ESTADÍSTICA DESCRIPTIVA

~ DAR SENTIDO A LOS DATOS ~

COMUNICACIÓN CIENTÍFICA

Ciro López Mendoza

ESTADÍSTICA DESCRIPTIVA

Dar sentido a los datos

COMUNICACIÓN
CIENTÍFICA

Ediciones Comunicación Científica se especializa en la publicación de conocimiento científico de calidad en español e inglés en soporte de libro impreso y digital en las áreas de humanidades, ciencias sociales y ciencias exactas. Guía su criterio de publicación cumpliendo con las prácticas internacionales: dictaminación de pares ciegos externos, autentificación antiplagio, comités y ética editorial, acceso abierto, métricas, campaña de promoción, distribución impresa y digital, transparencia editorial e indexación internacional.

Cada libro de la Colección Ciencia e Investigación es evaluado para su publicación mediante el sistema de dictaminación de pares externos y autentificación antiplagio. Invitamos a ver el proceso de dictaminación transparentado, así como la consulta del libro en Acceso Abierto.

www.comunicacion-cientifica.com

DOI.ORG/10.52501/cc.283

COMUNICACIÓN CIENTÍFICA PUBLICACIONES ARBITRADAS
HUMANIDADES, SOCIALES Y CIENCIAS

CC+I
COLECCIÓN
CIENCIA e
INVESTIGACIÓN

ESTADÍSTICA DESCRIPTIVA

Dar sentido a los datos

Ciro López Mendoza

COMUNICACIÓN
CIENTÍFICA

López Mendoza, Ciro

Estadística descriptiva: dar sentido a los datos / Ciro López Mendoza. — Ciudad de México: Comunicación Científica, 2024. (Colección Ciencia e Investigación).

188 páginas: gráficas; 27 × 21 centímetros

DOI: 1052501/cc.283

ISBN: 978-607-2628-61-8

1. Estadística descriptiva. 2. Investigación social. 3. Método estadístico.

LC: QA276.18 L67 DEWEY: 519.5 L67

Primera edición en Ediciones Comunicación Científica, 2025
Diseño de portada: Francisco Zeledón • Interiores: Guillermo Huerta
Ediciones Comunicación Científica, S. A. de C. V., 2025,
Av. Insurgentes Sur 1602, piso 4, suite 400,
Crédito Constructor, Benito Juárez, 03940, Ciudad de México,
Tel.: (52) 55-5696-6541 • Móvil: (52) 55-4516-2170
info@comunicacion-cientifica.com • www.comunicacion-cientifica.com
comunicacioncientificapublicaciones @ComunidadCient2

ISBN 978-607-2628-61-8
DOI 10.52501/cc.283

Esta obra fue dictaminada mediante el sistema de pares ciegos externos. El proceso transparentado puede consultarse, así como el libro en acceso abierto, en https://doi.org/10.52501/cc.283

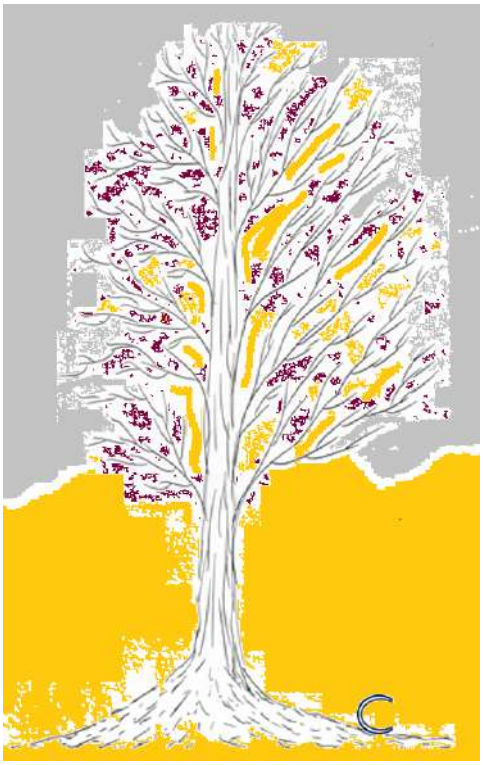


Imagen de CLM 2019

Índice

Resumen

Este libro te lleva de la mano en el estudio de conceptos básicos de estadística descriptiva a través un método sencillo y práctico denominado DPI (definición-procedimiento-interpretación) justo para aplicar con orden y estructura el método estadístico dentro de un proceso de investigación social. Se estudian las cinco etapas fundamentales del método estadístico, es decir, recolectar, contar, presentar, describir y analizar. En este sentido, profundiza de manera simple y clara en conceptos como variables y sus niveles de medición, cálculo de frecuencias, elaboración de tablas y gráficos. Particularmente se enfoca en la etapa de descripción donde se define la selección de medidas estadísticas aplicables, tanto a variables cualitativas como cuantitativas ya sea en series simples como series agrupadas. Además, se estudian los elementos indispensables para el cálculo del tamaño de la muestra; así como los tipos de muestreo generalmente utilizados.

Palabras clave: *investigación social, estadística descriptiva, método estadístico, muestreo, trabajo social.*

Imagen de CLM 2019

Dedicatoria

A todos los alumnos que me han acompañado en un espacio de formación presencial o virtual que me permitieron ser parte de su historia, gracias por tantas satisfacciones, por tantos rostros de "lo entendí".

A todos mis profesores que dedicaron su tiempo para formarme, y a los que dedican y entregan su vida a forjar profesionistas comprometidos con su persona, su profesión y con la sociedad.

Al Dr. Jesús Reynaga Obregón por su invaluable contribución a mi formación y al estudio de la estadística.

A la Dra. Rosario Silva Arciniega por proponerme como candidato a profesor, experiencia que hoy se sintetiza y materializa en este libro.

A quienes agradecen al "otro" a través de la acción de compartir.

A Jeny V. y Daniel A. por estar siempre, por estar aquí y por estar ahora. ¡Gracias!

Imagen de CLM 2019

Prólogo

El resultado de un esfuerzo siempre debe agradecerse a quien de alguna manera contribuyó a su logro, al profesor o al maestro que son uno y el mismo; el primero como la persona que ejerce o enseña una ciencia o arte, y el segundo como la persona que es práctica en una materia y la maneja con desenvoltura. Ambos preceptos describen lo que para mí es y significa el Dr. Jesús Reynaga Obregón, a quien tuve la oportunidad de conocer en la Escuela Nacional de Trabajo Social (ENTS) de la Universidad Nacional Autónoma de México (UNAM) hace algunos años, y quien, sin saberlo, despertó en mí el interés por la estadística. Debo agradecer la fineza para abordar los temas y sobre todo el detalle y la simpleza –sin soslayar la dificultad y el rigor del cálculo– para alcanzar resultados cuyo único fin era describir un contexto determinado –algo que aquí he dado en llamar método DPI (definición, procedimiento e interpretación), saber qué es, conocer y desarrollar el método de cálculo y darle sentido a los valores obtenidos–. Lamento mucho que este prologo no lo haya escrito él, espero haber interpretado de manera correcta sus aportaciones a la estadística y contribuir con otra partícula de arena a través de este libro. Si hay que reconocer a alguien la estructura e idea original de este libro es al Dr. Reynaga.

Lograr formas de abordar la estadística de manera tal que el resultado obtenido o valor resultante se vuelva eje principal del estudio de un problema –incluso que *hable* la estadística dado que *nos dice cosas*– es el mayor reto del presente trabajo, articular argumentos basados en procedimientos estadísticos se vuelve la premisa fundamental. Que esos argumentos sean generados por quienes hoy y ayer me permiten acompañarlos en ese pequeño e inmenso espacio llamado salón de clases se vuelve el logro más significativo y la razón de ser de todo docente, "que cada uno empiece a construir su propio criterio, su propio camino".

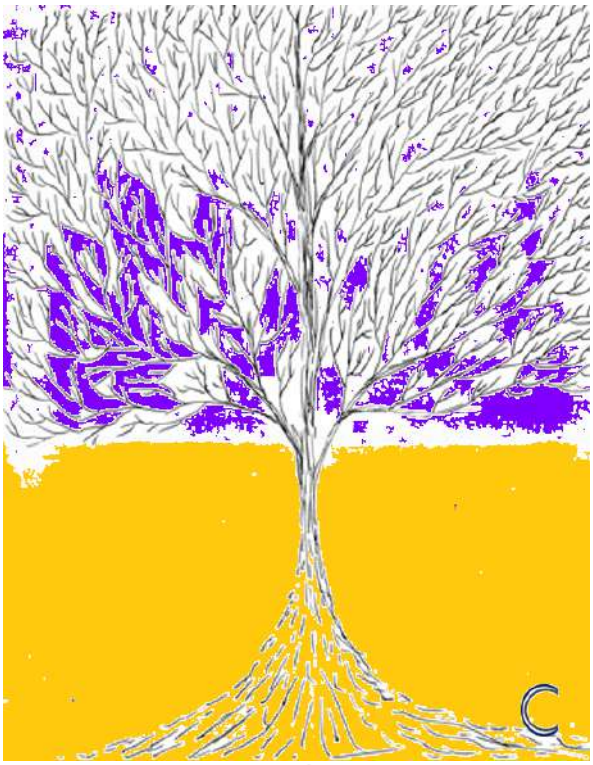


Imagen de CLM 2019

Introducción

El objetivo de este libro es compartir conocimiento relativo a técnicas estadísticas básicas que permitan fortalecer la formación de los profesionales de las ciencias sociales –en particular de trabajadores sociales–, y con ello fundamentar la toma de decisiones en cualquier entorno en el que se desempeñen. Se exponen las principales medidas estadísticas a través del método DPI (definición-procedimiento-interpretación); es decir, bajo un modo ordenado y sistemático de proceder se involucra al lector en el significado de cada concepto, lo lleva a aplicar una serie de fases para el cálculo de una prueba y, finalmente, le da las bases para interpretar los valores obtenidos respecto a una variable de estudio. Ello implica un proceso de entendimiento que parte de fijar con claridad el significado de una palabra (definición), pasa por el cálculo de la prueba (procedimiento) y termina cuando se da razón de ser y referencia al valor obtenido (interpretación).

El estudio de la estadística en este libro comienza con el abordaje de los aspectos de mayor relevancia en dos áreas primordiales, investigación y estadística. Este análisis parte de la definición de ambos preceptos y de su relación e importancia, continúa con el estudio del método estadístico donde se profundiza en cada una de sus cinco etapas; a saber: recolectar, contar, presentar, describir y analizar. Recolectar a través de variables y sus niveles de medición. Contar para determinar frecuencias en las modalidades o clases. Presentar, es decir, definir tablas y gráficos a utilizar para la exposición numérica y visual de los datos. Describir precisa la selección de medidas estadísticas aplicables a variables cualitativas y cuantitativas tanto en series simples como agrupadas. Y analizar implica comparar medidas de resumen.

Finalmente se estudian los elementos indispensables para el cálculo del tamaño de la muestra; así como los tipos de muestreo a utilizar al extraer la misma.

Capítulo 1
Investigación y estadística

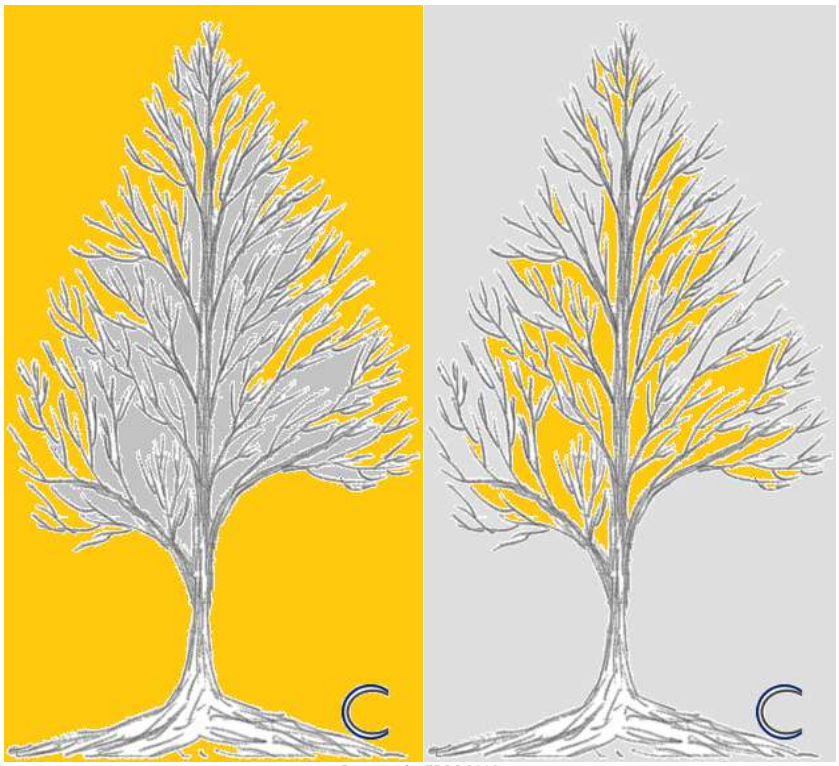


Imagen de CLM 2019

1.1. Investigación

El motor de la investigación es y será *dudar de lo establecido*, *de lo aceptado* y *de la supuesta evidencia* para generar nuevos procesos que refuten o enriquezcan eso que se ha dado como real y verdadero, con el objetivo de expandir el conocimiento y la comprensión del entorno en el que se desarrolla. En otras palabras, expresar *esas dudas* que genera una afirmación, para valorar *otras formas de abordaje* y provocar *otros acercamientos* capaces de crear nuevos saberes en favor de distintos entendimientos.

La ciencia se basa en la contrastación empírica de las teorías con la evidencia, a su vez las teorías se comprueban tratando de demostrar que son falsas; si no se logra esto se retiene la teoría (Elorza, 2008, p. 5). Así, el método de la ciencia –*en tanto de la investigación*–, como señala Popper (1974), es "el de las conjeturas audaces e ingeniosas seguidas por intentos rigurosos de refutarlas" (p. 20).

La etimología del término investigar proviene del latín *in* ('en') y *vestigare* ('hallar', 'inquirir', 'indagar', 'seguir vestigios'); investigar es entonces seguir sistemáticamente la huella, seguir el rastro de los hechos para explicarlos, implica *volver a buscar* dado que es fuente de nuevas líneas de investigación.

La **investigación** "es conjunto de procesos sistemáticos, críticos y empíricos que se aplican al estudio de un fenómeno o problema" (Hernández et al., 2014, p. 4), agregaría *natural* o *social*. Al profundizar en la definición se puede establecer que es un *conjunto* dada la unión de elementos que la conforman; es un *proceso*, ya que los elementos están vinculados en fases sucesivas; es *sistemática* porque las fases están enlazadas o relacionadas entre sí; es *crítica* debido a que se realiza de manera constante un análisis pormenorizado de elementos y fases que la conforman con objeto de evaluar y mejorar el proceso; y es *empírica* porque genera conocimiento desde la experiencia mediante la observación, recolección y el análisis de información.

Arias (2012) define a la **investigación científica** como un "proceso metódico y sistemático dirigido a la solución de problemas o preguntas científicas, mediante la producción de nuevos conocimientos, los cuales constituyen la solución o respuesta a tales interrogantes" (p. 22).

1.1.1. Propósitos de la investigación

La investigación cumple dos propósitos fundamentales: a) producir conocimiento y teorías **(investigación básica)** y b) resolver problemas prácticos **(investigación aplicada)** (Hernández et al., 2014; Cozby, 2004; Selltiz et al., 1980; Castañeda et al., 2002).

Cuando el objetivo de una investigación es acrecentar el conocimiento dentro de una determinada ciencia se dice que se trata de **ciencia pura o básica**. Cuando el objetivo es utilizar los conocimientos, descubrimientos y conclusiones de la investigación básica con fines prácticos o para solucionar un problema concreto se habla de **ciencia aplicada**. Muñoz (2012) profundiza en el tema al señalar que

> la **investigación básica** implica desarrollar teorías por medio del descubrimiento de generalizaciones o principios. Su finalidad consiste en recabar datos empíricos que sirvan para formular, ampliar o evaluar la teoría. Se centra en conocer y explicar. Esta investigación utiliza el experimento, que es la técnica científica de observación más potente, ya que la investigación se realiza en una situación de laboratorio dónde todos los elementos o variables de la situación, al estar creada por el investigador, están controlados. (párr. 4)

La **investigación aplicada**, establece Muñoz (2012),

> pretende la resolución de un problema práctico, inmediato. Se lleva a cabo con relación a problemas reales y en las condiciones en que aparecen (trabajo de campo) sitúa el énfasis en la resolución de un problema concreto, aquí y ahora, en una situación localizada. Se centra en prever o predecir y actuar. Este tipo de investigación comparte con la anterior que en ella también se puede servir del método científico para dar rigor y sistematizar el proceso. (párr. 5)

El origen de la investigación es la interrogante. Su propósito fundamental es generar nuevas preguntas, formar incógnitas recientes, descubrir originales formas de conocimiento, provocar acercamientos creativos y multidimensionales que fusionen y, al tiempo, diluyan las fronteras disciplinarias, para construir entendimientos diversos y diferenciados, y con ello generar otra pregunta e iniciar un nuevo proceso de investigación.

Para beneficio de la humanidad, el vocablo investigación es un concepto de esencia infinita.

1.2. Lo social

Lo social no es una propiedad, ni pertenece a una disciplina concreta, es ante todo una idea compleja y rica en espacios de atención. Desde diversas perspectivas se ha tratado de definir el concepto, por lo cual es importante considerar algunos acercamientos.

