursos humanos para la salud y educación médica en Colombia. Bogotá, 1994.

La variable supuestamente independiente se coloca horizontalmente, la dependiente se sitúa en la vertical.

Fuente: Hernández, R. (2010). Diseño de tablas (Figura 2).. Capítulo 10 Análisis de datos Cuantitativos. Fundamentos de la Metodología de la Investigación. México. Recuperado de: <https://es.scribd.com/doc/8447321/Sampieri-Metodologia-Inv-Cap-10-Analisis-de-Los-Datos>

3.3.5 Representación gráfica de los datos

Es la representación de los datos estadísticos mediante conceptos de longitud, área, volúmenes auxiliados por medio de figuras geométricas y sus propiedades, con el apoyo de los sistemas de coordenadas. También pueden acompañarse con el uso de rayados, sombreados o con colores para resaltar alguna parte en especial.

Se usan para enfatizar la presentación de datos estadísticos; cuando se quiere resaltar determinado hallazgo se acompaña de la tabla ya que los gráficos no dan exactitud, sino consistencia y forma, en cambio la tabla si ofrece la precisión con aproximación exacta de datos y estimaciones (Grisales: 2010)⁴¹.

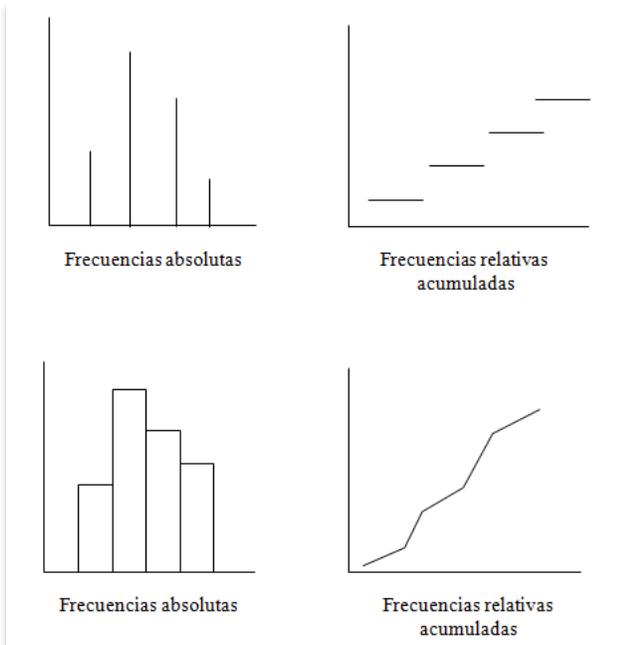
Los gráficos pueden ser de varios tipos:

- Diagramática en donde se incluyen las gráficas de puntos, las lineales (polígonos), gráfico de barras, circulares.
- Estereometría incluye cúbicos, piramidales, prismáticos.
- Pictogramas.

⁴¹ Grisales Romero Hugo. Estadística Aplicada en Salud Pública: Estadística Descriptiva y Probabilidad. Editorial L-Vieco e Hijas. 2002.



- Cartografías incluye mapas estadísticos.



Cuando las variables son discretas la representación se hace mediante Diagramas de frecuencias, estas frecuencias pueden ser absolutas o relativas acumuladas. En las variables continuas su representación gráfica se hace mediante Histogramas de frecuencias cuando se utilizan frecuencias absolutas.⁴² Si la frecuencia es relativa acumulada se utiliza una ojiva ascendente para graficar. Si en el histograma de frecuencias se unen los puntos medios de la parte superior de cada rectángulo se obtiene el Polígono de frecuencias.

Fuente: Hernández, R. (2010). Tipos de Curvas de Frecuencias (Figura 3).. Capítulo 10 Análisis de datos Cuantitativos. Fundamentos de la Metodología de la Investigación. México. Recuperado de: <https://ivanacal.files.wordpress.com/2013/05/distribuciones-de-frecuencias.pdf>

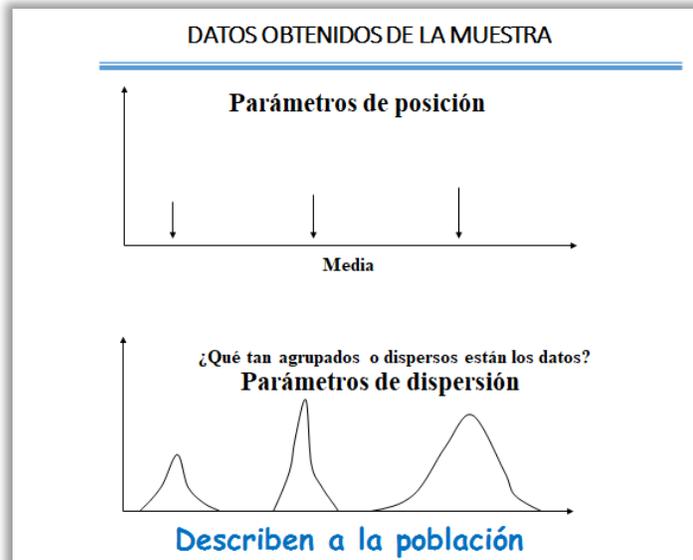
La estadística descriptiva utiliza diversas medidas para realizar la descripción de un fenómeno, denominándose estadísticas cuando se trabaja con muestras; o parámetros cuando se trabaja con poblaciones completas (Kuehl: 2001)⁴³. Estas medidas se pueden presentar en distribuciones unidimensionales, bidimensionales o pluridimensionales.

Las estadísticas se pueden clasificar de varias formas:

- Posición.
- Dispersión.
- Asimetría.
- Curtosis.

⁴² IBIDEM

⁴³ Kuehl R., Diseño de experimentos. Segunda edición. Thomson Learning. 2001



Fuente: Blalock, h. (1986): Estadística Social''. Fondo de Cultura Económica. México.1986. Retomado de <https://es.scribd.com/document/356083141/15MIS-Blalock-1-Unidad-4>

3.3.6 Estadísticas de tendencia central

En un conjunto de datos, las medidas de posición indican el lugar o posición relativa de un valor de la variable, a la cual tienden la mayoría de datos. En este grupo se incluyen la media aritmética, la mediana, la moda o modo, cuartiles, deciles, y percentiles (Marques de Cantú: 1991). ⁴⁴Se acostumbra llamar de tendencia central porque en un buen número de las distribuciones de datos se concentran sus valores en el centro, pero no es lo general, ellas pueden colocarse en cualquier punto del dominio de la variable.

ESTADÍSTICOS
DESCRIPTIVOS

**MEDIDAS DE
TENDENCIA CENTRAL Y
DE DISPERSIÓN**

Tendencia central:
posición de la medida
de una observación
(encuestado)

Dispersión:
Dispersión de la medida

Cuando se habla de medida en estadística se hace referencia a:

- La media
- La mediana
- La moda

Es particularmente importante la **media**. Sobre ésta gira todo el desarrollo de las **pruebas estadísticas de significancia** para la **prueba de hipótesis**.

⁴⁴ Marques de Cantú, María José. Probabilidades y Estadística. Para Ciencias Químico-Biológicas. Editorial McGraw-Hill 1991



La mediana en un conjunto ordenado de datos (de mayor a menor o viceversa) es el valor que divide el conjunto de valores en 2 partes iguales. No es afectada tan drásticamente por los valores extremos (Quesada: 1989)⁴⁵.

Cálculo de la mediana en datos no agrupados:

- Número impar de observaciones.

Ejemplo: En un listado ordenado de pulsaciones por minuto de 15 estudiantes: 62 64 65 66 68 70 71 71 72 72 80 80 80 80 83

La mediana corresponde al valor **71** que ocupa el puesto N. 8; es decir divide en 2 la distribución.

- Número par de observaciones: la mediana es igual al promedio aritmético de los dos valores centrales de la distribución, es decir la suma de los dos valores dividido por dos.

Ejemplo: En un listado ordenado de pulsaciones por minuto de 14 estudiantes 62 64 65 66 68 70 71 72 72 80 80 80 80 83

La mediana corresponde a $a = 71 + 72 / 2 = 71.5$

Cálculo de la mediana en datos agrupados

Obtener las frecuencias absolutas acumuladas.

Obtener el punto medio de las observaciones, a través de $N/2$ (N= número total de observaciones).

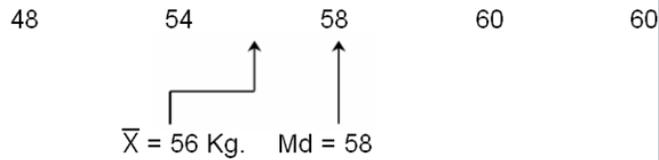
Localizar el resultado de $N/2$ en la columna de frecuencias acumuladas, buscar el valor de la variable que corresponde.

$N/2 = 57/2 = 28.5$, se busca en la columna de la frecuencia acumulada, la posición 28.5 y se encuentra en la tercera fila, por lo tanto el valor que se encuentra en la posición de la mediana es el 1.5.

⁴⁵ Quesada v., Isidoro, López. (1989) Curso y ejercicios de estadística, ed. Alhambra



Ejemplo: Mediana



Note que si el último peso en lugar de 60 Kg. hubiese sido 80 Kg. la mediana no se modificaría, mientras que la media aumentaría considerablemente.

VENTAJAS:

Es una medida fácil de calcular y no está afectada por los datos extremos.
Hay ocasiones en que es la única medida de tendencia central que puede calcularse; como cuando las distribuciones no tienen definidos los límites extremos.

DESVENTAJAS:

No es tan conocida como la media aritmética.
Es necesario ordenar los datos.
No permite cálculos matemáticos posteriores.
La mediana no se afecta con cambios de valores de los elementos que componen la distribución.

La moda

Es el valor de la variable que se presenta con mayor frecuencia. Una serie de datos puede tener una sola moda, dos modas, tres modas etc. o no tener ningún valor modal. (Milton: 2001)⁴⁶. En el ejemplo anterior sobre las pulsaciones la moda es igual a 80, es el valor que más se repite. La moda es muy útil en variables cualitativas, así en el ejemplo sobre tipo de transporte de los estudiantes la moda es la bicicleta. En distribuciones de variables discretas, por ejemplo:

⁴⁶ Milton, J. Susan (2001). Estadística para la Biología y Ciencias de la Salud. Editorial McGraw-Hill. Tercera Edición Española.



Figura 5. Tablas de frecuencia

VARIABLE	FRECUENCIA
0	8
1	12
2	30
3	20
4	10
TOTAL	80

Fuente: Hernández, R. (2010). Diseño de tablas (Figura 4).. Capítulo 10 Análisis de datos Cuantitativos. Fundamentos de la Metodología de la Investigación. México. Recuperado de: <https://es.scribd.com/doc/8447321/Sampieri-Metodologia-Inv-Cap-10-Analisis-de-Los-Datos>

La moda, el valor de la variable que más se repite es el 2. Si la distribución es de variables continuas tenemos. Si la distribución es de variables continuas tenemos:

VARIABLE	CLASE	
6.1 - 16	11	6
16.1 - 26	21	14
26.1 - 36	31	22
36.1 - 46	41	38
46.1 - 56	51	26
56.1 - 66	61	14
66.1 - 76	71	10
TOTAL	-	130



La moda, el valor de la variable que más se repite corresponde a la marca de clase 41 (punto medio del rango de la variable).

VENTAJAS:

Se puede considerar la mejor medida de tendencia central ya que indica el punto de mayor concentración de datos. En una distribución asimétrica, la Moda es la medida más representativa del grupo y si son muy diferentes la media aritmética y la Moda es preferible utilizar esta última.