Campos (2008) refiere que **lo social** se define como "un sistema constitutivo del ser humano, de relaciones, interacciones y comunicación para la convivencia entre sujetos, que presenta diversos ámbitos en forma dinámica e integral, abarca hasta lo que no parece social, en un momento histórico determinado de la realidad de la familia y las personas vulnerables; asume la cotidianidad; también tiene que ver con la política pública y las instituciones" (p. 67).

Lo social –argumenta Campos (2008)– es definido como los "procesos de relaciones e interacciones dadas a partir de la comunicación y el lenguaje que se manifiestan en significados compartidos entre sujetos. Todas aquellas relaciones que establecen las personas por su condición de seres sociales que hacen la vida humana. Implica la realidad interrelacional e interaccional entre los hombres y la sociedad en la cotidianidad" (p. 67).

Desde la propuesta de Carballeda (2004), citado por Campos (2008), la visión de **lo social** se plantea como algo constitutivo de la vida cotidiana, y requiere considerar la construcción de intercambios y reciprocidades dentro de un grupo de sujetos; se intenta comprender y explicar *lo social* desde la singularidad, centralizando la mirada en las subjetividades de los propios sujetos (p. 60).

En este sentido, **lo social** se define como el conjunto de interrelaciones e interacciones que presentan las personas en lo individual y/o en lo colectivo dentro de un contexto denominado sociedad. La interrelación se refiere a una correspondencia o conexión recíproca que existe entre personas o entre colectivos; por lo tanto, es una relación mutua como persona u organización –en su sentido amplio– dentro de un ente o conjunto de personas que comparten ciertas características y que tienen objetivos en común.

Lo social invariablemente hace referencia a la persona y sus relaciones con el entorno –sociedad– donde se desenvuelve, el cual incluye todo tipo de contacto con sus semejantes, así como, con la diversidad de formas de organización del ser humano.

1.3. Investigación social

La investigación social –plantea Moreno (2017)– "es un ejercicio que ha facilitado auscultar la realidad de los fenómenos humanos y sociales. Ha ampliado los horizontes de comprensión y reestructurado el sentido que se le ha otorgado al ser humano y sus prácticas. Qué es, para qué sirve y cómo se puede hacer ha sido tema de debate desde los albores de la humanidad, pues la respuesta al qué, condiciona la forma o los procedimientos en que se hará y el alcance científico y social que podría tener" (párr. 1).

Ander-Egg (2011) plantea que la **investigación social** se refiere al estudio sistemático de hechos, procesos o acontecimientos que emergen en el entramado social. En el contexto de las metodologías de intervención social —como el trabajo social, la animación sociocultural y la educación— este concepto adquiere una dimensión praxeológica, dado que su propósito es generar conocimiento orientado a la acción transformadora, sustentado en objetivos operativos y finalidades concretas. Asimismo, la concibe como un procedimiento metódico, compuesto por etapas sucesivas y articuladas lógicamente, que exige una continua verificación empírica de los fenómenos examinados (pp. 20, 23).

La investigación social es un proceso en el que se aplican métodos y técnicas al estudio de situaciones-problema de carácter social. Implica la búsqueda de conocimiento respecto a las instituciones, los grupos o las personas a partir del estudio detallado de sus relaciones e interacciones sociales. La investigación social parte del planteamiento de un problema, pregunta o interrogante que requiere ser estudiada en el marco de un proceso metodológico cuyo centro son las personas y sus relaciones e interacciones.

Considerando los conceptos hasta aquí abordados, se entiende por **investigación social** al conjunto de procesos sistemáticos, críticos y empíricos que se aplican al estudio de las interrelaciones e interacciones que presentan las personas en lo individual y/o en lo colectivo dentro de un contexto denominado sociedad. El abordaje según su propósito se puede desarrollar a partir de la investigación básica; o bien, desde de la investigación aplicada.

Desde lo social investigar es adentrarse al estudio o tratado de lo humano, a la constante creación de interrogantes y a la búsqueda incesante de respuestas que minimicen la incertidumbre del *qué*, el *cómo*, el *cuándo*, el *dónde* y, en particular, el *para qué* de la persona, de su ser y existencia en la sociedad de la cual es origen y esencia.

1.4. Estadística

En su informe final denominado *Génesis y evolución histórica de los conceptos de probabilidad y estadística como herramienta metodológica*, Ángel (2006) sostiene que la palabra *estadística* tiene sus antecedentes históricos en los censos, recuentos o inventarios, de personas o bienes, realizados aún antes de Cristo (p. 25).

Ángel (2006) refiere que el vocablo *statistik* proviene de la palabra italiana *statista* y fue utilizado por primera vez por Gotfried A. Achenwall en 1752, quien usó la palabra estadística para esta rama del conocimiento con el significado de "ciencia de las cosas que pertenecen al estado". Achenwal fundó la Escuela de Göttingen y es conocido por los alemanes como el padre de la estadística, reconocimiento que él atribuye a Hermann Conring (1606-1681) (Ángel, 2006, p. 25).

Por otro lado, Ángel (2006) retoma lo indicado por Cortada et al. (1968) al señalar que según estos autores la palabra "estadística" parece derivar de *status*, que en el latín medieval tenía el sentido de "estado político". Usado por primera vez en el Hamlet de Shakespeare. Luego se usó en tratados de política económica y significaba la exposición sistemática y ordenada de las características más notables de un Estado (p. 25). El término latino *status* significa "estado" o "situación"; esta etimología aumenta el valor intrínseco de la palabra, por cuanto la estadística revela el sentido cuantitativo de las más variadas situaciones.

Ángel (2006) enfatiza que según Levin (2024) fue el Dr. E. A. W. Zimmermann quien introdujo el término *statistics* en Inglaterra y que su uso fue popularizado por sir John Sinclair en su obra *Statistical accountof Scotland* (Informe estadístico sobre Escocia, 1791-1799). Sin embargo, mucho antes del siglo XVIII la gente ya utilizaba y registraba datos (p. 25).

Por otra parte, en su artículo "El progreso de la Estadística y su utilidad en la evaluación del desarrollo", Barreto-Villanueva (2012) sostiene que la estadística en la investigación social tradicionalmente ha existido en dos vertientes metodológicas: la cualitativa, que privilegia la recopilación de información sobre valores objetivos medibles a través de técnicas cualitativas como la observación, la entrevista, la participación grupal, los grupos de enfoque, entre otros métodos; y la cuantitativa, que se auxilia de la recopilación y análisis de datos. Actualmente se combinan ambas metodologías, aplicando preceptos estadísticos a las técnicas tradicionales de investigación cualitativa. Precisamente, señala Barreto-Villanueva (2012), uno de los atractivos de la estadística consiste en eso, en su versatilidad de aplicación. Así, los procesos de investigación se han fortalecido y consolidado a lo largo de las últimas décadas gracias, en parte, a los avances tecnológicos que han dotado de potencia el tratamiento informático de los datos, ya sean cualitativos o cuantitativos (párr. 6).

1.4.1 Definición

La estadística se aborda desde diversos preceptos según el autor que se considere. A continuación, se presentan algunas aportaciones que contribuyen al estudio del tema.

Ríus et al. (1998) señalan que la **estadística**

> se ocupa de los métodos y procedimientos para recoger, clasificar, resumir, hallar regularidades y analizar los *datos*, siempre y cuando la variabilidad e *incertidumbre* sea una causa intrínseca de los mismos; así como de realizar *inferencias* a partir de ellos, con la finalidad de ayudar a la toma de *decisiones* y en su caso formular predicciones. (p. 14)

Para Triola (2009) la **estadística** "es un conjunto de métodos para planear estudios y experimentos, obtener datos y luego organizar, resumir, presentar, analizar, interpretar y llegar a conclusiones basadas en los datos" (p. 4).

De acuerdo con Martínez (2019), la **estadística,** concebida en singular, representa un cuerpo sistemático de procedimientos metodológicos, normativos y analíticos orientados a la observación, clasificación, descripción, cuantificación e interpretación del comportamiento colectivo (p. 2).

Si se considera el proceso metodológico empleado por Reynaga et al. López (1996), la **estadística** se puede definir como el conjunto de procedimientos para recolectar, contar, presentar, describir y analizar un conjunto de datos.

En resumen, la **estadística** estudia los métodos y procedimientos diseñados para cubrir dos funciones: describir e inferir.

En este último punto es importante hacer algunas consideraciones dado lo sintético del término. Al definir la estadística a través de las funciones describir e inferir –en ese orden y no al revés– implica que debemos conocer que integra cada término; en ese sentido, al precisar que la estadística es un proceso metodológico primero debemos *describir*; es decir, establecer las características generales de un conjunto de datos lo que nos permite saber qué variables se consideran, cómo es la distribución de frecuencias, cómo se deben presentar los datos y qué particularidades ostentan los mismos, cuyos hallazgos nos permitirán avanzar a una segunda función –o rama de la estadística– *analizar* lo que significa tomar como base las medidas calculadas previamente resultado de la estadística descriptiva para compararlas entre sí o con un contexto determinado en busca de aprobar o no aprobar hipótesis que el investigador se planteé en un determinado de estudio.

1.4.2. Clasificación

Como refieren Flores y Lozano (1998), al clasificar la estadística "es posible hablar de dos términos y metodologías diferentes. Por una parte, si se trata de la recolección, presentación y caracterización de un conjunto de datos que arrojan como resultado la descripción de las diversas características de una población o muestra, tiene lugar una metodología llamada **estadística descriptiva**. Así, esas experiencias, enriquecidas con los conceptos de la teoría de probabilidades, hacen posible la estimación de características de una población, validación de distribuciones o la toma de decisiones sobre algún factor de la población, sin conocerla enteramente y basándose sólo en los resultados de un muestreo. Esta metodología se llama **estadística inferencial**" (p. 2).

La estadística se divide en **estadística descriptiva**, que son los métodos que implican recopilación, caracterización y presentación de un conjunto de datos con el fin de describir varias de sus características; y en **estadística inferencial**, que son los métodos que permiten hacer estimación de una característica de la población o de toma de decisiones respecto a una población, con base solo en los resultados obtenidos de una muestra (Barreto-Villanueva, 2012, párr. 19-20).

De esta manera se puede establecer que la estadística se divide en:

Estadística descriptiva: la rama de la estadística que recolecta, cuenta, presenta y establece las características generales un subconjunto de datos denominado muestra.

Estadística inferencial: la rama de la estadística que establece los métodos y procedimientos para estimar las características de un conjunto total (denominado población), basándose en datos de un subconjunto de observaciones (llamado muestra).

En resumen, como indica Ángel (2006), "se habla de la ciencia de la recolección y análisis de datos para la toma de decisiones, transformando datos en información. Su método comienza con la recolección de datos respecto a un fenómeno; luego mediante la **estadística descriptiva** se resume lo medular de la información. La **inferencia estadística** extiende las conclusiones obtenidas de la muestra a la población de la que ella es parte, además de postular modelos que se ajusten a los datos" (p. 30).

1.4.3. Objeto

Como lo hace notar Holguín (1981), el objeto de la estadística es "resumir los datos más destacados de los elementos que componen un conjunto, logrando así aprehender más fácilmente su contenido" (p. 17).

En la opinión de Ángel (2006), el objetivo de la estadística es el de obtener conclusiones basadas en datos experimentales.

Desde el punto de vista de Chou Ya Lun (1991) la función principal de la estadística es elaborar principios y métodos que nos ayuden a tomar decisiones frente a la incertidumbre.

1.4.4. Utilidad

La estadística se emplea como un valioso auxiliar en los más diversos campos del conocimiento y en las más variadas de las ciencias, tanto básicas como aplicadas; difícilmente se podría encontrar un área de actividad cognitiva en el que la estadística no tenga elementos por aportar. Gorgas, Cardiel y Zamorano (2011) puntualizan la utilidad de la estadística al señalar que resuelve multitud de problemas que se plantean en ciencia:

- **Análisis de muestras.** Se elige una muestra de una población para hacer inferencias respecto a esa población a partir de lo observado en la muestra (sondeos de opinión, control de calidad, etc.)
- **Descripción de datos.** Procedimientos para resumir la información contenida en un conjunto (amplio) de datos.
- **Contraste de hipótesis.** Metodología estadística para diseñar experimentos que garanticen que las conclusiones que se extraigan sean válidas. Sirve para comparar las predicciones resultantes de las hipótesis con los datos observados (medicina eficaz, diferencias entre poblaciones, etc.)
- **Medición de relaciones** entre variables estadísticas (contenido de gas hidrógeno neutro en galaxias y la tasa de formación de estrellas, etc.)
- **Predicción.** Prever la evolución de una variable estudiando su historia y/o relación con otras variables. (p. 4)

La utilidad de la estadística se resume en la aplicación de sus dos ramas fundamentales: la descripción y la inferencia de datos; es decir, la recopilación de información, la determinación de frecuencias, selección y creación de esquemas de presentación, el establecimiento de medidas que reflejen las características generales de un subconjunto de datos, para finalmente a partir de esa descripción de la muestra se establezcan mecanismos para la estimación de datos respecto a la población de estudio.

1.5. Relación e importancia de la estadística y la investigación

La estadística y la investigación se relacionan y al mismo tiempo adquieren valor e importancia porque son:

- Metodologías de constante exploración y descubrimiento.
- Medios para examinar y entender la operación de los fenómenos sociales y naturales.
- Métodos que brindan puntos de vista y procedimientos técnicos que revelan detalles particulares del problema de estudio.
- Principios, procedimientos, técnicas y métodos de carácter universal.

- Bases insoslayables para generar conocimiento.
- Elementos con metodología propia.
- Métodos para la toma de decisiones aún y cuando se asuma cierta incertidumbre.

La relación entre la investigación y la estadística encuentra su razón de ser en la creativa conjunción de métodos, procedimientos y técnicas que permitan adentrarse en los problemas de estudio, y con ello enriquecen el conocimiento que se pudiera tener de los mismos para generar nuevas perspectivas de acercamiento y entendimiento.

Pensar en hacer investigación de tipo cuantitativo implica no soslayar el uso de la estadística y el procesamiento de datos, dado que son la base y principio de la descripción y base para la contrastación de hipótesis.

1.6. El vínculo entre la investigación y la estadística: el método estadístico

La investigación y la estadística cumplen una función fundamental en el estudio de los problemas sociales o naturales; su vínculo y enlace es el método estadístico específicamente si se consideran las características que posee el enfoque cuantitativo de investigación, como lo señala López (2003) basado en el esquema metodológico de Hernández et al. (2014), a partir de las siguientes fases:

1) Concebir la idea a investigar.
2) Plantear el problema de investigación: establecer objetivos; preguntas y justificación.
3) Elaborar el marco teórico: revisión, obtención y consulta de la literatura; extracción y recopilación de la información de interés y construcción del marco teórico.
4) Definir si la investigación se inicia como exploratoria, descriptiva correlacional o explicativa y hasta qué nivel se llegará.
5) **Establecer hipótesis: detectar las variables; definir conceptual y operacionalmente las variables**.
6) Seleccionar el diseño apropiado de investigación: diseño experimental; preexperimental o cuasiexperimental.
7) **Selección de la muestra: determinar el universo y extraer la muestra.**
8) Recolección de los datos: elaborar el instrumento de medición y aplicarlo; calcular validez y confiabilidad del instrumento; codificar los datos.
9) **Analizar los datos: seleccionar las pruebas estadísticas; elaborar el problema de análisis y realizar el mismo.**
10) **Presentar los resultados: elaborar el reporte de investigación y presentar el reporte de investigación.** (p. 29)

La estadística se enmarca en el proceso de investigación al obtener datos. La estadística proporciona elementos teórico-prácticos que permiten estudiar los fenómenos sociales; su principal referente, como lo establece López (2003), es el **método estadístico**, es decir, "la secuencia sistemática de procedimientos para el manejo de los datos derivados de la investigación con el propósito de comprobar hipótesis" (p. 70).

El **método estadístico** se vuelve base fundamental para el estudio sistemático y detallado de un problema y su aplicación secuencial se estructura a partir de cinco etapas, a saber: recolección, recuento, presentación, descripción y análisis. El objetivo primordial es extraer de los datos el máximo de información, y con ello tomar de decisiones con la menor incertidumbre posible.

Capítulo 2
El método estadístico

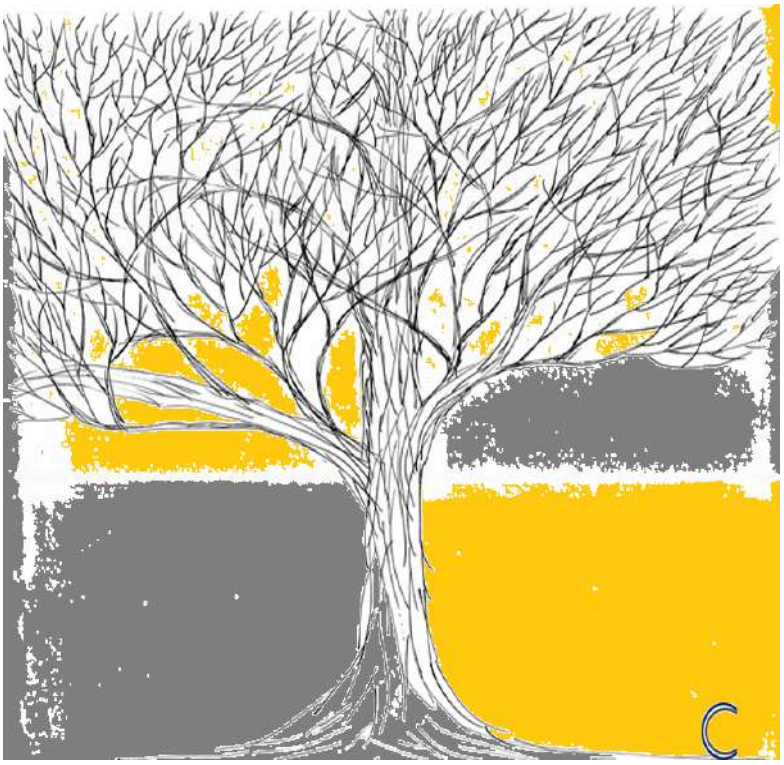


Imagen de CLM 2019

2.1. Concepto

En el estudio de los fenómenos sociales, la investigación y la estadística convergen a través del método estadístico, el cual permite asumir una secuencia sistemática de pasos para ir más allá de simplemente conocer los problemas, es decir, tener bases firmes para la toma de decisiones.