En series polimodales (varias modas), la moda permite dividir la distribución con fines de estratificación.

DESVENTAJAS:

Es difícil calcular la moda en una serie agrupada y las aproximaciones de su cálculo no son de mucha confianza.

No puede ser usado fácilmente en procesos algebraicos posteriores.

No es sensible a cambios de valores de la distribución, a menos que se afecte su propio valor.

Media aritmética

Es un único valor de la variable que se obtiene de sumar todos los valores de la serie y dividir por el total de ellos (Wayne:1998) ⁴⁷. Es el estadístico de posición más conocido.

Media Aritmética

- Cuando se calcula en una población por N individuos se simboliza por μ "mu" y viene dada por la fórmula.

$$\mu = \frac{(x_1 + x_2 + \dots + x_N)}{N} = \sum_{i=1}^{i=N} x_i$$

X_i = Valor de la observación i (cada individuo de la pob.)

N = Tamaño de la población

Imagen 6. Fuente: Blalock, h. (1986): Estadística Social''. Fondo de Cultura Económica. México.1986. Retomado de <https://es.scribd.com/document/356083141/15MIS-Blalock-1-Unidad-4>

⁴⁷ Wayne W. Daniel. (1998). Bioestadística: Base para el análisis de las ciencias de la salud. Noriega editores. Editorial Limusa. Tercera edición. México.



Media Aritmética.

- Cuando se calcula en una muestra con n individuos, es el estimador de la media μ de la población, y se simboliza por "x barra" y viene dada por la fórmula:

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

X_i = Valor de la observación i (cada individuo encuestado)

n = Tamaño de la muestra

Ejemplo: Media Aritmética

- Supongamos una muestra de 5 chicos cuyos pesos al cumplir 15 años son: 48, 54, 58, 60, 60 kg. La media de estos pesos vale:

$$\bar{X}_M = \frac{48 + 54 + 58 + 60 + 60}{5} = 56 \text{kg}$$

Imagen 7. Fuente: Blalock, h. (1986): *Estadística Social*. Fondo de Cultura Económica. México.1986. Retomado de <https://es.scribd.com/document/356083141/15MIS-Blalock-1-Unidad-4>

VENTAJAS:

- 1) Esta medida se define en forma rígida por una ecuación matemática muy fácil de entender. En casos en que no se conozcan los valores individuales se puede obtener, por ejemplo 30 personas se ganan \$ 1.800.000, en promedio cada persona se gana \$ 60.000.
- 2) La media aritmética es muy estable en el muestreo.



- 3) Es altamente sensible a cualquier cambio en la distribución.
- 4) Permite cálculos matemáticos posteriores (como promedios ponderados, promedios de promedios)

DESVENTAJAS:

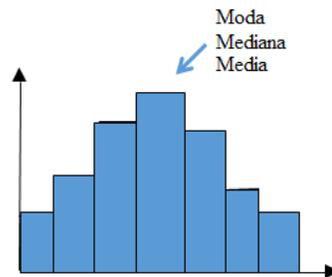
Es afectada por los valores muy grandes o muy pequeños, por lo tanto valores muy extremos de la distribución pueden afectar la representatividad del promedio aritmético con respecto a los valores de la distribución.

En una distribución marcadamente asimétrica en donde el promedio aritmético, la mediana y la moda difieren en forma apreciable, se debe considerar que el promedio no es el único valor representativo de la distribución.

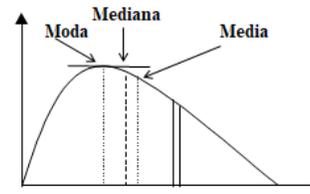
Cuando la distribución tiene forma de U, es decir es parabólica, el promedio corresponde a los valores menos comunes en la serie y por lo tanto da una idea irreal de la distribución.

Figura 7. Distribución Simétrica

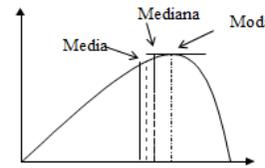
En una distribución simétrica de los datos. La moda, la mediana y la media, coinciden



De otra forma se tienen distribuciones asimétricas



A la izquierda



A la derecha

Fuente: Hernández, R. (2010). Tipos de Frecuencias (Figura 6). Capítulo 10 Análisis de datos Cuantitativos. Fundamentos de la Metodología de la Investigación. México. Recuperado de: <https://ivanacal.files.wordpress.com/2013/05/distribuciones-de-frecuencias.pdf>



3.3.7 Estadísticas de dispersión

Las estadísticas de dispersión son aquellas que describen como se agrupan o dispersan los datos alrededor de un promedio (Walpole: 2007)⁴⁸. Permiten conocer si el promedio representa adecuadamente la distribución considerada, cuando menor sea la dispersión más representativos será el promedio, la mediana o la moda. Miden el grado de homogeneidad de los datos; cuando los datos son iguales las medidas de dispersión son iguales a cero; cuando existe mucha heterogeneidad las medidas de dispersión serán grandes.

Además sirven para el cálculo del tamaño de muestra, a menor variabilidad menos tamaño de muestra requerido. Las estadísticas de dispersión o variabilidad tienen sentido sólo acompañando las medidas de posición o de tendencia central.

En general se consideran como estadísticas de dispersión: El rango, la varianza, la desviación estándar, y el coeficiente de variación relativa.

El rango: es la diferencia entre el dato mayor y el dato menor de la distribución de frecuencias; también es llamado recorrido (Walpole: 1998)⁴⁹. Cuanto más grande sea el rango, mayor será la dispersión de los datos. Se calcula:

$$X_{\text{máx.}} - X_{\text{min}}$$

La varianza: se define como la media aritmética de los cuadrados de las desviaciones respecto a su media. Por lo tanto se expresa en unidades de medidas elevadas al cuadrado

Varianza

Mide el grado de variación de los valores de una variable en la población.