En este apartado se abordan definiciones básicas del método estadístico, en particular, lo relacionado con sus etapas; así como, con la importancia de su seguimiento secuencial con relación al proceso de investigación social.

El método estadístico aporta una referencia lógica para el tratamiento de los datos apoyado en cinco etapas fundamentales: recolectar, contar, presentar, describir y analizar.

El **método estadístico** es la secuencia sistemática de procedimientos para el estudio de los datos cualitativos y cuantitativos derivados de la investigación, con el propósito de rechazar o no una hipótesis.

2.2. Etapas

El método estadístico como conjunto de procedimientos enmarca cinco etapas secuenciales: **(1) recolectar, (2) contar, (3) presentar, (4) describir y (5) analizar.**

2.2.1. Recolectar

En esta etapa se estudia un concepto fundamental para la estadística y la investigación social, tal es el caso de las variables desde su concepto, pasando por su clasificación hasta sus niveles de medición, en este sentido, se entiende por **recolectar** a la etapa donde se acopia la información cualitativa y cuantitativa a través de variables estableciendo sus niveles de medición.

En palabras de Reynaga et al. (1996), **recolectar** es la fase donde

> se recoge la información cualitativa y cuantitativa señalada en el diseño de la investigación. En vista que los datos recogidos suelen tener diferentes magnitudes o intensidades en cada elemento observado (por ejemplo, el peso y la talla de un grupo de personas) a dicha información o datos también se le conoce como variables. Por lo anterior puede decirse que esta etapa del método estadístico consiste en la medición de las variables. (p. 33)

En esta etapa se recupera, acopia, reúne o recolecta la materia prima de la investigación y la estadística; a la vez que se establece su nivel de medición de las variables.

2.2.2. Contar

En la etapa de **contar** se determinan los valores que se obtienen en cada una de las modalidades o clases de la variable a medir; es decir, se define la frecuencia que tiene cada una de las opciones de respuesta de cada variable de estudio. En un procedimiento electrónico se entiende por contar al procedimiento donde los datos son sometidos a revisión –del registro correcto de los datos–, clasificación y cómputo numérico.

Como lo señala Reynaga et al. (1996), el ***recuento*** o el contar

> consiste en la cuantificación de la frecuencia con que aparecen las diferentes características medidas en los elementos en estudio; por ejemplo, el número de personas de sexo mujer y el de personas de sexo hombre o el número de niños con peso menor de 3 kilos y el número de niños con peso igual o mayor a dicha cifra. (p. 34)

2.2.3. Presentar

Presentar significa mostrar de manera numérica o visual la frecuencia de los datos a través de tablas o gráficos. dependiendo del nivel de medición de la o las variables a exponer.

Para Reynaga et al. (1996) **presentar** consiste en

> elaborar cuadros o tablas; así como, gráficos que permitan una inspección precisa y rápida de los datos. La elaboración de cuadros que también suelen llamarse tablas tiene por propósito acomodar los datos de manera que se pueda efectuar una revisión numérica precisa de los mismos. La elaboración de gráficos tiene por propósito facilitar la inspección visual rápida de la información. (p. 34)

2.2.4. Describir

De acuerdo con Reynaga et al. (1996), en la etapa de **describir** "la información es resumida en forma de medidas que permiten expresar de forma sintética las principales propiedades numéricas de los datos" (p. 34).

La condensación de la información, en forma de medidas llamadas de resumen, tiene por propósito facilitar la comprensión global de las características fundamentales de los agrupamientos de datos.

Describir implica el cálculo de medidas estadísticas de resumen que representan las características generales de un conjunto de datos; implica obtener medidas de tendencia central, posición, dispersión y de distribución.

2.2.5. Analizar

En la etapa de **analizar** –establece Reynaga et al. (1996)–, mediante fórmulas estadísticas y el uso de tablas específicamente diseñadas, se efectúa la comparación de las medidas de resumen previamente calculadas, es decir, medidas basadas en una o varias muestras resultado la etapa de descripción.

Analizar implica proporcionar los métodos para estimar las características de un grupo total (población), basándose en datos de un conjunto pequeño (muestra) de observaciones.

Capítulo 3
El método estadístico: recolectar

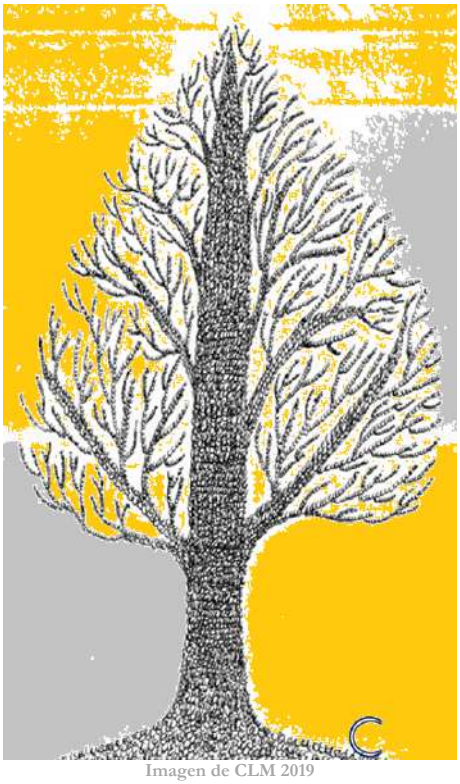


Imagen de CLM 2019

3.1. Recolectar

En la etapa de **recolectar** se acopia la información cualitativa y cuantitativa a través de variables estableciendo su nivel de medición.

Recolectar información implica decidir el tipo de instrumento que se utilizará para recuperar información, además de definir las preguntas que se incluirán en el mismo. Cada pregunta que conforma un instrumento involucra en sí misma una variable y, por tanto, la definición de cuál será su nivel de medición y tratamiento estadístico.

En la recolección de información a través de variables es imperativo valorar qué se va a preguntar y con ello determinar qué aporta cada cuestionamiento a la solución del problema de investigación. Si la pregunta no contribuye a resolver la incógnita estudiada será mejor valorar o replantear la misma, para tener claro si debe o no formar parte del instrumento de recolección de datos.

3.1.1. Variables

Para acercarnos a la definición de una variable cabe preguntarse si un dato ¿presenta cambios en los elementos que lo describen?, ¿sus modalidades o valores se pueden medir?, ¿asume distintas magnitudes o intensidades?, ¿muestra características diferenciadas?, ¿se expresa con atributos diversos?, ¿expone variaciones en las propiedades que admite? Si las respuestas a las interrogantes planteadas son afirmativas, entonces se trata de una variable.

Por otro lado, la palabra variable proviene del latín *variabilis* "que varía o puede variar", inestable, inconstante y mudable. Son características, cualidades, propiedades o atributos que pueden adoptar diferentes valores, magnitudes o intensidades en los diversos sujetos en que se miden. Morán y Alvarado (2010) sintetizan la definición al señalar que una **variable** "es una propiedad que puede variar y cuya variación es susceptible de medirse" (p. 40), Hernández et al. (2014) complementan la definición al agregar que una variable también puede observarse (p. 123).

Al respecto Bologna (2013) llama **variable** a "una característica de las unidades de análisis que puede asumir diferentes valores en cada una de ellas. Se llaman unidades de análisis a los entes individuales acerca de los que se analizan sus cualidades" (p. 19) o características. Por su parte, Ritchey (2004) establece que una **variable** es un "fenómeno medible que varía (cambia) a través del tiempo, o que difiere de un lugar a otro o de un individuo a otro" (p. 18).

Finalmente, Arias (2012) orienta su definición a su aplicación en el proceso de investigación al señalar que una **variable** "es una característica o cualidad; magnitud o cantidad, que puede sufrir cambios, y que es objeto de análisis, medición, manipulación o control en una investigación" (p. 57).

3.1.1.1. Clasificación

La naturaleza de los datos es de gran importancia a la hora de elegir el tratamiento estadístico apropiado para abordar su estudio. En este sentido, las variables se clasifican **estadística y metodológicamente**. Las primeras en consideración a su **nivel de medición**, las segundas debido a su **orden de precedencia**.

a) **Estadísticamente o por su nivel de medición**, las variables se clasifican en: **cualitativas y cuantitativas.**

Variables cualitativas. Este tipo de variables representan una cualidad o atributo que clasifica a cada caso en una de varias categorías. Estas a su vez se clasifican en **nominal u ordinal**.

Variables cuantitativas. Son las variables que asumen valores o magnitudes que pueden medirse o expresarse numéricamente.

En el nivel cuantitativo, medir significa –además de asignar un atributo a una unidad de análisis– saber cuánto mayor o menor está una escala de otra, es decir, específica la distancia o intervalo entre valores (el valor 70 es el doble del valor de 35).

b) **Metodológicamente o por orden de precedencia**, las variables se clasifican en: **independiente y dependiente.**

Variable independiente: es la variable manipulada (el predictor) para determinar sus efectos (predicciones) sobre la variable dependiente. Es la variable de un experimento controlada en forma sistemática por el investigador. Es la causa, origen, razón que establece o controla el investigador. Ritchey (2004) considera que una **variable independiente** o predictora está relacionada o predice la variación de la variable dependiente.

Variable dependiente: es el resultado o variable criterio que está relacionada con cambios en la variable independiente. Esta variable en un experimento es medida por el investigador para determinar el efecto de una variable independiente. Es la consecuencia, efecto o resultado que mide el investigador. Ritchey (2004) señala que la **variable dependiente** es la variable cuya variación se quiere explicar.

Por su parte, Arias (2012) sostiene que las **variables independientes** son

> las causas que generan y explican los cambios en la variable dependiente. En los diseños experimentales la variable independiente es el tratamiento que se aplica y manipula en el grupo experimental. En tanto que las **variables dependientes** son aquellas que se modifican por acción de la variable independiente. Constituyen los efectos o consecuencias que se miden y que dan origen a los resultados de la investigación. (p. 59)

Tabla 1. ***Variable independiente y variable dependiente***

Afirmación 1	
El presupuesto familiar depende de los ingresos de los integrantes	
Variable independiente	**Variable dependiente**
Ingresos de los integrantes	Presupuesto familiar

Afirmación 2	
El método de estudio al que es sometido un grupo de estudiantes aumenta su promedio de calificaciones	
Variable independiente	**Variable dependiente**
Método de estudio	Promedio de calificaciones

Fuente: elaboración propia.

3.1.1.2. Niveles de medición

Por nivel de medición las variables se clasifican en **cualitativas** y **cuantitativas**. Las variables cualitativas a su vez se dividen en **nominales** y **ordinales**, y las variables cuantitativas en **discretas** y **continuas**. Las escalas de las variables cualitativas reciben el nombre de **modalidades**. Las escalas de las variables cuantitativas reciben el nombre de **valor** o **clase**. En este sentido, una variable no es sino el conjunto de las distintas modalidades o clases definidas por una escala.

Nominales: son aquellas en las que los datos se ajustan por categorías que no mantienen una relación de orden entre sí. Significa simplemente asignar un atributo o característica a una unidad de análisis sin importar jerarquía (color, estado civil, ocupación, religión).

De acuerdo con Celis de la Rosa y Labrada (2014) una **variable nominal** es "aquella cuyas características se define por un nombre y no implica ser más o menos (+ *o* -) que la característica definida por un nombre diferente" (p. 4). Rustom (2012), por su parte, señala que una **variable nominal** es "aquella cuyos valores son nombres o códigos sin una relación de orden intrínseco entre ellos" (p. 12).

Tabla 2. ***Ejemplos de variable nominal***

Variable	Escala de modalidades
Tipo de basura	Orgánica Inorgánica
Sexo	Hombre Mujer
Color de ojos	Negro Café Azul Verde
Profesión	Trabajo social Psicología Sociología Pedagogía
Estado civil	Soltero Casado Divorciado Viudo
Nacionalidad	Mexicana Estadounidense Canadiense

Fuente: elaboración propia.

Las modalidades representan la forma en cómo se va a medir una variable, generalmente basadas en la definición conceptual de la misma.

Ordinales: son aquellas en las que existe un orden o jerarquía entre las categorías. Significa asignar un atributo a una unidad de análisis cuyas categorías pueden ser ordenadas en una serie creciente o decreciente.

Para Rustom (2012) una **variable ordinal**, corresponde a "aquella cuyos valores son nombres o códigos, pero con una relación de orden intrínseco entre ellos, es decir, sus valores conllevan un ordenamiento de mejor a peor o de mayor a menor" (p. 12).

Por su parte, Arias (2012) establece que el **nivel de medición ordinal** es "la escala en la que se establece un orden jerárquico entre variables cualitativas o categorías. En esta escala no se indica la magnitud de la diferencia entre las categorías, ni se aplican las operaciones matemáticas básicas" (p. 64).

Celis de la Rosa y Labrada (2014) sostienen que las **variables ordinales** son aquéllas cuyas características pueden recibir algún orden subjetivo. Su característica principal es que, al ser clasificadas de alguna manera, se puede asumir que se es más o menos *(+ o -)* que las otras, aunque se desconozca qué tanto más o qué tanto menos (p. 4).

Tabla 3. ***Ejemplos de variables ordinales***

Variable	Escala de modalidades
Calidad del agua	Buena Regular Mala
Calificación de un servicio prestado	Excelente Bueno Regular Malo
Escolaridad	Primaria Secundaria Preparatoria Licenciatura Posgrado
Puesto	Director general Director de área Subdirector de área Jefe de departamento Jefe de servicio
Lugar de llegada en un maratón	Primero Segundo Tercero

Fuente: elaboración propia.

Discretas: son aquellas que no admiten todos los valores intermedios en un rango. Son el resultado de contar, y asumen valores numéricos enteros.

Para Celis de la Rosa y Labrada (2014) las **variables discretas** "solo se admiten valores individuales en números enteros. Además de que un rasgo distintivo de estas variables es que la unidad no puede fraccionarse, porque pierde su naturaleza" (p. 5).

Calduch (2014) señala que las **variables discretas** corresponden a aquellas propiedades cuya medición solo puede asumir valores específicos dentro de un conjunto numérico limitado, habitualmente compuesto por números enteros (p. 90).

Tabla 4. ***Ejemplos de variables discretas***

Variable	Escala de: valor o clase
Número de hijos	1 2 3
Número de hermanos	1 2 3 4
Edad (en años cumplidos)	19 20 21 22
Número de solicitudes de empleo presentadas	1 2 3
Número de asignaturas no aprobadas	1 2 3 4
Número de consultas otorgadas por un médico	1 2 3
Número de visitas domiciliarias realizadas al día	1 2 3 4

Fuente: elaboración propia.

Para el tratamiento estadístico, y en particular para las variables de tipo cuantitativo, incluidas las variables discretas, se recomienda que los datos se recaben sin agrupar; es decir, como una pregunta abierta para posteriormente crear intervalos, sobre todo cuando el comportamiento de estos puede ser muy disperso y se requiere en un ejercicio posterior agruparlos y hacer intervalos. Algunos ejemplos pueden ser el número de seguidores en redes sociales, el número de "me gusta" a un comentario, el número de personas en un grupo, el número de veces que se ha compartido o reenviado una foto, etc.

Continuas: son aquellas que admiten cualquier valor dentro de un rango numérico determinado. Son el resultado de medir, por lo que asumen valores numéricos fraccionados o decimales. Pueden tomar, entonces, cualquier valor de un determinado intervalo.

Por lo que se refiere a **variables continuas**, Celis de la Rosa y Labrada (2014) señalan que son "aquellas en las cuales la escala de medición se puede dividir en una cantidad infinita de valores entre dos puntos cualquiera" (p. 5).

De acuerdo con Estuardo (2012), las **variables cuantitativas continuas** son "respuestas numéricas que surgen de un proceso de medición, las cuales pueden tomar valores entre dos números enteros" (p. 6).

Salazar y Del Castillo (2018) establecen que las **variables cuantitativas continuas** son "aquellas que presentan números relativos o racionales (fraccionados, porcentuales y/o decimales) siendo esta medición aproximada (p. 21).

Tabla 5. ***Ejemplos de variables cuantitativas continuas***

Variable	Escala de valor o clase
Estatura	1.75 1.80 1.90 2.00
Peso	3.100 3.200 3.300 3.400
Hora de llegada a clase	8.00 8.15 8.20 8.30
Tiempo de realizado en un maratón	1.59 2.05 2.10 2.20
Distancia recorrida (km)	6.25 6.30 6.40 6.50

Fuente: elaboración propia.

Para el tratamiento estadístico, al igual que las variables discretas, se recomienda recuperar los datos de las variables cuantitativas continuas sin agrupar, para que posteriormente se puedan crear intervalos y hacer la revisión estadística correspondiente.

3.1.1.3. Identificación de variables

En el estudio y tratamiento de las variables, lo más importante es identificar, en un contexto determinado, cuáles son sus niveles de medición generalmente presentes en un instrumento de recolección o en una hipótesis, como se muestra a continuación.

Hipótesis: un profesor suponía que el tiempo destinado a estudiar aumentaba la calificación obtenida por los alumnos en la asignatura.

a) Desde su clasificación metodológica, ¿cuál es la variable independiente en la hipótesis?

Tiempo destinado al estudio

b) Desde su clasificación metodológica, ¿cuál es la variable dependiente en la hipótesis?

Calificación

c) Desde el punto de vista estadístico, ¿cómo se clasifica la variable independiente?

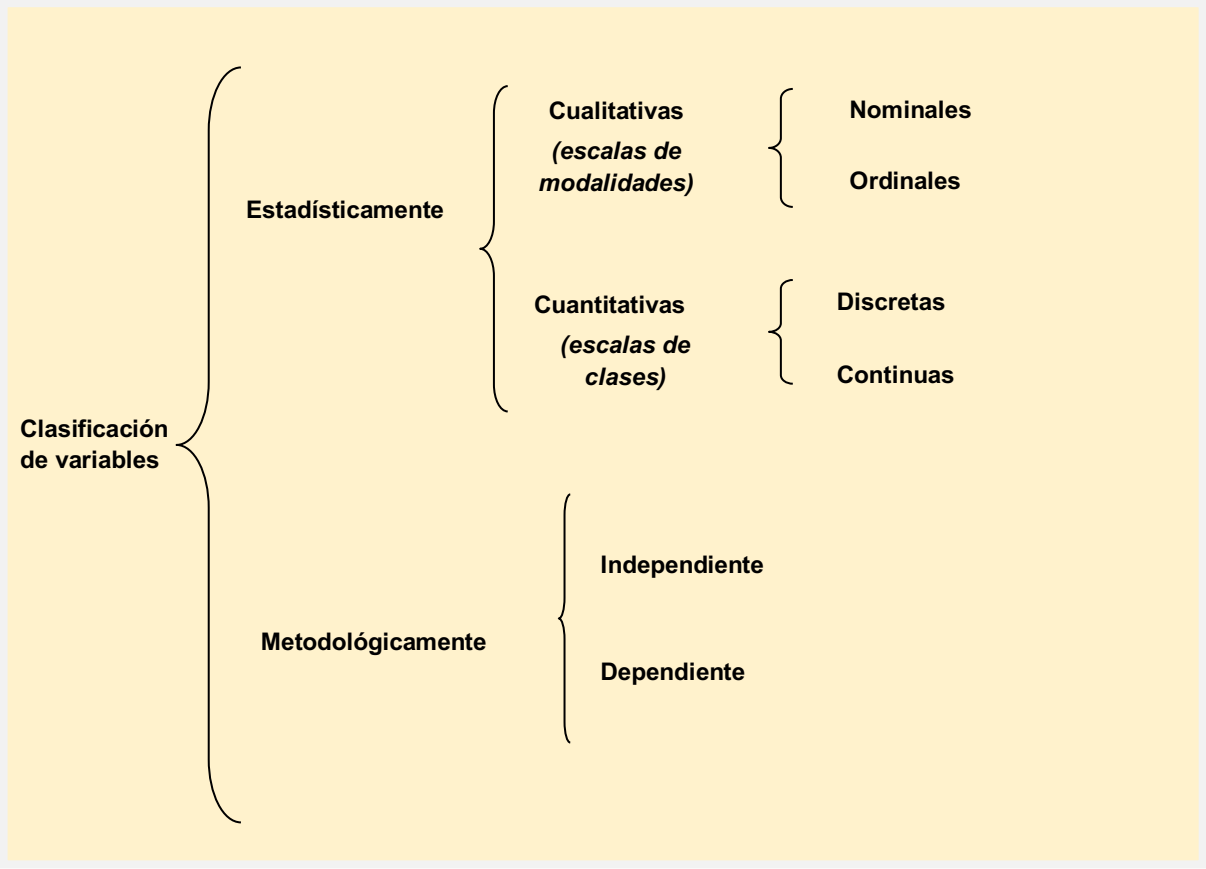
Para determinar su clasificación estadística es necesario definir la variable de estudio; en este caso, se entenderá por *Tiempo destinado a estudiar* al número de horas destinadas por una persona para revisar el contenido de una asignatura; entonces se medirá con escalas de clases de 2 a 3 horas, de 4 a 5 horas, de 6 a 7

horas. Por lo tanto, su clasificación desde el punto de vista estadístico es: **variable cuantitativa discreta con escala de clases**.

d) Desde el punto de vista estadístico, ¿cómo se clasifica la variable dependiente?

Para determinar su clasificación estadística es necesario definir la variable de estudio; en este caso, se entenderá por *Calificación* a la puntuación obtenida resultado de la valoración de los criterios de acreditación, expresada en categorías de MB, B, S, NA y NP; entonces, se medirá con escalas de modalidades de: Muy bien; Bien, Suficiente, No aprobatoria y No presentó. Por lo tanto, se trata de una **variable cualitativa ordinal con escala de modalidades.**

Figura 1. ***Clasificación de variables y niveles de medición***



Fuente: elaboración propia.

Una vez revisados los temas anteriores podemos concluir que las variables se clasifican estadística y metodológicamente (figura 1). Estadísticamente en cualitativas y cuantitativas. Las variables cualitativas a su vez se dividen en nominales y ordinales; y las variables cuantitativas, en discretas y continuas. Las variables nominales son aquellas que asumen un nombre o una característica, las variables ordinales asumen un nombre o una característica, pero con un orden, nivel de importancia o jerarquía. Las variables discretas asumen valores numéricos enteros y las variables continuas valores numéricos fraccionados o con decimales. Metodológicamente las variables se clasifican en independiente y dependiente. La variable independiente es aquella que establece o define el investigador, causa, origen o razón de estudio, y la variable dependiente es el resultado, efecto o consecuencia que mide el investigador.

Capítulo 4
El método estadístico: contar

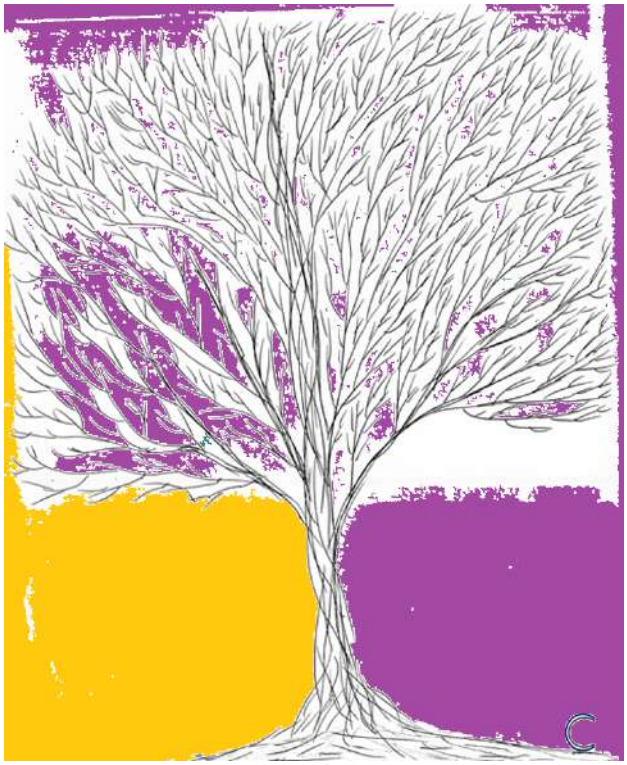


Imagen de CLM 2019

4.1. Contar

En esta etapa, **contar** significa determinar los valores que se obtienen en cada una de las modalidades o clases de la variable a medir; es decir, define la frecuencia –o número de veces que se repite un valor– en cada una de esas modalidades o clases de la variable de estudio.

En un tratamiento electrónico –mediante el empleo de un procesador estadístico u hoja del cálculo– se entiende por contar al procedimiento donde los datos son sometidos a revisión, clasificación y cómputo numérico.

Contar es el segundo paso del método estadístico, consiste en asociar las modalidades o clases con el número de veces en que aparece un dato dentro de un conjunto de observaciones, a estos valores se les denomina frecuencias.

4.2. Frecuencia absoluta

La **frecuencia absoluta** se define como el número de veces que se repite un dato. Esta es una medida estadística que establece la cantidad de veces que se repite un evento al realizar un número determinado de experimentos aleatorios.

En concordancia con Gorgas, Cardiel y Zamorano (2011), la **frecuencia absoluta** se define como "el número de veces que aparece repetido el valor en cuestión de la variable estadística en el conjunto de las observaciones realizadas" (p. 13).

Para Córdova y Cortés (2010) la **frecuencia absoluta de un dato** es "el número de veces que se repite ese dato, también se presenta la frecuencia absoluta de un intervalo que se refiere al número de datos que pertenecen a ese intervalo. Se denota por **f**" (p. 35).

La **frecuencia absoluta** es útil para conocer las características de una muestra o una población y es la base para determinar la frecuencia relativa. El total de las frecuencias absolutas es igual a la suma total de datos recabados de la muestra o de la población.

4.3. Frecuencia relativa

La **frecuencia relativa** es el porciento de veces en que ocurre un dato. Es la transformación en porcentaje de la frecuencia absoluta. La **frecuencia relativa** es una medida estadística que se obtiene al dividir la frecuencia absoluta de cada dato entre el número total de datos, y cuyo resultado se multiplica por cien.

Para Gorgas, Cardiel y Zamorano (2011) la **frecuencia relativa** se define como el "cociente entre la frecuencia absoluta y el número de observaciones realizadas *N*, es un dato el cual se puede expresar en tantos por cientos del tamaño de la muestra, para lo cual basta con multiplicar por 100" (pp. 14-15).

Por su parte, Córdova y Cortés (2010) señalan que la **frecuencia relativa**

> de un dato se obtiene al dividir la frecuencia absoluta de cada dato entre el número total de datos. De un intervalo se obtiene al dividir la frecuencia absoluta de cada intervalo entre el número total de datos. Se denota por f_r . (p. 36)

La frecuencia relativa determina la proporción o porcentaje que tiene algún valor u observación en la población o en la muestra. Esta frecuencia permite hacer comparaciones entre muestras de tamaños distintos. Se expresa como un valor decimal, proporción o como un porcentaje.

4.4. Frecuencia absoluta acumulada

La **frecuencia absoluta acumulada** es la frecuencia absoluta de cada uno de los datos más la suma de los valores anteriores a dicha suma. Esta frecuencia es el resultado de sumar o de reunir las frecuencias absolutas anteriores que presenta una población o muestra.

Una vez determinada la frecuencia absoluta; la frecuencia absoluta acumulada se obtiene al sumar las frecuencias absolutas de una clase o modalidad de la población o muestra con la anterior, así sucesivamente hasta llegar a acumular el total de la población o muestra.

Para Córdova y Cortés (2010) la **frecuencia absoluta acumulada** es

> la suma de las frecuencias absolutas de todos los datos anteriores, incluyendo también la del dato mismo del cual se desea su frecuencia acumulada. De un intervalo es la suma de las frecuencias absolutas de todos los intervalos de clase anteriores, incluyendo la frecuencia del intervalo mismo del cual se desea su frecuencia acumulada. La última frecuencia absoluta acumulada deberá ser igual al número total de datos. Se denota por f_a. (p. 35)

4.5. Frecuencia relativa acumulada

La **frecuencia relativa acumulada** es la frecuencia relativa de cada uno de los datos más la suma de los valores anteriores a dicha suma. Una vez determinada la frecuencia relativa, la frecuencia relativa acumulada se obtiene al sumar las frecuencias relativas de una clase o modalidad de la población o muestra con la anterior, así sucesivamente hasta llegar a acumular el total de la población o muestra.

Finalmente, Córdova y Cortés (2010) establecen que la **frecuencia relativa acumulada** es

> la suma de las frecuencias relativas de todos los datos anteriores, incluyendo también la del dato mismo del cual se desea su frecuencia relativa acumulada de un intervalo es la suma de las frecuencias relativas de todos los intervalos de clase anteriores incluyendo la frecuencia del intervalo mismo del cual se desea su frecuencia relativa acumulada. La última frecuencia relativa acumulada deberá ser igual a la unidad. Se denota por f_{ra}. (p. 36)

Capítulo 5
El método estadístico: presentar

Imagen de CLM 2019

5.1. Presentar

Presentar significa mostrar de manera numérica o visual la frecuencia de los datos a través de tablas o gráficos, dependiendo del nivel de medición de la o las variables a exponer.

Presentar también implica determinar el tipo de tabla y gráfico que deberá corresponder a cada variable según su nivel de medición; así como establecer las normas mínimas para mostrar de manera numérica o visual los datos.

Una tabla o cuadro presenta de manera ordenada, jerarquizada, resumida y organizada la información para facilitar la comprensión de los datos. Mientras que un gráfico es la representación visual de los datos; expresa los resultados clave de una variable de estudio a través de formas o figuras (líneas, puntos, barras, círculos, etc.)

5.1.1. Tablas

La Organización de las Naciones Unidas (ONU) y la Comisión Económica para Europa (CEE) (2009a), en el libro *Cómo hacer comprensibles los datos. Parte 1: una guía para escribir sobre números,* sostienen que las tablas "deben presentar los números de forma concisa y bien organizada para ayudar al análisis de datos, minimizar los números en el texto estadístico y eliminar la necesidad de discutir variables no significativas o no esenciales para la línea argumental" (p. 10).

Estos organismos internacionales señalan que las tablas deben facilitar a los lectores la localización y comprensión de los datos, para ello sugieren presentar tablas pequeñas; utilizar un decimal para la mayoría de los datos y en casos específicos, dos o más decimales para ilustrar sutiles diferencias en una serie de datos; así como clasificar los datos por orden o por jerarquías para que sean fácilmente asimilables. También proponen mostrar las cifras más altas y las más bajas y otros datos atípicos; justificar a la derecha las cifras para reforzar su magnitud; evitar celdas en blanco; y considerar las tablas grandes o complejas como material de apoyo o anexo.

Finalmente, para el título de una tabla se considera que debe ser corto y describir el tema o mensaje contenido en la misma.

Por otra parte, la ONU y la CEE retoman a Miller (2004) que recomienda en la construcción de tablas lo siguiente:

- Que sea fácil para la audiencia encontrar y entender los números en las tablas.
- Que tanto el formato como el título de la tabla sean discretos y sencillos, de forma que la atención se centre en los puntos sustanciales expresados por los datos más que en la propia estructura de la tabla.

En el libro *Cómo hacer comprensibles los datos. Parte 2: una guía para presentar estadísticas* la Organización de las Naciones Unidas (ONU) y la Comisión Económica para Europa (CEE) (2009b, p. 14) establecen cinco elementos de apoyo que son necesarios para la descripción de los datos mostrados en una tabla:

- **El título de la tabla** debe hacer una descripción clara y precisa de los datos. Debe responder las tres preguntas *qué*, *dónde* y *cuándo*. Debe ser breve y conciso y evitar el uso de verbos.

- **Los encabezados de las columnas**, expuestos en la parte superior de la tabla, deben indicar qué datos hay presentes en cada columna de la tabla y proporcionar los metadatos necesarios (por ejemplo, unidad de medida, periodo de tiempo o área geográfica).
- **Los encabezados de las filas**, en la primera columna de la tabla, deben identificar los datos presentes en cada fila de la tabla.
- **Las notas a pie**, en la parte inferior de la tabla, pueden proporcionar cualquier información adicional necesaria para comprender y utilizar correctamente los datos (definiciones, por ejemplo).
- **La línea que contiene la fuente**, en la parte inferior de la tabla, debe indicar la fuente de los datos, es decir, la organización que elaboró los datos y el método de recogida de datos (por ejemplo, censo de población o encuesta de población activa).

La siguiente figura indica cómo estos componentes de una tabla deben ser mostrados, según la ONU-CEE.

Figura 2. ***Componentes de una tabla***

Título de la tabla

	Encabezados de columnas
Encabezados de filas	**Datos**

Notas a pie
Fuente:
Fuente: ONU-CEE (2009b).

Para asegurar que las tablas sean fáciles de entender, la ONU y la CEE (2009b, pp. 15-16) proponen considerar lo siguiente:

- Evita texto innecesario.
- Presenta los datos por orden cronológico en el caso de series temporales o usando una clasificación estándar. Para series temporales más largas, puede ser más apropiado utilizar el orden cronológico inverso (es decir, comenzando con el periodo más reciente y yendo hacia atrás), como el desempleo mensual.
- Usa mínimamente las posiciones decimales.
- Utiliza separadores de miles. El uso de un espacio en lugar de un símbolo puede evitar el problema de tener que traducir entre lenguas.
- Alinear los números hacia la coma decimal (o a la derecha en la ausencia de cifras decimales) hace claramente apreciable su valor relativo. No centres los números en una columna, a menos que todos ellos tengan la misma extensión.

- No dejar ninguna celda de datos vacía. Los valores que faltan deben ser identificados como *no disponible* o *no aplicable*. La abreviatura *NA* se puede aplicar a cualquier caso, por lo que hay que definirla.

5.1.2. Gráficos

La Organización de las Naciones Unidas (ONU) y la Comisión Económica para Europa (CEE) (2009a) en el libro *Cómo hacer comprensibles los datos. Parte 1: una guía para escribir sobre números,* consideran que los gráficos pueden ser muy efectivos para expresar resultados clave o para ilustrar una presentación. La efectividad se produce cuando el gráfico tiene un mensaje claro, visual, con un titular explicativo. Si un gráfico intenta decir demasiado se convierte en un rompecabezas que requiere demasiado esfuerzo para descifrarlo. En el peor de los casos, se vuelve engañoso. Sugieren hacer todo lo posible para que la audiencia pueda comprender fácilmente el propósito del investigador (p. 8).