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$

Nótese que el punto de referencia para estimar las desviaciones es la media

Se define como la media de la suma de cuadrados de las diferencias entre cada valor de la variable (X_i) y la media aritmética de la distribución (μ). Cuando se calcula en una población con N individuos, la varianza se simboliza por σ^2 . Si se calcula para una muestra se denota con S^2

⁴⁸ Walpole R., Myers R., Myers S., Ye K. (2007). Probabilidad y Estadística para ingeniería y ciencias. Octava Edición. Pearson, Prentice Hall.

⁴⁹ IBIDEM



$V = \frac{\sum (X - M)^2}{N}$ Varianza de una población

$S = \frac{\sum (X - \bar{X})^2}{n-1}$ Varianza de una muestra, con datos no agrupados
 $V = \frac{\sum (X - \bar{X})^2}{f/n}$ Varianza de una muestra con datos agrupados

Los resultados obtenidos del cálculo de la varianza son elevados al cuadrado, lo cual es diferente al valor real de variable, se hace necesario calcular una medida más real que nos permita comparar las diferencias; lo que se logra obteniendo la desviación estándar, que es la raíz cuadrada de la varianza.

$$S = \sqrt{V}$$

Varianza

¿Por qué es necesario elevar los valores al cuadrado?

- Es lógico que un índice de dispersión esté basado en las diferencias $x_i - \mu$, ya que estas diferencias serán tanto mayores cuanto más alejadas estén las observaciones x_i de la media μ . Dado que al sumar estas diferencias se anularían entre sí, al ser unas positivas (para valores x_i superiores a μ) y otras negativas (para valores x_i inferiores a μ), en la definición de índice de dispersión debemos quitarle el signo. Una buena manera de lograrlo es elevando estas diferencias al cuadrado. Por este motivo la varianza se define como la **media del cuadrado de estas diferencias**.

La desviación estándar

Es la raíz cuadrada de la varianza y su notación es:

$$S = \sqrt{S^2}$$

Caracteriza la dispersión (o el grado de homogeneidad) de una distribución

Su símbolo es σ cuando se trata de la desviación estándar de la población.

Se mide en la misma escala que la variable.

La desviación estándar significa cuánto en promedio difieren los datos de la distribución con respecto al dato promedio (Grisales: 2002).⁵⁰

⁵⁰ Grisales Romero Hugo. (2002). Estadística Aplicada en Salud Pública: Estadística Descriptiva y Probabilidad. Editorial L-Vieco e Hijas.



Otra estadística de dispersión es el **coeficiente de variación**, se utiliza cuando se quiere comparar la variabilidad que presentan dos series de datos. Se refiere a la comparación de la desviación estándar con respecto al promedio (2008)⁵¹, se expresa en términos porcentuales⁵². Esta medida tiene el inconveniente que presenta resultados diferentes cuando las distribuciones tienen diferentes promedios e iguales varianzas; es decir con igual dispersión tienen diferentes coeficientes de variación.

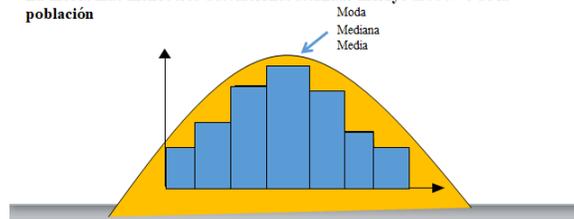
Figura 8. Distribución coeficiente de variación

Si la variable original se distribuye siguiendo la ley normal, entonces la desviación estándar indica lo siguiente:

La media más menos una desviación estándar incluye al 68.3% de la población.

La media más menos dos desviaciones estándar incluye al 95.4% de la población.

La media más menos tres desviaciones estándar incluye al 99.7% de la población



Fuente: Hernández, R. (2010). Tipos de Curvas de Frecuencias (Figura 7).. Capítulo 10 Análisis de datos Cuantitativos. Fundamentos de la Metodología de la Investigación. México. Recuperado de: <https://ivanacal.files.wordpress.com/2013/05/distribuciones-de-frecuencias.pdf>

Medidas de asimetría. Estas estadísticas indican la dirección que toman los datos respecto a un eje. Se dice que una distribución es simétrica si los datos se distribuyen en igual forma a lado y lado del eje (Milton: 2001)⁵³, en cuyo caso la simetría es igual a cero; puede ser positiva si los datos se agrupan a la derecha del eje o negativa si se agrupan a la izquierda.

Curtosis. Es un indicador de lo plana o *picuda* que es la curva. Cuando la curtosis es cero quiere decir que la curva es normal; si es positiva quiere decir que la curva o polígono es más levantada, y si es negativa, la distribución es achatada o plana.

- La distribución normal, se denomina así a la distribución simétrica de datos con su moda, mediana y promedio en el centro; el área bajo la curva corresponde al 100%.

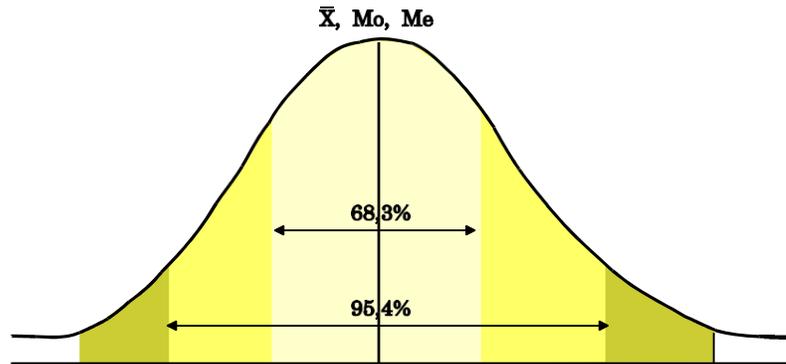
⁵¹ Anderson D., Sweeney D., Williams T. (2008). Estadística para la administración y economía. Décima edición. Cengage Learning.