Para hacer un buen gráfico estadístico, la ONU y la CEE (2009ª, p. 8) sugieren:

- Hacer la imagen grande para representar muchos datos;
- Los datos deben transmitir una conclusión o una idea;
- Destacar los datos y evitar información innecesaria o que distraiga (celdas en blanco, ruido o basura en el gráfico);
- Presentar diseños visualmente lógicos.

Para lograr claridad en los gráficos, la Organización de las Naciones Unidas (ONU) y la Comisión Económica para Europa (CEE) (2009a, p. 19) señalan que se debe:

- Preferentemente, utilizar línea continua en lugar de otros estilos; y en los rellenos o fondos de gráfico un solo color y sin trama.
- Evitar marcadores de datos en los gráficos de línea.
- Utiliza los valores de los datos solo si estos no interfieren en la visualización del gráfico principal.
- Comenzar en cero la escala del eje de ordenadas Y.
- Utilizar una sola unidad de medida por gráfico.
- Utilizar diseños bidimensionales para datos bidimensionales.
- Hacer fácilmente comprensible el texto presente en el gráfico:
  - No utilizar abreviaturas.
  - Evitar los acrónimos.
  - Escribir etiquetas de izquierda a derecha.
  - Emplear una gramática correcta.
  - Evitar leyendas excepto en los mapas.

Además, la Organización de las Naciones Unidas (ONU) y la Comisión Económica para Europa (CEE) (2009b, p. 19) indican que cuando se realizan representaciones visuales, se debe pensar en:

- **El grupo-objetivo:** diferentes formas de presentación pueden ser adecuadas para diferentes públicos (por ejemplo, de negocios o del mundo académico, especialistas o población general).
- **El papel de la gráfica en la presentación:** analizar el cuadro completo o centrar la atención en puntos clave pueden requerir diferentes tipos de presentaciones visuales.
- **Cómo y dónde será presentado el mensaje:** determinar si será un análisis largo y detallado, o una presentación rápida.
- **Aspectos contextuales que pueden distorsionar la comprensión:** por ejemplo, usuarios de datos expertos o novatos.
- **Si es mejor solución el análisis textual o una tabla de datos.**
- **Consideraciones sobre accesibilidad:**
    - Ofrecer alternativas de texto para los elementos no textuales como gráficos e imágenes.
    - No depender únicamente del color: ¿quitando el color, la presentación sigue siendo comprensible?, ¿tienen las combinaciones de colores suficiente contraste?, ¿valen los colores para daltónicos (rojo/verde)?
    - Asegurar que el contenido de reproducción puede ser controlado por el usuario (por ejemplo, pausar gráficos animados).
- **Coherencia entre las representaciones visuales de datos:** asegurar que los elementos de las representaciones visuales estén diseñados de forma coherente y usa las convenciones habituales cuando sea posible (por ejemplo, azul para representar el agua en un mapa).
- **Tamaño, duración y complejidad:** ¿la presentación es fácil de entender?, ¿es demasiado para el público en una sesión dada?
- **Posibilidad de una mala interpretación:** prueba la presentación con compañeros, amigos o alguna persona del grupo objetivo, para ver si se capta el mensaje pretendido.

Por su parte, Reynaga et al. (1996, p. 75) señala que en la **presentación de una tabla o gráfico** se deben aplicar al menos las siguientes **normas mínimas**:

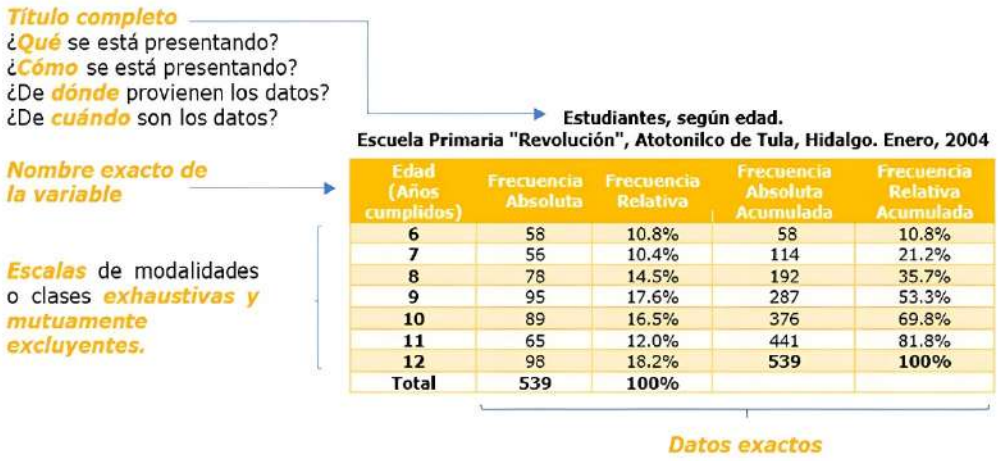
a) **Título completo** que responda a las preguntas:
    1) ¿**Qué** se está presentando? Responde a la población sujeto de estudio (personas, organizaciones, mujeres, estudiantes, etc.)
    2) ¿**Cómo** se está presentando? Responde a la variable o variables a mostrar.
    3) ¿De **dónde** provienen los datos? Responde al lugar donde se aplica el estudio.
    4) ¿De **cuándo** son los datos? Responde a la fecha en que se realiza el estudio.

b) **Nombre exacto de la variable**: debe representar las escalas de modalidades o clases definidas de la variable a medir.

c) **Escalas** de modalidades o clases **exhaustivas y mutuamente excluyentes**. Cada escala de modalidad o clase debe representar a una categoría o valor que no se incluya en otra.

d) **Datos exactos**. Cada frecuencia absoluta, relativa, absoluta acumulada o relativa acumulada debe ser exacta y correcta dado el carácter numérico que representa.

Figura 3. *Ejemplo de tabla*

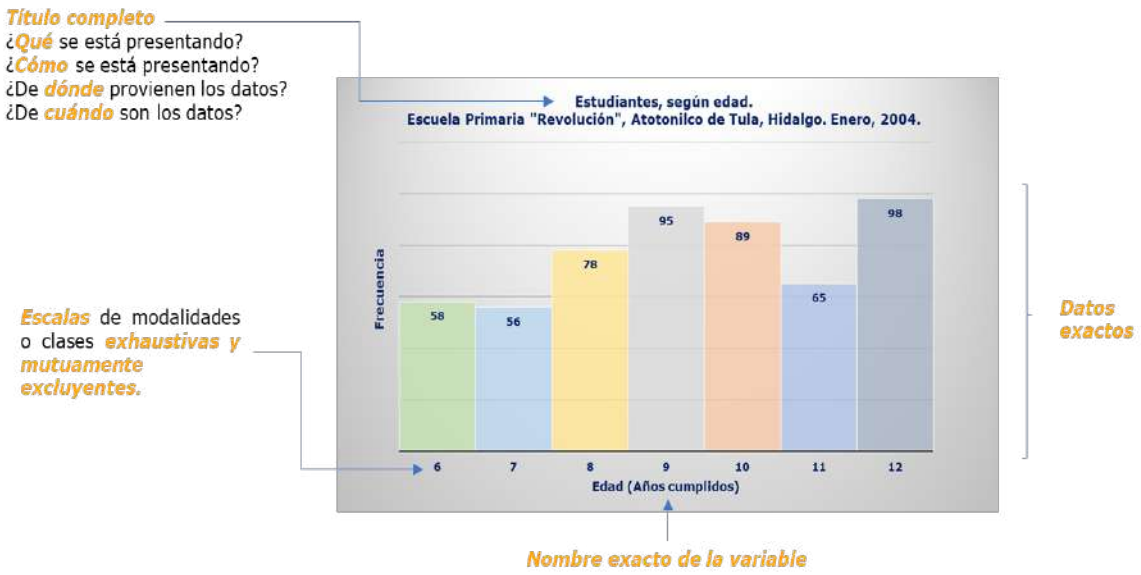


Estudiantes, según edad.
Escuela Primaria "Revolución", Atotonilco de Tula, Hidalgo. Enero, 2004

Edad (Años cumplidos)	Frecuencia Absoluta	Frecuencia Relativa	Frecuencia Absoluta Acumulada	Frecuencia Relativa Acumulada
6	58	10.8%	58	10.8%
7	56	10.4%	114	21.2%
8	78	14.5%	192	35.7%
9	95	17.6%	287	53.3%
10	89	16.5%	376	69.8%
11	65	12.0%	441	81.8%
12	98	18.2%	539	100%
Total	539	100%		

Fuente: elaboración propia.

Figura 4. *Ejemplo de gráfica*



Fuente: elaboración propia.

En la presentación de información en tablas o gráficos el investigador deberá decidir cuál de las dos opciones refleja de mejor manera el comportamiento de los datos. Se sugiere no generar una tabla y un gráfico por cada una de las variables de estudio. Una variable responde a un nivel de medición y este a su vez implica un tipo de tabla y gráfico a utilizar para la presentación de datos.

5.2. Tablas y gráficos para variables cualitativas

Una **variable cualitativa nominal u ordinal** se presenta en una **tabla de distribución de frecuencias** y en un **gráfico de barras o columnas separadas; circular, de sectores o pastel o un pictograma**.

5.2.1. Tabla de distribución de frecuencias y gráfico de barras o columnas separadas

Para Triola (2009) una **distribución de frecuencias (o tabla de frecuencias)** "es una lista de valores de los datos (ya sea de manera individual o por grupos de intervalos), junto con sus frecuencias (o conteos) correspondientes" (p. 43). Ritchey (2004) por su parte señala que una **distribución de frecuencias** es un "listado de todas las puntuaciones observadas de una variable y la frecuencia *(f)* de cada puntuación (o categoría)" (p. 51).

Por otra parte, Salazar y Del Castillo (2018) refieren que un **gráfico de barras o columnas separadas**

> es un tipo de gráfico que constan de dos ejes, de los cuales se escoge a uno de ellos para representar a la variable de estudio de acuerdo a la distribución de frecuencias generada y el otro para representar la frecuencia de cada categoría, si es el eje vertical en el que se marcan las frecuencias el diagrama será de **columnas**, por el contrario si es en el eje horizontal donde se representan las frecuencia, el diagrama será de **barras**. (p. 35)

Un **gráfico de barras separadas** es un esquema que se integra de una serie de barras verticales u horizontales, donde la extensión de la barra representa la frecuencia proporcional (o relativa) de una categoría de una variable con nivel de medición nominal u ordinal. Las barras también pueden representar frecuencias absolutas de una categoría.

Los gráficos de barra separadas son especialmente eficaces para ilustrar la competencia entre categorías o comparar magnitudes entre categorías. Aris y Martínez (2013) consideran que este tipo de gráficas permiten comparar directamente las alturas observadas por cada categoría en una variable dando una imagen rápida de las formas más frecuentes de presentación (párr. 6).

Ejemplo: en el municipio de Atotonilco de Tula en el estado de Hidalgo, México, se realizó una investigación sobre la calidad de vida de las personas mayores. Se entrevistó a un total de 628 personas mayores. Una de las variables incluidas en el estudio fue el estado civil con los siguientes resultados: 156 solteros, 238 casados, 176 divorciados y 58 viudos.

Para la presentación de la variable de estudio es importante considerar las siguientes preguntas: ¿cuál es la variable de estudio?, ¿cuál es su nivel de medición?, ¿qué tipo de tabla le corresponde?, ¿qué tipo de gráfico se debe utilizar?, ¿cuáles son las normas que se deben aplicar para la presentación de tablas y gráficos?, entre otras.

En el ejemplo: la variable de estudio es de tipo *cualitativo nominal con escala de modalidades*; por lo tanto, se presenta en una *tabla de distribución de frecuencias* y en un gráfico denominado de *barras separadas*, como se muestra a continuación:

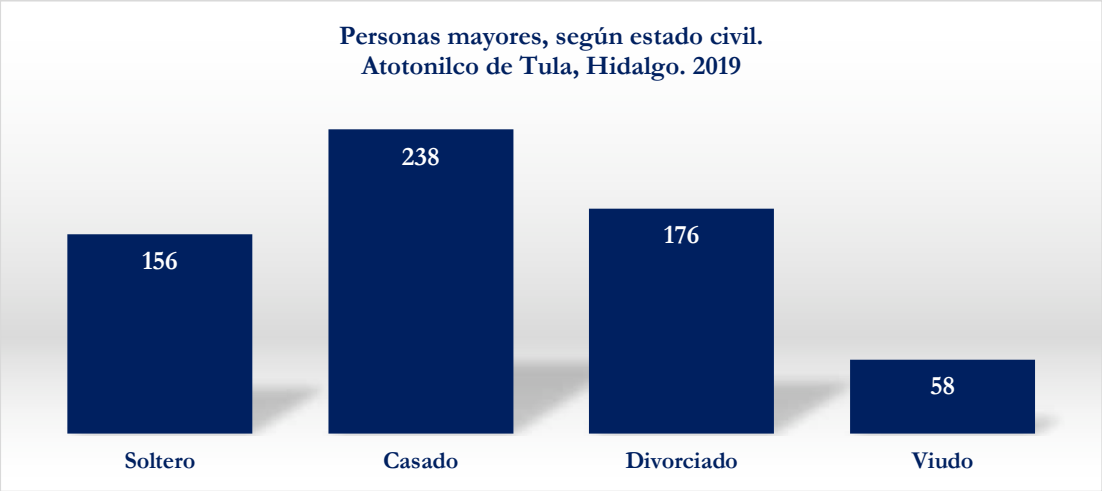
Tabla 6. *Distribución de frecuencias*

Personas mayores, según estado civil.
Atotonilco de Tula, Hidalgo. 2019

Estado civil	Frecuencia absoluta	Frecuencia relativa	Frecuencia absoluta acumulada	Frecuencia relativa acumulada
Soltero	156	25 %	156	25 %
Casado	238	38 %	394	63 %
Divorciado	176	28 %	570	91 %
Viudo	58	9 %	628	100 %
Total	628	100%	-	-

Fuente: Elaboración propia.

Figura 5. Gráfico de *barras separadas*



Fuente: elaboración propia.

5.2.2. Tabla de distribución de frecuencias y gráfico circular o de sectores o de pastel

De acuerdo con Ritchey (2004) un **gráfico de pastel** es un "círculo que se divide (o rebana) desde su punto central, donde cada espacio representa la frecuencia proporcional –o relativa– de determinada categoría de una variable nominal u ordinal" (p. 76)

Por su parte, Elorza (2008) considera que una **gráfica circular** es

> uno de los métodos gráficos más simples y usuales, y un instrumento auxiliar de análisis y presentación de la información, que resulta muy valioso para el investigador. Éste, como un diagrama en forma de círculo, es particularmente útil para visualizar las diferencias de frecuencia entre algunas categorías de nivel nominal. (p. 20)

Es especialmente útil para transmitir de manera sencilla y rápida un sentido de equidad, tamaño relativo o desigualdad al comparar categorías (señala Ritchey, 2004, p. 76). Se recomienda que las categorías no excedan de cinco para facilitar su comprensión y aumentar su aporte de información.

Estuardo (2012) señala que un gráfico circular "se usa para mostrar el comportamiento de las frecuencias relativas, absolutas o porcentuales de las variables. Dichas frecuencias son representadas por medio de sectores circulares, proporcionales a las frecuencias" (p. 17).

Ejemplo: la Universidad Nacional Autónoma de México de 2017 a 2018 tenía una población escolar de 348 617 estudiantes de los siguientes niveles de estudio: 30 310 de posgrado, 204 191 de licenciatura y 114 116 de bachillerato.

Para la presentación de la variable de estudio es importante considerar las siguientes preguntas: ¿cuál es la variable de estudio?, ¿cuál es su nivel de medición?, ¿qué tipo de tabla le corresponde?, ¿qué tipo de gráfico se debe utilizar?, ¿cuáles son las normas que se deben aplicar para la presentación de tablas y gráficos?, entre otras.

En el ejemplo: La variable de estudio es de tipo *cualitativo ordinal con escala de modalidades*; por lo tanto, se presenta en una *tabla de distribución de frecuencias* y en un gráfico denominado *circular, de sectores o de pastel*, como se muestra a continuación:

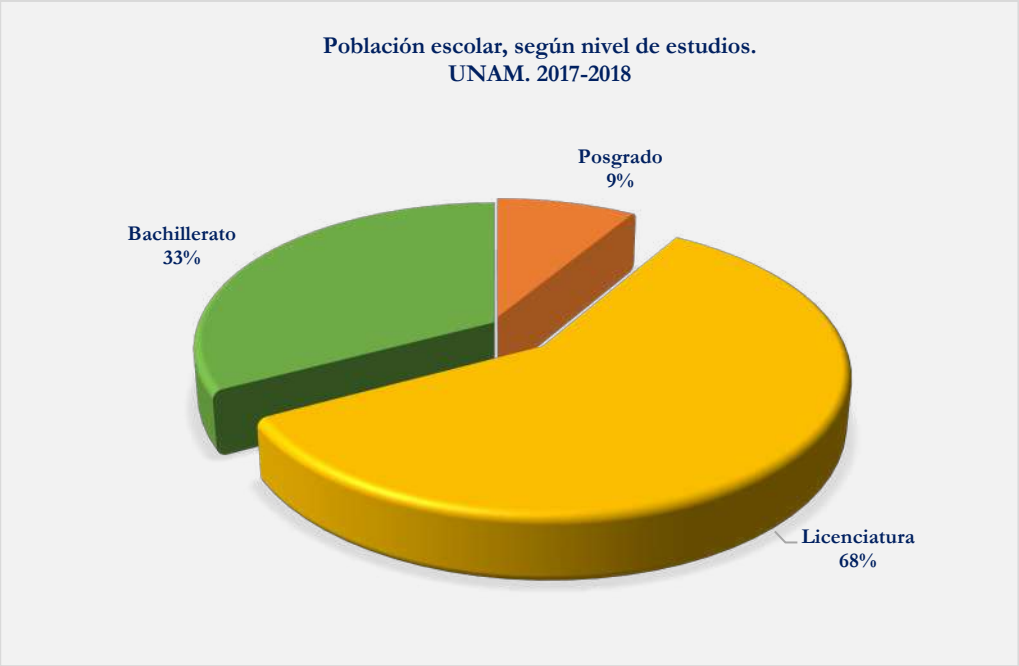
Tabla 7. ***Tabla de distribución de frecuencias***

Población escolar, según nivel de estudios.
UNAM. 2017-2018

Nivel	Frecuencia absoluta	Frecuencia relativa	Frecuencia absoluta acumulada	Frecuencia relativa acumulada
Posgrado	30 310	9 %	30 310	9 %
Licenciatura	204 191	58 %	234 501	67 %
Bachillerato	114 116	33 %	**348 617**	**100 %**
Total	**348 617**	**100 %**	-	-

Fuente: UNAM (2019).