⁵² Daniel, Wayne W. Bioestadística: Base para el análisis de las ciencias de la salud. Noriega editores. Editorial Limusa. Tercera edición. México. 1998

⁵³ Milton, J. Susan, Estadística para la Biología y Ciencias de la Salud. Editorial McGraw-Hill. Tercera Edición España 2001



Figura 9. Distribución normal



Fuente: Hernández, R. (2010). Tipos de Curvas de Frecuencias (Figura 8).. Capítulo 10 Análisis de datos Cuantitativos. Fundamentos de la Metodología de la Investigación. México. Recuperado de: <https://ivanacal.files.wordpress.com/2013/05/distribuciones-de-frecuencias.pdf>

En esta distribución se debe cumplir: $X \pm 1DE$ contiene el 68.3% de los datos $X \pm 2DE$ contiene el 95.4% de los datos $X \pm 3DE$ contiene el 99.7% de los datos Distribución de puntuaciones Z:

Las puntuaciones Z son transformaciones que se pueden hacer a los valores o puntuaciones obtenidas, con el fin de analizar su distancia respecto a la media, en unidades de desviación estándar (Wisniewski: 2001) ⁵⁴ Esta estandarización de los valores permite comparar distribuciones diferente pero la forma de medición es la misma. Se deben conocer de las distribuciones el promedio y la desviación estándar.

3.3.8 Estadística inferencial

La estadística inferencial se utiliza para dos procedimientos:

- a) Estimar parámetros.
- b) Probar hipótesis.

⁵⁴ Wisniewski P., Velasco G. (2001). Problemario de probabilidad. Thomson Learning.



Estimación de parámetros. Los parámetros no pueden ser calculados, porque no se recolectan todos los datos de la población; por lo tanto se calculan las estadísticas y a partir de ellas se estiman los parámetros (Sarabia: 1993)⁵⁵.

De otra parte, la existencia de las distribuciones muestrales, requiere que para inferir valores poblacionales a partir de estadísticas muestrales, no se acepte de inmediato hacerlo con el valor exacto calculado como promedio (o proporción, o diferencia de promedios, o diferencia de proporciones), sino a partir de un intervalo que contenga con mayor confianza el valor poblacional buscado.

Las estadísticas son estimativos del parámetro, pero en ocasiones se obtienen dos valores que contienen dicho parámetro, el intervalo entre estos dos valores se denomina intervalo de confianza. Un intervalo de confianza son dos valores límites dentro de los cuales se puede esperar se encuentre un parámetro determinado, como por ejemplo la proporción, o el promedio de una población con un determinado nivel de confiabilidad.

Por lo anterior es necesario definir el nivel de confiabilidad (1-alfa) del intervalo de confianza en donde alfa es el nivel de significancia estadística de los datos. Los valores de nivel de confiabilidad más utilizados son 90%, 95%, 99%.

El nivel de significancia se refiere al nivel de certeza fijado por el investigador, la probabilidad que acepta de equivocarse, es decir que lo observado se deba al azar. Para el cálculo del intervalo de confianza, en una distribución en que se conozca el promedio y la desviación estándar se utiliza la siguiente ecuación:

$$IC = \bar{X} \pm Z(1 - \alpha / 2) \cdot dx$$

Cuando se conocen datos de proporciones o tasas:

$$IC = \hat{p} \pm Z(1 - \alpha / 2) \cdot d^{\wedge}p$$

La interpretación del intervalo de confianza se da en los siguientes términos: con una confiabilidad de 1-alfa se estima que el parámetro poblacional desconocido está contenido en los dos valores límites que aparecen al restar y sumar al estadística (promedio, proporción muestral), el producto del valor Z por el respectivo error estándar (Devore: 2008)⁵⁶.

⁵⁵ Sarabia Viejo., Mate Jiménez c. (1993) "problemas de probabilidad y estadística. Elementos teóricos, cuestiones, aplicaciones con statgraphics", ed. Clagsa,

⁵⁶ Devore J. (2008). Probabilidad y Estadística para ingeniería y ciencias. Séptima edición. Cengage Learning.



Prueba de hipótesis. Cuando el investigador necesita decidir con respecto a una población examinando una muestra de ella, en un diseño analítico de casos y controles, cohortes o experimentales utiliza la técnica de prueba de hipótesis. En el proceso investigativo una hipótesis se define como una suposición que se plantea para explicar ciertos hechos o eventos y se emplea como base para desarrollar una investigación, mediante la cual se busca demostrarla o refutarla (rechazar o no rechazar la hipótesis) (Meyer: 1998)⁵⁷

Esta hipótesis se traduce en una de tipo estadístico definida como una suposición acerca de un parámetro o de otro valor estadístico de una población, hipótesis alterna; para hacer la comparación se utiliza la hipótesis de nulidad, la cual supone que no hay diferencia real entre los grupos de valores que se quieren comparar y se supone que las diferencias observadas son debidas a variaciones aleatorias de los datos. Para decidir si se rechaza o no se rechaza la hipótesis de nulidad, se calcula un valor y se compara con otro que se encuentra en tablas estadísticas, las cuales indican la probabilidad de cometer un error al aceptar o rechazar la hipótesis de nulidad. Si el valor calculado es mayor al valor obtenido de las tablas, la hipótesis de nulidad se rechaza y la diferencia encontrada entre los dos grupos de valores se declara estadísticamente significativa.

En los informes clínicos de tipo científico se acostumbra señalar el nivel más bajo de las tablas con el que fue rechazada la hipótesis de nulidad; este valor se señala con una letra p, seguido del valor de las tablas $p < 0.01$ que quiere decir: existe una probabilidad $<$ del 1% que las diferencias observadas entre las dos series de valores se deban a variaciones aleatorias de los datos; es decir tienen una significancia estadística del 1% es decir existe el 99% de probabilidad de ser correcto el rechazo de la hipótesis nula (Ritchey: 2002)⁵⁸.