Figura 6. Gráfica circular, de sectores o de pastel



Fuente: UNAM (2019).

5.2.3. Tabla de distribución de frecuencias y gráfico pictograma

Los **pictogramas o pictográficos**, como señala Ritchey (2004), son "figuras con formas de objetos que se colocan de manera que el número o el tamaño de los objetos representen la frecuencia proporcional de una categoría" (p. 80). El número o el tamaño de los objetos también puede representar la frecuencia absoluta.

En palabras de Estuardo A. (2012) un **pictograma** es

> un gráfico cuyo uso es similar al de sector circular, pero la frecuencia es representada por medio de una figura o dibujo que identifique a la variable en estudio. Este gráfico se utiliza para mostrar producciones en una serie cronológica. (p. 18)

Por su parte, para Kelmansky D. (2009) un **pictograma** "es un gráfico de barras que se reemplaza por figuras. Las figuras representan cantidades o porcentajes" (p. 49). Al igual que los gráficos de barras separadas, los pictogramas suelen usarse para comparar magnitudes entre categorías.

Ejemplo: el estado de Hidalgo, México en 2019 tenía una población de 3 050 720 personas, de los cuales 1 576 562 eran mujeres y 1 474 158 hombres.

Para la presentación de la variable de estudio es importante considerar las siguientes preguntas: ¿cuál es la variable de estudio?, ¿cuál es su nivel de medición?, ¿qué tipo de tabla le corresponde?, ¿qué tipo de gráfico se debe utilizar?, ¿cuáles son las normas que se deben aplicar para la presentación de tablas y gráficos?, entre otras.

En el ejemplo, la variable de estudio es de tipo *cualitativo nominal con escala de modalidades*; por lo tanto, se presenta en una *tabla de distribución de frecuencias* y en un gráfico denominado *pictograma*, como se muestra a continuación:

Tabla 8. Tabla de distribución de frecuencias

Población, según sexo.
Estado de Hidalgo, México, 2019.

Nivel	Frecuencia absoluta	Frecuencia relativa	Frecuencia absoluta acumulada	Frecuencia relativa acumulada
Mujer	1 576 562	52 %	1 576 562	52 %
Hombre	1 474 158	48 %	**3 050 720**	**100.00 %**
Total	**3 050 720**	**100 %**	-	-

Fuente: elaboración propia.

Figura 7. Gráfica pictograma

Fuente: elaboración propia.

5.2.4. Tabla de doble entrada, de asociación o de contingencias y gráfico de barras segmentadas o apiladas

Para la presentación de manera simultánea de **dos variables** de tipo cualitativo **nominal u ordinal** se utiliza una **tabla de doble entrada, de asociación o de contingencias** y un **gráfico barras segmentadas o apiladas**.

Una **tabla de doble entrada, de asociación o de contingencias** es una matriz o cuadro que permite organizar y sistematizar la información a partir de columnas horizontales y verticales que concentran y relacionan una serie de datos.

Los **gráficos de barras segmentadas o apiladas 100 %** establecen el porcentaje del total de cada grupo. Esto permite comparar diferencias relativas entre grupos. Cada barra muestra el porcentaje del total de cada grupo dividido por la representación en porcentaje de cada categoría frente a la cantidad total en cada grupo, lo que facilita identificar diferencias relativas entre los valores de cada grupo.

Ejemplo: en un estudio sobre cultura organizacional realizado en el Hospital AMM de la CDMX en 2019, se encontró que de un total de 2276 profesionales de la salud 841 eran solteros, de los cuales 524 pertenecían al área quirúrgica y 317 al área médica; en tanto que 1435 eran casados, 823 del área quirúrgica y 612 del área médica.

Para la presentación de la variable de estudio es importante considerar las siguientes preguntas: ¿cuáles son las variables de estudio?, ¿cuáles son sus niveles de medición?, ¿qué tipo de tabla les corresponde?, ¿qué tipo de gráfico se debe utilizar?, ¿cuáles son las normas que se deben aplicar para la presentación de tablas y gráficos?, entre otras.

En el ejemplo: las variables de estudio son de tipo *cualitativo nominal con escala de modalidades*; por lo tanto, se presenta en una *tabla de doble entrada o de contingencias o de asociación* y en un gráfico denominado *de barras segmentadas*, como se muestra a continuación:

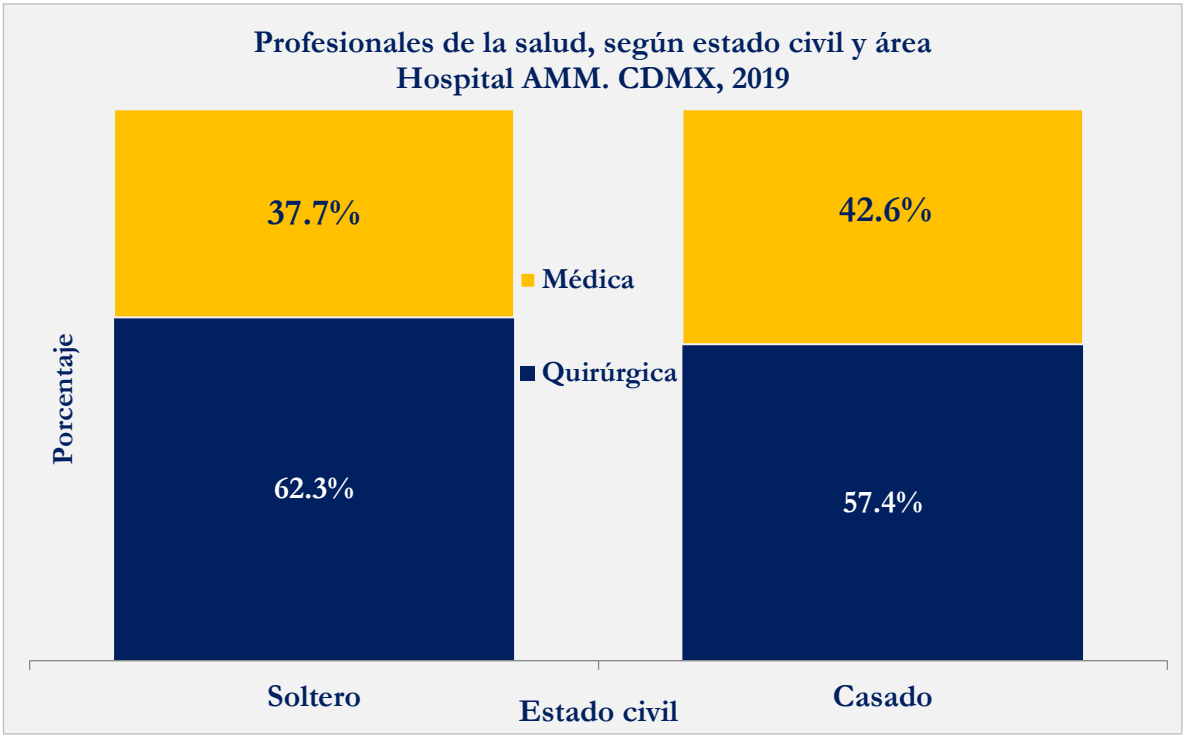
Tabla 9. ***Tabla de doble entrada, de asociación o de contingencias***

Profesionales de la salud, según estado civil y área.
Hospital AMM. CDMX, 2019.

Estado civil	Área		Total
	Quirúrgica	Médica	
Soltero	524 (62.3 %)	317 (37.7 %)	841 (100 %)
Casado	823 (57.4 %)	612 (42.6 %)	1435 (100 %)
Total	1347 (59.2 %)	929 (40.8 %)	2276 (100 %)

Fuente: elaboración propia.

Figura 8. *Gráfico barras segmentadas o apiladas*



Fuente: elaboración propia.

5.3. Tablas y gráficos para variables cuantitativas

Una **variable cuantitativa discreta** se presenta en una **tabla de distribución de frecuencias** y en un **gráfico** denominado **histograma.** Una **variable cuantitativa continua** también se presenta en una **tabla de distribución de frecuencias** y en un **gráfico** denominado **polígono de frecuencias**.

Los gráficos de histograma y polígono de frecuencias se utilizan de manera indistinta para presentar variables de tipo cuantitativo sean discretas o continuas. El gráfico de polígono de frecuencias acumuladas u ojiva y el de caja o bigotes también se pueden utilizar para este tipo de variables.

5.3.1. Tabla de distribución de frecuencias y gráfico histograma

De acuerdo con Ritchey (2004) un **histograma** es un

> diagrama de 90 grados que presenta las puntuaciones de una variable cuantitativa discreta (aunque también se pueden utilizar para niveles de medición de tipo continuo) a lo largo del eje horizontal, y la frecuencia de cada puntuación, en una columna al eje vertical. (p. 82)

El histograma es útil cuando se trata de representar distribución de frecuencias cuya variable es discreta –o continua– y viene dada por intervalos o clases.

Para Córdova y Cortés (2010) un **histograma** es

> una gráfica en forma de barras que consta de dos ejes, uno horizontal, llamado eje de la variable en observación, en donde situamos la base de una serie de rectángulos o barras contiguas; es decir, que no van separadas, y que se rotula con los límites inferiores de cada clase o intervalo excepto el último que deberá llevar también el límite superior, centradas en la marca de clase. Y un eje vertical llamado eje de las frecuencias, en donde se miden las alturas que vienen dadas por la frecuencia del intervalo que representa. Todos los intervalos deben tener la misma longitud. (p. 42)

Por su parte, Gorgas et al. (2001) señalan que un **histograma**

> es un conjunto de rectángulos adyacentes, cada uno de los cuales representa un intervalo de clase. La base de cada rectángulo es proporcional a la amplitud del intervalo. Es decir, el centro de la base de cada rectángulo ha de corresponder a una marca de clase. La altura se suele determinar para que el área de cada rectángulo sea igual a la frecuencia de la marca de clase correspondiente. Por tanto, la altura de cada rectángulo se puede calcular como el cociente entre la frecuencia (absoluta o relativa) y la amplitud del intervalo. En el caso de que la amplitud de los intervalos sea constante, la representación es equivalente a usar como altura la frecuencia de cada marca de clase, siendo este método más sencillo para dibujar rápidamente un histograma. (p. 18)

Ejemplo: en la escuela secundaria técnica 13 de Atitalaquia, Hidalgo, en enero de 2002 se midió la variable de peso a 80 estudiantes. Con objeto de resumir la información, los datos se agruparon en intervalos de clase.

Para la presentación de la variable de estudio es importante considerar las siguientes preguntas: ¿cuál es la variable de estudio?, ¿cuál es su nivel de medición?, ¿qué tipo de tabla le corresponde?, ¿qué tipo de gráfico se debe utilizar?, ¿cuáles son las normas que se deben aplicar para la presentación de tablas y gráficos?, entre otras.

En el ejemplo: la variable de estudio es de tipo *cuantitativo discreta con escala de clases*; por lo tanto, se presenta en una *tabla de distribución de frecuencias* y en un gráfico denominado *histograma*, como se muestra a continuación:

Tabla 10. ***Tabla de distribución de frecuencias***

Estudiantes, según peso
Escuela secundaria técnica 13, Atitalaquia, Hidalgo. Enero, 2002.

Peso (kilogramos)	Frecuencia absoluta	Frecuencia relativa	Frecuencia absoluta acumulada	Frecuencia relativa acumulada
52-57	1	1 %	1	1 %
58-63	1	1 %	2	2 %
64-69	7	9 %	9	11 %
70-75	13	16 %	22	27 %
76-81	24	30 %	46	57 %
82-87	18	23 %	64	80 %
88-93	8	10 %	72	90 %
94-99	5	6 %	77	96 %
100-105	3	4 %	80	100 %
Total	80	100 %		

Fuente: elaboración propia.

Figura 9. ***Gráfica histograma***

Fuente: elaboración propia.

5.3.2. Tabla de distribución de frecuencias y gráfico de polígono de frecuencias

Un **gráfico de polígono de frecuencias** –señala Ritchey (2004)– es un "diagrama de 90 grados con valores de nivel de medición continuos –incluso discretos– señalados en el eje horizontal, las frecuencias de las puntuaciones se presentan al centro de cada clase y se conectan mediante líneas rectas" (p. 83).

Córdova y Cortés (2010) señalan que un **polígono de frecuencias** es

> una gráfica del tipo de las gráficas de líneas trazadas sobre las marcas de clase, (de ahí el nombre de polígono), y se traza uniendo con segmentos de recta, de izquierda a derecha, las parejas ordenadas que se forman, al considerar como abscisa la marca de clase (eje horizontal) y como ordenada la frecuencia del intervalo representado (eje vertical); la primera y última parejas ordenadas se unen mediante un segmento de recta al eje horizontal, con las que serían la marca de clase anterior y posterior respectivamente si estas existieran. (p. 43)

De acuerdo con Ritchey (2004) los polígonos de frecuencias son especialmente útiles para comparar dos o más muestras. Mientras los histogramas presentan volumen y destacan el eje horizontal, los polígonos de frecuencias muestran picos y ponen la atención en el eje vertical. Los polígonos de frecuencias comunican un sentido de tendencia o movimiento.

Ejemplo: en la escuela primaria Revolución de Atotonilco de Tula, Hidalgo, en enero de 2018, se midió la variable gasto diario de traslado de 100 estudiantes. Con objeto de resumir la información se agruparon los datos en intervalos de clase.

Para la presentación de la variable de estudio es importante considerar las siguientes preguntas: ¿cuál es la variable de estudio?, ¿cuál es su nivel de medición?, ¿qué tipo de tabla le corresponde?, ¿qué tipo de gráfico se debe utilizar?, ¿cuáles son las normas que se deben aplicar para la presentación de tablas y gráficos?, entre otras.

En el ejemplo: la variable de estudio es de tipo *cuantitativo continua con escala de clases*; por lo tanto, se presenta en una *tabla de distribución de frecuencias* y en un gráfico denominado *polígono de frecuencias*, como se muestra a continuación:

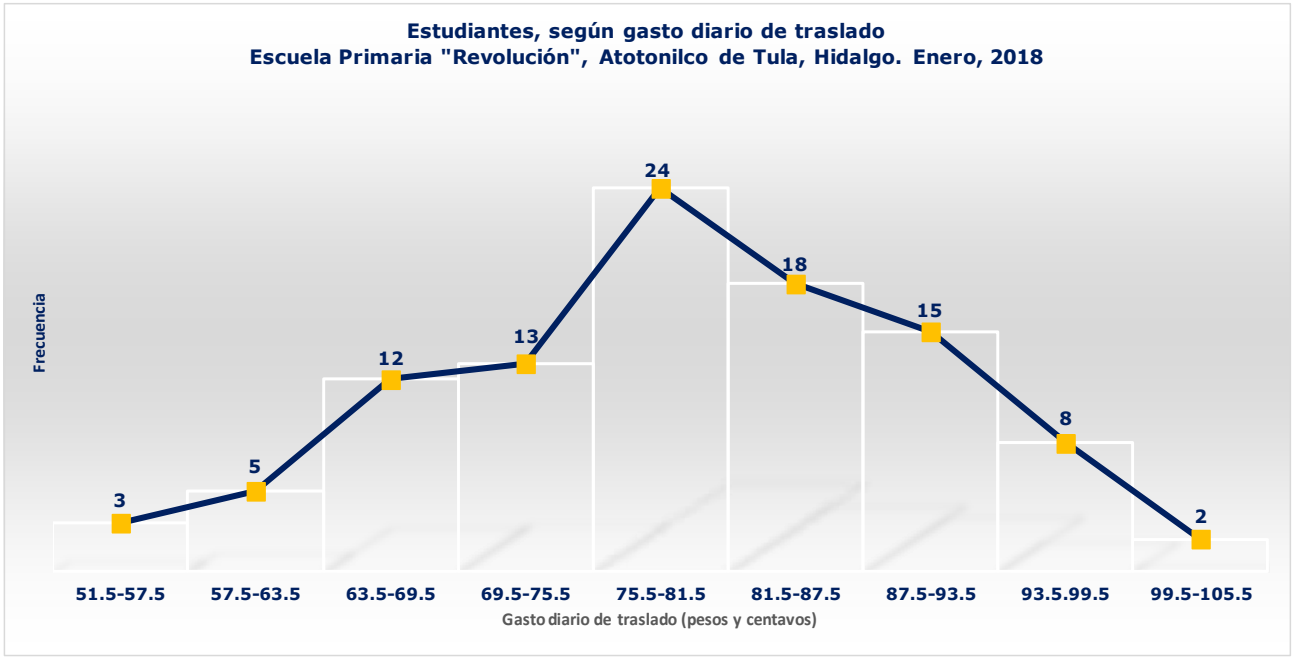
Tabla 11. *Tabla de distribución de frecuencias*

Estudiantes, según gasto diario de traslado
Escuela primaria *Revolución*, Atotonilco de Tula, Hidalgo. Enero, 2018

Gasto diario de traslado	Frecuencia absoluta	Frecuencia relativa	Frecuencia absoluta acumulada	Frecuencia relativa acumulada	Límite inferior real	Límite superior real	Centro de clase
51.5-57.5	3	3 %	3	3 %	51.5	57.5	54.5
57.5-63.5	5	5 %	8	8 %	57.5	63.5	60.5
63.5-69.5	12	12 %	20	20 %	63.5	69.5	66.5
69.5-75.5	13	13 %	33	33 %	69.5	75.5	72.5
75.5-81.5	24	24 %	57	57 %	75.5	81.5	78.5
81.5-87.5	18	18 %	75	75 %	81.5	87.5	84.5
87.5-93.5	15	15 %	90	90 %	87.5	93.5	90.5
93.5.99.5	8	8 %	98	98 %	93.5	99.5	96.5
99.5-105.5	2	2 %	100	100 %	99.5	105.5	102.5
Total	100	100 %	-	-	-	-	-

Fuente: elaboración propia.