Errores en la prueba de hipótesis

- **ERROR TIPO I:** A veces las diferencias observadas entre los promedios o proporciones de dos muestras pueden ser lo suficientemente grandes para rechazar la hipótesis de nulidad, pero esta diferencia puede ser sin embargo debida al azar. El rechazar la hipótesis de nulidad cuando esta es cierta se conoce como error tipo I, tipo alfa (α) o nivel de significancia estadística. Por tanto entre más pequeño sea el nivel de significancia utilizado, mayores serán las probabilidades de haber rechazado una hipótesis nula que era falsa.

⁵⁷ Meyer P. (1998) Probabilidad y Aplicaciones estadísticas. Edición revisada. Addison Wesley Logman.

⁵⁸ Ritchey F., (2002). Estadística para las ciencias sociales. Mc Graw Hill.



NATURALEZA	DECISIÓN DEL INVESTIGADOR	
	ACEPTACIÓN	RECHAZO
HO: VERDADERA	Correcto	Incorrecto Error I (Alfa α)
HO: FALSA	Incorrecto Error II (Beta β)	

- ERROR TIPO II:** También por azar, puede ocurrir que las muestras presenten diferencias pequeñas que concuerden con lo previsto por la hipótesis de nulidad y al observar los resultados se acepte como cierta la hipótesis de nulidad aunque en realidad se trate de poblaciones que difieren en el parámetro medido. Esta situación se conoce como error tipo II o tipo beta (β).

Para la prueba de hipótesis debe conocerse la naturaleza de los datos que son la base de los procedimientos, ya que esto determina la prueba particular a emplearse, se debe especificar si los datos se refieren a conteos o medidas. (Steel: 1988)⁵⁹

Procedimiento para la prueba de hipótesis estadística.

En el proceso de prueba de hipótesis se trabaja con dos hipótesis que deben enunciarse explícitamente. Se utiliza el siguiente procedimiento:

- Precisar las hipótesis: **HIPÓTESIS NULA:** Ho: se plantea en términos de igualdad; es la hipótesis que se rechaza o no. **HIPÓTESIS ALTERNA:** Ha: se plantea en términos de diferencias; mayor o menor. Define si la prueba es de 1 o 2 colas.
- Definir el nivel de significancia: valor de alfa, se refiere a la máxima probabilidad especificada con el fin de hacer mínimo el error tipo I. Generalmente se fija antes de seleccionar la muestra puede tener valores de 1%, 5% o 10%; corresponde a un área bajo la curva normal denominada región crítica o zona de rechazo. Este dato sirve para obtener el dato tabular.
- Seleccionar la prueba de significancia a emplearse. El tipo de prueba depende del diseño que se utilice, de acuerdo a la hipótesis planteada. Puede ser:
 - Prueba Z de la curva normal (para muestras grandes mayores de 30) o prueba de T student muestras pequeñas menores o iguales a 30 individuos; Prueba de

⁵⁹ Steel R., Torrie J. (1998). Bioestadística. Segunda edición. Mc Graw Hill.



correlación y regresión.

d) Distribución de la estadística de la prueba, la gráfica puede ser de una o dos colas.

e) Construir la regla de decisión

$Z_c < 0 = Z_t$ No rechazo Hipótesis nula (H_0) $Z_c > Z_t$ Rechazo Hipótesis nula (H_0).

El valor de Z_t se obtiene de acuerdo al nivel de significancia, generalmente es de 5%.

* Si la hipótesis nula no se rechaza se dice que los datos sobre los cuales se basa la prueba, no proporcionan evidencia suficiente que provoque el rechazo.

f) Realizar el estudio. La estadística de prueba calculada se compara con las regiones de rechazo o no.

g) Calcular el nivel de probabilidad (valor de p).

h) Decisión estadística: consiste en el rechazo o no de la H_0 .

i) Decisión administrativa o clínica depende de la decisión estadística.

Prueba de Chi Cuadrado. El *chi cuadrado* es una prueba no paramétrica, es decir, puede ser utilizada en distribuciones no normales (Montgomery: 2006)⁶⁰. Las variables son generalmente discretas y categóricas; se utiliza para evaluar hipótesis de asociación entre dos variables, para su cálculo se utilizan tablas de contingencia.

Procedimiento para calcular Chi Cuadrado:

Planteamiento de hipótesis: H_0 : A y B son independientes. H_a : A y B están asociados.

Nivel de significancia (alfa) y grados de libertad (L-1) (K-1) de acuerdo al número de filas y columnas de las tablas de comparación

L: # de filas; K: # de columnas.

c) Distribución de la estadística de prueba.

d) Regla de decisión.

$X^2_c < 0 = X^2_t$ No rechazo H_0 . $X^2_c > X^2_t$ rechazo H_0 .

e) Cálculo de estadística de prueba: $X^2_c = \sum \frac{(O-E)^2}{E}$.

a b Y1

c d Y2

X1 X2 n

Con los totales horizontales y verticales de los datos observados, se calculan los datos esperados en cada una de las casillas:

⁶⁰ Montgomery D. (2006)., Diseño y análisis de experimentos. Segunda edición. Limusa Wiley.



El valor esperado en a: (X_1Y_1/n) El valor esperado en b: (X_2Y_1/n) El valor esperado en c: (X_1Y_2/n) El valor esperado en d: (X_2Y_2/n)

- f) Decisión estadística.
- g) Decisión administrativa o clínica.

Cuando el diseño de la investigación se basa en una muestra aleatoria simple, se utiliza análisis simple en el cual puedo utilizar la tabla de 2x2 con variables dicotómicas, para buscar probable asociación. Si se utiliza diseño pareado para controlar variables de confusión o cuando se estratifican las variables se debe utilizar el análisis estratificado; cuando el número de variables a estudiar es igual o mayor a dos se utiliza el análisis multivariado.

Análisis multivariado. Este tipo de análisis se utiliza para distribuciones bidimensionales, dos variables, para determinar si existe alguna relación entre ellas y cuantificar dicha relación; estas variables pueden ser ambas discretas, continuas o una discreta y la otra continua (Steel:1988)⁶¹. Para su utilización se requiere el uso de paquetes estadísticos.

Se llama X_i la primera variable (los valores varían de 1 hasta n), Y_i la segunda variable (los valores varían de 1 hasta n). Siempre se tomaran pares de observaciones (X_i, Y_i) .

Los datos de la distribución bidimensional pueden representarse gráficamente en un par de ejes coordenados. Tomando el eje de las abscisas para la primera variable (X) y al eje de las coordenadas, para los valores de la segunda variable (Y). En un plano cartesiano se presentan tantos puntos como pares de observaciones se tengan; a cada punto corresponde 1 par de observaciones, a esta representación gráfica se le denomina Diagrama de esparcimiento o nube de puntos.

Regresión. Todo investigador generalmente cuenta con una muestra de observaciones, basado en su análisis le interesa llegar a conclusiones en la población de la cual obtuvo la muestra (Montgomery: 2006)⁶². Por lo tanto cuando en el análisis se va a utilizar el modelo de regresión lineal simple, se debe estar seguro que dicho modelo es al menos una representación aproximada de la población.

El término de regresión se refiere a un modelo matemático que permite unir algunos puntos de la nube o conjunto de puntos mediante un ajuste rectilíneo, parabólico, exponencial o cualquier otra línea que represente el conjunto. Estos puntos pueden estar agrupados alrededor de la línea o presentar diferencias, presentándose errores en el caso que el ajuste no sea el adecuado.

⁶¹ Steel R., Torrie J. Bioestadística. Segunda edición. Mc Graw Hill. 1988

⁶² Montgomery D., Diseño y análisis de experimentos. Segunda edición. Limusa Wiley. 2006



Se dice que la curva que hace mínima la suma de los cuadrados de las desviaciones entre los puntos dados y dicha línea es la mejor.

En otras palabras, si dos variables están relacionadas se busca un procedimiento que permita estimar el valor de una variable para diferentes valores de la otra variable.

La relación existente entre las variables se puede clasificar en la siguiente forma:

- 1) Dependencia causal unilateral: esta relación se da cuando una de las variables influye en la otra, pero no al contrario.
- 2) Interdependencia: se presenta cuando la influencia entre las dos variables es recíproca, es decir una dependencia bilateral.
- 3) Dependencia indirecta: dos variables pueden mostrar una correlación a través de una tercera variable que influye en ellas.
- 4) Concordancia: se presenta en dos variables independientes a las cuales se les determina la correlación que pueda existir.
- 5) Covariación casual: cuando la correlación que se presenta entre las dos variables es totalmente casual o accidental.

Coefficiente de correlación. El coeficiente de correlación r mide la fuerza de asociación lineal entre 2 variables numéricas y su valor puede variar entre -1 y $+1$ (Montgomery:2006)⁶³

Correlación perfecta, cuando $r=1$ o -1

Correlación excelente, cuando r es mayor de 0.90 y menor de 1 o ($-1 < r < -0.90$)
 Correlación aceptable, cuando r se encuentra entre 0.80 y 0.90 . o ($-0.90 < r < -0.80$)
 Correlación regular, cuando r se encuentra entre 0.60 y 0.80 o ($-0.80 < r < -0.60$)
 Correlación mínima, cuando r se encuentra entre 0.30 y 0.60 o ($-0.60 < r < -0.30$)

No hay correlación para r menor de 0.30 o ($-0.30 < r < 0$)

Otra medida que nos ayuda a explicar la relación entre dos variables es el coeficiente de determinación r^2 varía entre 0 y 1 , se interpreta como el porcentaje de variación de Y que es explicado o que está relacionado con el cambio de X . Simplemente se calcula elevando al cuadrado el coeficiente de correlación.

$$Y=a +bx$$

a b coeficientes de regresión.

a coordenada de origen.

b pendiente de la recta.

⁶³ Montgomery D., Diseño y análisis de experimentos. Segunda edición. Limusa Wiley, 2006



Permite hallar el valor de Y conociendo un valor de X (estimadora). Esta ecuación puede ser lineal, cuadrática, exponencial, la anterior se refiere al modelo lineal.

La diferencia entre regresión y correlación es la siguiente:

1. Regresión busca una ecuación para estimar valores de Y (variable dependiente) según unos valores de X (variable independiente).
2. Correlación calcula la fuerza de asociación.

Existen otras técnicas de análisis estadístico para asociación entre variables como el análisis de Varianza, la cual se utiliza para análisis de más de dos variables, de las cuales la variable dependiente debe ser continua y las variables independientes que son discretas. Es un proceso secuencial de adición de variables para calcular los datos de varianza para cada uno de las variables independientes, después de esto se calcula un dato que mide la importancia relativa que tiene cada una de estas.



Cierre de la unidad

¡Felicidades! Has concluido las tres unidades de la asignatura de Métodos de Investigación Cuantitativa. La complejidad y diversidad de la problemática en materia de salud hacen indispensable que los profesionales de la salud de hoy, posean las herramientas necesarias para aportar en la solución de las necesidades que en esta materia demanda la población, y es en este escenario que la epidemiología cobra un rol central. Los métodos de investigación cuantitativa es una herramienta esencial en tu formación académica, pues ahora has logrado adquirir una serie de conocimientos y habilidades que te permitirán mantener una actitud reflexiva y de permanente perfeccionamiento, mediante la generación de conocimientos encausados por su curiosidad científica, que lo inducen a indagar el porqué de las problemáticas clínicas y epidemiológicas.