Figura 10. *Gráfica de polígono de frecuencias*



Fuente: elaboración propia.

5.3.3. Tabla de distribución de frecuencias y gráfico de polígono de frecuencias acumuladas u ojiva o diagrama de Galton

Un gráfico de polígono de frecuencias acumuladas u ojiva o diagrama de Galton es la representación gráfica de las frecuencias relativas acumuladas o frecuencias porcentuales acumuladas de una variable. Una **gráfica ojiva** es un polígono frecuencias acumuladas que permite visualizar cuántas observaciones se encuentran por encima o debajo de un valor determinado.

Como afirma Elorza (2008), "a la representación gráfica de frecuencias acumuladas (sumadas progresivamente) se denomina **polígono de frecuencias acumuladas** y también recibe el nombre de **ojiva o diagrama de Galton"** (p. 25).

Para el trazado de esta gráfica, en primer lugar, se ubica el fenómeno de interés; es decir, la variable de estudio sobre el eje horizontal, en tanto que las frecuencias relativas acumuladas o frecuencias absolutas acumuladas sobre el eje vertical. Elorza (2008) señala que este gráfico se obtiene

> uniendo mediante una línea continua, los puntos cuyas ordenadas representan las frecuencias acumuladas (relativas o absolutas) de los intervalos y su abscisa, el límite superior real (LSR) de cada uno de ellos. La frecuencia acumulada de cada intervalo representa el número total de casos o porcentaje, dentro y debajo de un intervalo de clase en particular. (p. 25)

Ejemplo: en la escuela primaria Revolución de Atotonilco de Tula, Hidalgo, en enero de 2018, se midió la variable de gasto diario de traslado 100 estudiantes. Con objeto de resumir la información, se agruparon los datos en intervalos de clase.

Para la presentación de la variable de estudio es importante considerar las siguientes preguntas: ¿cuál es la variable de estudio?, ¿cuál es su nivel de medición?, ¿qué tipo de tabla le corresponde?, ¿qué tipo de gráfico se debe utilizar?, ¿cuáles son las normas que se deben aplicar para la presentación de tablas y gráficos?, entre otras.

En el ejemplo: la variable de estudio es de tipo *cuantitativo continua con escala de clases*; por lo tanto, se presenta en una *tabla de distribución de frecuencias* y en un gráfico denominado *polígono de frecuencias acumuladas, u ojiva o diagrama de Galton*, como se muestra a continuación:

Tabla 12. *Tabla de distribución de frecuencias*

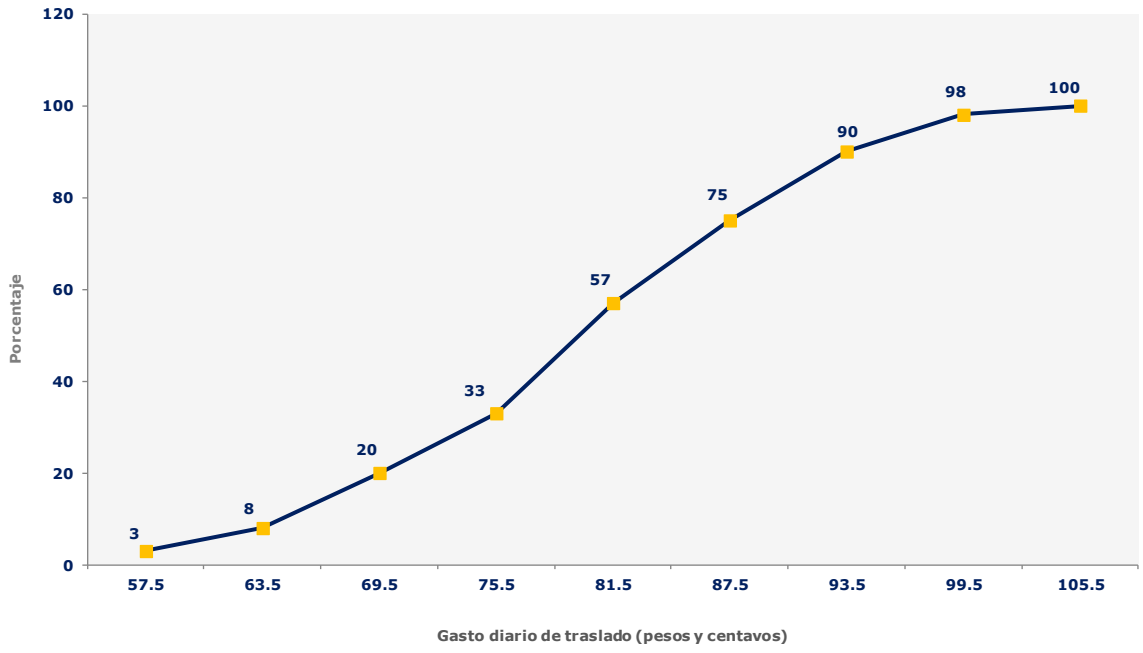
Estudiantes, según gasto diario de traslado.
Escuela primaria Revolución, Atotonilco de Tula, Hidalgo. Enero, 2018.

Gasto diario de traslado	Frecuencia absoluta	Frecuencia relativa	Frecuencia absoluta acumulada	Frecuencia relativa acumulada	Límite inferior real	Límite superior real	Centro de clase
51.5-57.5	3	3 %	3	3 %	51.5	57.5	54.5
57.5-63.5	5	5 %	8	8 %	57.5	63.5	60.5
63.5-69.5	12	12 %	20	20 %	63.5	69.5	66.5
69.5-75.5	13	13 %	33	33 %	69.5	75.5	72.5
75.5-81.5	24	24 %	57	57 %	75.5	81.5	78.5
81.5-87.5	18	18 %	75	75 %	81.5	87.5	84.5
87.5-93.5	15	15 %	90	90 %	87.5	93.5	90.5
93.5.99.5	8	8 %	98	98 %	93.5	99.5	96.5
99.5-105.5	2	2 %	100	100 %	99.5	105.5	102.5
Total	100	100 %	-	-	-	-	-

Fuente: elaboración propia.

Figura 11. *Polígono de frecuencias acumuladas u ojiva o diagrama de Galton*

Estudiantes, según gasto diario de traslado
Escuela primaria Revolución, Atotonilco de Tula, Hidalgo. Enero, 2018



Fuente: elaboración propia.

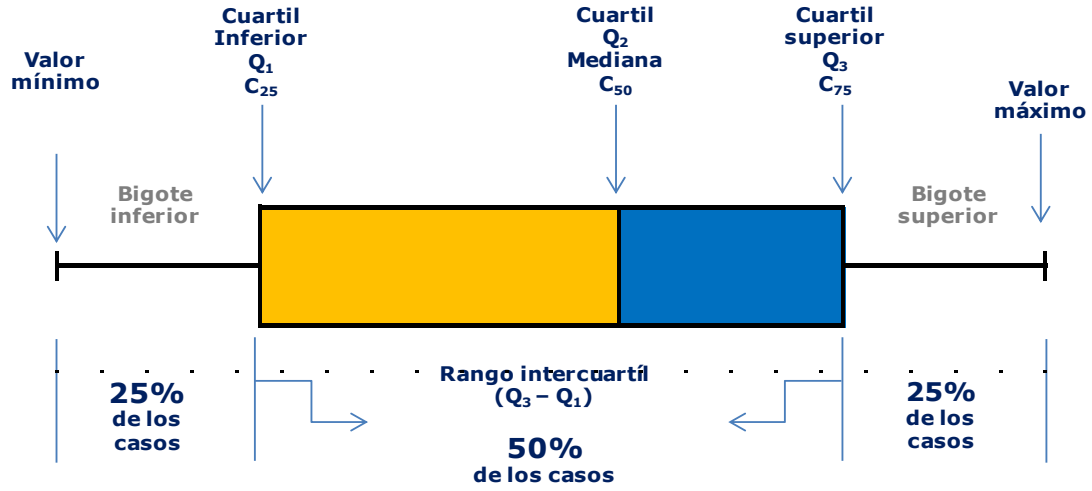
5.3.4. Tabla de distribución de frecuencias y gráfico de caja y bigotes

Un **gráfico de caja y bigotes** es la representación gráfica que despliega cuartiles (Q), así como puntuaciones mínimas y máximas de una distribución. Es una técnica útil para detectar valores extremos de una distribución.

En este sentido, Elorza (2008) señala que es "una forma de representar gráficamente los cuantiles, así como los valores extremos (el mínimo y el máximo) de un conjunto de datos, es una caja rectangular ubicada en un eje, vertical u horizontal" (p. 53).

Por su parte, De la Rosa y Labrada (2014) establecen que este tipo de gráfico de cuadro y línea es una manera útil de resumir datos presentados en percentiles. En tanto que Ritchey (2004) considera que es "una técnica gráfica muy útil para detectar valores extremos" (p. 89).

Figura 12. ***Gráfico de caja y bigotes***



Fuente: elaboración propia.

Ejemplo: en la ENTS-UNAM, en enero de 2019, se realizó una encuesta sobre formación extracurricular, una de las variables de estudio fue edad. El grupo estaba constituido por 20 estudiantes.

Para la presentación de la variable de estudio es importante considerar las siguientes preguntas: ¿cuál es la variable de estudio?, ¿cuál es su nivel de medición?, ¿qué tipo de tabla le corresponde?, ¿qué tipo de gráfico se debe utilizar?, ¿cuáles son las normas que se deben aplicar para la presentación de tablas y gráficos?, entre otras.

En el ejemplo: la variable de estudio es de tipo *cuantitativo discreto con escala de clases*; por lo tanto, se presenta en una *tabla de distribución de frecuencias* y en un gráfico denominado *de caja y bigotes*, como se muestra a continuación:

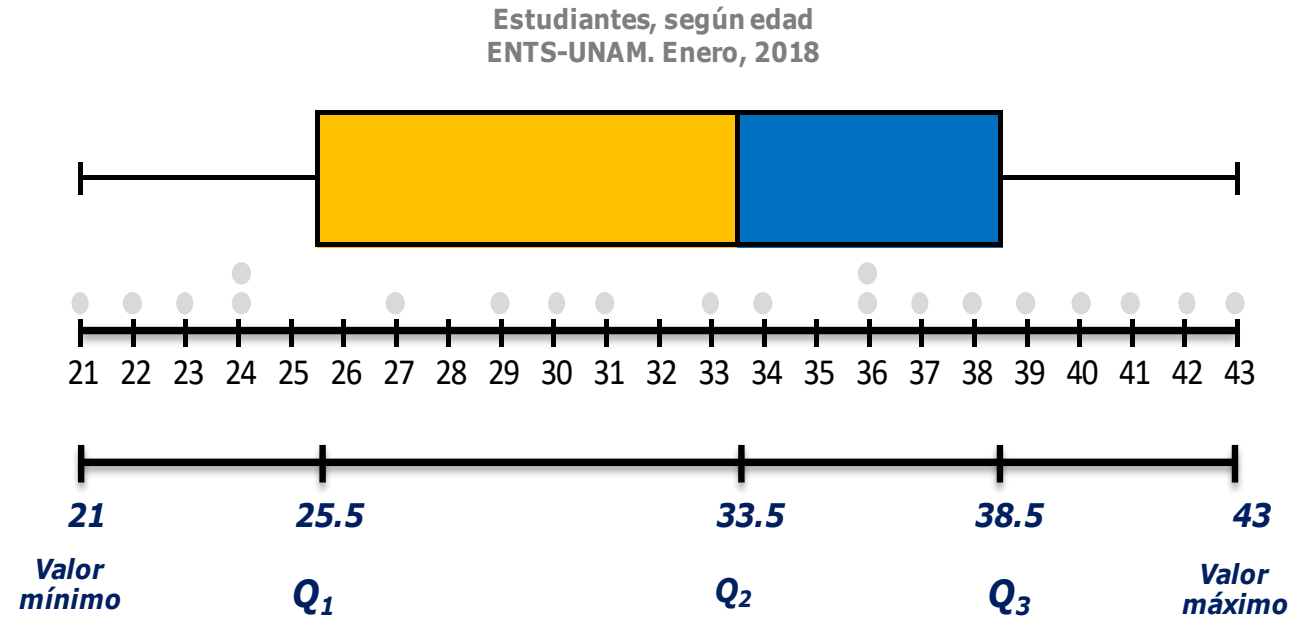
Tabla 13. ***Tabla de distribución de frecuencias***

Estudiantes, según edad.
ENTS-UNAM. Enero, 2018.

Valor	21	22	23	24	27	29	30	31	33	34	36	37	38	39	40	41	42	43	Total
Frecuencia	1	1	1	2	1	1	1	1	1	1	2	1	1	1	1	1	1	1	20

Fuente: ENTS-UNAM (2018).

Figura 13. ***Gráfico de caja y bigotes***



Fuente: ENTS-UNAM (2018).

5.3.5. Tabla de distribución de frecuencias y gráfico de gráfico de correlación

Para la presentación de manera simultánea de **dos variables** de tipo cuantitativo **discreto o continuo** se utiliza una **tabla de lista simple de datos** y un **gráfico de correlación**.

Una **tabla de lista simple de datos** es una lista de datos sobre la relación de dos conjuntos de valores o variables respecto a un sujeto, objeto de medición.

Celis de la Rosa y Labrada (2014) establecen que un **diagrama de puntos o correlación** "se utiliza para representar gráficamente la asociación que existe entre dos variables cuantitativas medidas en el mismo sujeto" (p. 55).

Un diagrama de puntos o correlación establece la relación existente entre dos variables cuantitativas. Se utiliza para conocer si existe una correlación en términos de magnitud –que tan intensa es– y dirección –positiva o negativa–.

Ejemplo: en el hospital AMM en enero de 2019 se realizó un estudio relacionado con el desempeño institucional de sus investigadores; en el mismo se incluyeron las variables: artículos producidos y antigüedad. El grupo estaba constituido por 7 investigadores.

Para la presentación de la variable de estudio es importante considerar las siguientes preguntas: ¿cuáles son las variables de estudio?, ¿cuáles son sus niveles de medición?, ¿qué tipo de tabla les corresponde?, ¿qué tipo de gráfico se debe utilizar?, ¿cuáles son las normas que se deben aplicar para la presentación de tablas y gráficos?, entre otras.

En el ejemplo: las variables de estudio son de tipo *cuantitativo discreto con escala de clases*; por lo tanto, se presenta en una *tabla de lista simple de pares datos* y en un gráfico denominado *de correlación*, como se muestra a continuación:

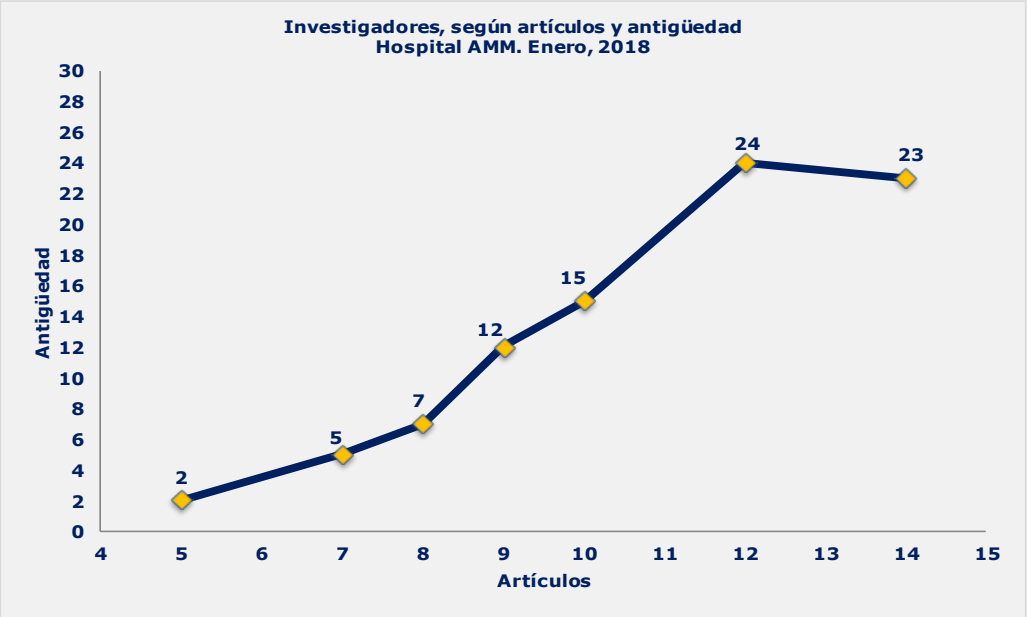
Tabla 14. ***Tabla de lista simple de datos***

Investigadores, según artículos y antigüedad
Hospital AMM. Enero, 2018

Investigador	Artículos	Antigüedad
1	5	2
2	7	5
3	8	7
4	9	12
5	10	15
6	12	24
7	14	23

Fuente: Hospital AMM (2018).