Para el especialista en la promoción y educación para salud es esencial adquirir los conocimientos y habilidades tendientes a la incorporación del método científico a su quehacer profesional, de esta forma se espera que con los conocimientos adquiridos en la asignatura de métodos de investigación cuantitativa puedas complementar los conocimientos y habilidades que te permitan formular de propuestas de intervención tendientes a que la población con la cual trabaja mejore su calidad de vida y alcance niveles superiores de bienestar.

Por último, recuerda que aprender a investigar no es seguir unos preceptos metodológicos, supone la conjunción de por lo menos tres aspectos. En primer lugar es menester desarrollar una actitud permanente de observación, curiosidad, indagación y crítica de la realidad con el fin de encontrar nuevas maneras de resolver los problemas con los cuales nos enfrentamos en la cotidianidad. Pero ante todo se requiere, en segundo lugar, una sólida formación general, un creciente dominio de conocimientos sobre un área específica de la realidad, en particular aquel sobre la cual pretende desarrollar su práctica investigativa y aportar nuevo conocimiento.



Fuentes de consulta

Anderson D., Sweeney D., Williams T. (2008). Estadística para la administración y economía. Décima edición. Cengage Learning.

Argibay, J.C. (2006), "Técnicas psicométricas. Cuestiones de validez y confiabilidad".

Argibay, Juan Carlos. (2009). Muestra en investigación cuantitativa. Subjetividad y procesos cognitivos, 13(1), 13-29. Recuperado en 07 de febrero de 2018, de http://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S1852-73102009000100001&lng=es&tlng=es.

Aramburú, Carlos Eduardo (2001). Métodos y técnicas de investigación social. Gerencia social. Diseño, monitoreo y evaluación de proyectos sociales. Lima-Perú: Universidad del Pacífico. ISBN 9972-603-32-6..

"Introducción al Diseño de una Encuesta". Disponible en : <http://www.estadistica.mat.uson.mx/Material/etapas.pdf>

Arnau Gras, J. (1982), Psicología experimental, México, Trillas

Bello, León Darío. (2000). Estadística como apoyo a la Investigación. Editorial L.Vieco e Hijas Ltda

Bonilla Castro, Elsy y Rodríguez Sehk, Penélope. Más allá del dilema de los métodos. La investigación en ciencias sociales. 3ª Ed. Santafé de Bogotá, Ediciones Uniandes, 1997. Cap. 4. El proceso de la investigación cualitativa. Pág. 119-145.

Castro Posada. J.A. (1989). Técnicas de investigación en las ciencias del comportamiento. Salamanca: Universidad Pontificia.

Cordero R. (2005) "El mundo de las encuestas. Reflexiones sobre su desarrollo e importancia" (2005), Universidad Diego Portales.

Cortada de Kohan, N. (1994), Diseño estadístico, Buenos Aires, Editorial Universitaria

Craig, J.R. y Metze, L.P. (1982). Métodos de la investigación psicológica. México: Interamericana (Ed. Original 1979)



Devore J. (2008). Probabilidad y Estadística para ingeniería y ciencias. Séptima edición. Cengage Learning. En: Subjetividad y Procesos Cognitivos, 8, 15-33.

Gardner, R.C. (2003), Estadística para Psicología usando SPSS para Windows,

Gregorio Rodriguez Gomez, javier Gil Flores, Eduardo Garcés Jimenez, Metodología de Investigación Cualitativa, Editorial Algibe, 1999, Pág. 205

Grisales Romero Hugo. (2002). Estadística Aplicada en Salud Pública: Estadística Descriptiva y Probabilidad. Editorial L-Vieco e Hijas.

Hernández Sampieri, Roberto; Carlos Fernández Collado y Pilar Baptista Lucio. (2010) Fundamentos de la Metodología de la Investigación.— México: McGraw-Hill / Interamericana de España, S.A.U. 613 p.

Joberg S. (2002), Metodología de la Investigación social. Editorial Trillas, México. 2002

Kerlinger, F.N. (1988). Investigación del comportamiento. México: McGraw-Hill

Larios Osorio, Vicente (2001) "¿Cómo hacer una encuesta?". U.A.Q. México

León, O.G. y Montero, I. (2003), Métodos de investigación en psicología y educación. Madrid, McGraw-Hill/Interamericana de España.

Meyer P. (1998) Probabilidad y Aplicaciones estadísticas. Edición revisada. Addison Wesley Logman.

Murdock, George P. (1994), Guía para la clasificación de los datos culturales, Editado por la Universidad Autónoma Metropolitana, México, 2ª edición.

Polit, D.F. y Hungler, B.P. (1997), Investigación científica. En ciencias de la salud,

Quesada v., Isidoro, López. (1989) Curso y ejercicios de estadística", ed. Alhambra

Ritchev F., (2002). Estadística para las ciencias sociales. Mc Graw Hill.



Sarabia Viejo., Mate Jiménez c.. (1993) "problemas de probabilidad y estadística. Elementos teóricos, cuestiones, aplicaciones con statgraphics", ed. Clagsa,

Steel R., Torrie J. (1998). Bioestadística. Segunda edición. Mc Graw Hill.

Thiebaut, Carlos (1999) Conceptos fundamentales de la Filosofía, Alianza Editorial, Madrid.

Walpole R., Myers R., Myers S., Ye K. (2007). Probabilidad y Estadística para ingeniería y ciencias. Octava Edición. Pearson, Prentice Hall.

Wayne W. Daniel. (1998). Bioestadística: Base para el análisis de las ciencias de la salud. Noriega editores. Editorial Limusa. Tercera edición. México.

Wisniewski P., Velasco G. (2001). Problemario de probabilidad. Thomson Learning.

Zorrilla S., Torres M., Luiz A., Alcino P (2000). Metodología de la Investigación. Mc Graw Hill, México.