Figura 14. ***Gráfico de correlación***



Fuente: Hospital AMM (2018).

Capítulo 6
El método estadístico: describir (series simples)

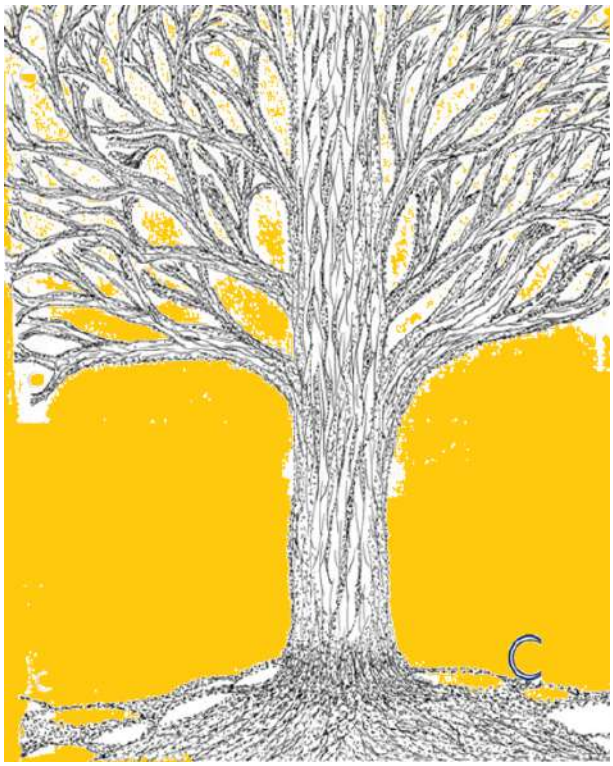


Imagen de CLM 2019

6.1. Describir: medidas de resumen en series simples

Describir implica el cálculo de medidas estadísticas que representan las características generales de un conjunto de datos basados en una muestra; implica obtener medidas de resumen para variables cualitativas como son proporciones, porcentajes, razones y tasas; así como medidas de resumen para variables cuantitativas, tal es el caso de medidas de tendencia central (moda, mediana, media), de posición (centil, cuartil, decil), de dispersión (rango, rango intercuartil, desviación estándar, varianza, coeficiente de variación) y de distribución (sesgo y curtosis).

En este apartado se estudiarán las medidas estadísticas de resumen que, como señala Kelmansky (2009), "permiten tener una idea rápida de cómo son los datos" (p. 120).

Por otra parte, no hay que perder de vista que lo más importante al estudiar datos es conocer *qué nos dicen* cada uno de ellos respecto del problema a estudiar, dado que cada medida de resumen calculada muestra una característica de los datos de estudio.

Para cumplir con el objetivo anterior, en la descripción de medidas de resumen aplicaremos el *Método DPI* (definición-procedimiento-interpretación); es decir, emplearemos un proceso de entendimiento que parte de fijar con claridad el significado de una palabra (definición), pasa por el cálculo de la prueba (procedimiento) y termina cuando se da razón de ser y referencia al valor obtenido (interpretación).

6.2. Para variables cualitativas

Las medidas estadísticas aplicables a variables de tipo cualitativo son **proporciones, porcentajes, razones y tasas**. La moda también es aplicable a este nivel de medición (se incluye en el apartado de variables cuantitativas).

6.2.1. Proporciones y porcentajes

Definición: es la importancia de un subconjunto con respecto al conjunto al que pertenece. Para Reynaga et al. (1996) una **proporción** es la "comparación, a través de una división, entre un subconjunto y el conjunto al que pertenece" (p. 89). Kelinger (1981) define a la **proporción** como "una fracción donde el numerador es una de dos o más frecuencias observadas y el denominador la suma de las frecuencias observadas" (p. 190).

Para Ritchey (2004) una **proporción** es "parte de la cantidad total o número de observaciones expresada en forma decimal" (p. 4). El resultado se multiplica por 100 y se convierte en **porcentaje**, es decir, en un símbolo matemático que representa una cantidad dada como una fracción en 100 partes iguales, dado que un **porcentaje** significa "por cien", y es igual a una proporción multiplicada por 100.

Procedimiento: determinar la proporción a obtener y de conformidad con la definición dividir subconjunto entre conjunto al que pertenece.

Ejemplo: en un estudio sobre deserción escolar a nivel licenciatura se encontró que en 1996 de 895 estudiantes que ingresaron a la carrera de trabajo social 179 abandonaron sus estudios durante el primer ciclo escolar.

$$\text{Proporción de alumnos que abandonaron sus estudios} = \frac{179}{895} = 0.20\ x\ 100 = 20.0\ \%$$

Interpretación: la importancia de los alumnos que abandonaron sus estudios durante el primer ciclo escolar, respecto al total de estudiantes que ingresaron a la carrera de trabajo social durante 1996, es de 0.20 o 20 %, lo que equivale a una quinta parte de la población de estudio. Uno de cada cinco alumnos abandona sus estudios durante el primer año escolar.

6.2.2 Razón

Definición: es la relación o existencia de un elemento con respecto a otro. Para Reynaga et al. (1996) una **razón** es la "comparación, a través de una división, entre dos conjuntos o grupos de elementos de diferente o igual naturaleza" (p. 88). Por su parte, para Hernández et al. (2014) una **razón** es simplemente "la relación entre dos categorías" (p. 300).

Procedimiento: Determinar la razón a obtener y de conformidad con la definición dividir entre dos elementos de iguales o diferentes.

Al respecto, Celis de la Rosa y Labrada (2014) señalan que

> cuando la serie que se está examinando consta sólo de dos categorías, o el interés de la investigación se dirige únicamente a dos categorías, se pueden utilizar las razones para resumir la información. Para ello, se divide la totalidad de individuos que tengan una característica (de preferencia el grupo de mayor tamaño) entre el grupo que tenga la otra característica. De esta manera, y a diferencia de las proporciones, las lecturas del numerador no se incluyen en el denominador. Su fórmula es: $R= a/b$. En la que R representa la razón, a simboliza el número de elementos con la característica de interés y b, el número de elementos con una característica diferente. Hay que notar que a/b no necesariamente son el total del universo. (p. 33)

Ejemplo: en la escuela primaria *Revolución* del municipio de Atotonilco de Tula, en el estado de Hidalgo se encontró que existía una población de 850 alumnos y 50 becas disponibles.

$$\text{Razón alumnos/becas} = \frac{850}{50} = 17$$

Interpretación: en la escuela primaria *Revolución* del municipio de Atotonilco de Tula, en el estado de Hidalgo existen 17 alumnos por cada beca disponible.

6.2.3 Tasa

Definición: es la magnitud o probabilidad de ocurrencia de un evento dentro de una población determinada. Para Reynaga et al. (1996) una **tasa** es la "comparación, a través de una división, entre el número de veces que ocurre un evento y la población en la que ocurre tal evento" (p. 90). El resultado se multiplica por un constante múltiplo de 10.

Por su parte, Hernández et al. (2014) definen una **tasa** como "la relación entre el número de casos, frecuencias o eventos de una categoría y el número total de observaciones, multiplicada por un múltiplo de

10, generalmente 100 o 1000" (p. 293). Mientras que para Ritchey (2004) una **tasa** es "la frecuencia de ocurrencia con relación a un número "base" especificado de sujetos en una población" (p. 7).

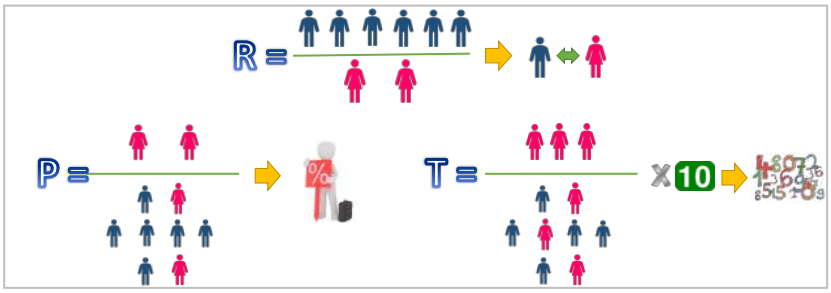
Procedimiento: determinar la tasa a obtener y de conformidad con la definición dividir el número de veces que ocurre el evento entre la población en la que ocurre tal evento.

Ejemplo: en la Escuela Nacional de Trabajo Social de la UNAM, en el ciclo escolar 2020-1, ingresaron a licenciatura 817 alumnos. De ellos 47 son originarios del norte de la República mexicana.

$$\text{Tasa de alumnos originarios del norte de la República mexicana} = \frac{47}{817} = 0.057 \boldsymbol{x\ 1000 = 57}$$

Interpretación: en la Escuela Nacional de Trabajo Social de la UNAM, en el ciclo escolar 2020-1, 57 de cada 1000 alumnos son originarios del norte de la República mexicana.

Figura 15. ***Proporción, razón, tasa***



Fuente: elaboración propia.

6.3. Para variables cuantitativas

Las medidas estadísticas aplicables a variables de tipo cuantitativo discretas o continuas se clasifican en: **de tendencia central** (moda, mediana, media), **de posición** (centil, cuartil, decil), ***de dispersión*** (rango, rango intercuartil, desviación estándar, varianza, coeficiente de variación) y ***de distribución*** (sesgo y curtosis).

6.3.1. De tendencia central

Las **medidas de tendencia central** indican en torno a qué valor se distribuyen los datos; o bien, como señala Rustom (2012), "cumplen el propósito de indicar el valor alrededor del cual se distribuyen los datos, es decir, una especie de centro de gravedad de estos" (p. 17).

Las medidas estadísticas aplicables a variables de tipo cuantitativo clasificadas como de tendencia central son **moda, mediana, media**.

6.3.1.1. Moda

Definición: es el valor de mayor frecuencia u ocurrencia en un conjunto de datos. Según Reynaga et al. (1996) es "el valor que en una serie se repite con mayor frecuencia" (p. 96).

Hernández et al. (2014) definen a la **moda** como "la categoría o puntuación que ocurre con mayor frecuencia" (p. 286). Por su parte, para Celis de la Rosa y Labrada (2014) la **moda** es "el valor que más se repite en un grupo de datos y profundizan al señalar que un grupo de datos puede tener más de una moda –pueden ser bimodal o multimodal–. Esta medida se puede utilizar tanto para variables cualitativas como para cuantitativas" (p. 45).

Ejemplo: un grupo de prácticas de especialización de la Escuela Nacional de Trabajo Social de la UNAM acudió a la Escuela Preparatoria Regional de Apaxco en el Estado de México, a impartir un curso de educación ambiental. Al finalizar el curso y con objeto de crear conciencia sobre el destino de la basura; el grupo de prácticas solicitó a cada participante pesar la basura orgánica que se generaba en su casa durante una semana.

Tabla 15. ***Estudiantes, según peso de basura orgánica generada.***

Apaxco, Estado de México. Enero, 2018.

Número de alumnos	Peso de basura orgánica generada (kg)	Frecuencia relativa	Frecuencia absoluta acumulada	Frecuencia relativa acumulada
1	12.5	4.2 %	12.5	4.2 %
2	12.7	4.2 %	25.2	8.4 %
3	12.9	4.3 %	38.1	12.7 %
4	13.2	4.4 %	51.3	17.1 %
5	13.7	4.6 %	65	21.7 %
6	14.3	4.8 %	79.3	26.5 %
7	14.3	4.8 %	93.6	31.3 %
8	14.3	4.8 %	107.9	36.1 %
9	14.3	4.8 %	122.2	40.9 %
10	14.9	5.0 %	137.1	45.9 %
11	15.2	5.0 %	152.3	50.9 %
12	15.5	5.2 %	167.8	56.1 %
13	15.7	5.3 %	183.5	61.4 %
14	15.9	5.3 %	199.4	66.7 %
15	15.9	5.3 %	215.3	72 %
16	16.1	5.4 %	231.4	77.4 %
17	16.4	5.5 %	247.8	82.9 %
18	16.6	5.6 %	264.4	88.5 %
19	16.9	5.7 %	281.3	94.2 %
20	17.5	5.8 %	**298.8**	**100 %**
Total	**298.8**	**100 %**	-	-

Fuente: elaboración propia.

Procedimiento: buscar el valor que más se repite o con mayor frecuencia que los demás.

En el ejemplo: el valor 14.3 kg. se repite en cuatro ocasiones.

Interpretación: el peso de basura orgánica generada que más se repite o con mayor frecuencia en el grupo de estudiantes de la Escuela Preparatoria Regional de Apaxco en el Estado de México, es de 14.3 kg.

6.3.1.2. Mediana (Md)

Definición: en una serie de valores ordenados de mayor a menor o viceversa, es aquel valor que divide de manera exacta a una serie de datos en dos partes de igual tamaño. Es decir, la mitad de los datos o el 50 % de

los mismos se localizan por debajo de la mediana, y la otra mitad, el otro 50 %, por encima de la mediana. Levin y Levin (2004) considera a la **mediana** como la medida de tendencia central que "corta a la distribución en dos partes iguales" (p. 40).

Hernández et al. (2014) señalan que la **mediana** es "el valor que divide la distribución por la mitad" (p. 286). Celis de la Rosa y Labrada (2014), por su parte, consideran que el valor percentilar más conocido es la **mediana**, que se define como "aquel valor que se encuentra en la mitad de una población cuyos valores están ordenados según su magnitud" (p. 44).

Ejemplo: un grupo de prácticas de especialización de la Escuela Nacional de Trabajo Social de la UNAM acudió a la Escuela Preparatoria Regional de Apaxco en el Estado de México, a impartir un curso de educación ambiental. Al finalizar el curso y con objeto de crear conciencia sobre el destino de la basura; el grupo de prácticas solicitó a cada participante pesar la basura orgánica que se generaba en casa su durante una semana.

Tabla 16. ***Estudiantes, según basura orgánica generada***

Apaxco Estado de México. Enero, 2018

Orden	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Kg	12.5	12.7	12.9	13.2	13.7	14.3	14.3	14.3	14.3	14.9	15.2	15.5	15.7	15.9	15.9	16.1	16.4	16.6	16.9	17.5

Fuente: elaboración propia.

Procedimiento: ordenar la serie de menor a mayor o viceversa y localizar el valor que la divida en dos partes de igual tamaño, de tal manera que en una parte quede el 50 % de los datos y en la otra el 50 % restante.

Tabla 17. ***Mediana serie simple***

Mediana (Md) **Serie simple**	**Par**	**Media de los valores centrales**
	Impar	$$Md = \frac{n+1}{2}$$ Donde: n = tamaño de la muestra *1 y 2* = constantes ***Define el lugar donde se encuentra la mediana***

Fuente: elaboración propia.

Serie par:

Figura 16. Mediana serie simple

Estudiantes, según basura orgánica generada
Apaxco Estado de México. Enero, 2018

Orden	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Kg	12.5	12.7	12.9	13.2	13.7	14.3	14.3	14.3	14.3	14.9	15.2	15.5	15.7	15.9	15.9	16.1	16.4	16.6	16.9	17.5

$$Mediana = \frac{14.9 + 15.2}{2} = 15.05$$

Fuente: elaboración propia.

En el ejemplo: la mediana de los valores centrales es 15.05 kg.

Interpretación: la mitad de los estudiantes genera 15.05 kg de basura orgánica o menos y la otra mitad 15.05 kg o más.

Serie impar:

Figura 17. *Mediana serie impar*

Estudiantes, según basura orgánica generada
Apaxco, Estado de México. Enero, 2018

Orden	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Kg	12.5	12.7	12.9	13.2	13.7	14.3	14.3	14.3	14.3	14.9	15.2	15.5	15.7	15.9	15.9	16.1	16.4	16.6	16.9

$$\frac{n+1}{2} = \frac{19+1}{2} = 10\ (Lugar)$$

Fuente: elaboración propia.

En el ejemplo: el valor que ocupa el lugar número 10 es 14.9 kg.

Interpretación: la mitad de los estudiantes genera 14.9 kg de basura orgánica o menos y la otra mitad 14.9 kg o más.

6.3.1.3. Media ($\bar{x}$)

Definición: Es el valor único que resumen a toda una serie de datos. Reynaga et al. (1996) establece que una **media o promedio** "es el valor que tendrían todos los datos de una serie numérica si ellos fueran de igual valor" (p. 100). Bolongna (2013) enfatiza que la **media** es "el valor de la variable que anula la suma de los desvíos en torno suyo" (p. 101).

Si bien Celis de la Rosa y Labrada (2014), Mendenhall, Beaver y Beaver (2010), Triola (2009) o Córdova y Cortés (2010), Elorza (2008), Salazar y Del Castillo (2018), Estuardo (2012), Kelmansky (2009), Hernández et al. (2014) o Ritchey (2004) definen a la media o promedio como el valor que resulta de dividir la suma de todos los valores de una variable entre el número total de datos, la definición que establece Reynaga et al. (1996) facilita la interpretación de datos al darle sentido al valor calculado. Este valor se simboliza como $\bar{x}$.

Ejemplo: un grupo de prácticas de especialización de la Escuela Nacional de Trabajo Social de la UNAM acudió a la Escuela Preparatoria Regional de Apaxco en el Estado de México a impartir un curso de educación ambiental. Al finalizar el curso, y con objeto de crear conciencia sobre el destino de la basura, el grupo de prácticas solicitó a cada participante pesar la basura orgánica que se generaba en casa durante una semana.

Tabla 18. ***Estudiantes, según basura orgánica generada***

Apaxco Estado de México. Enero, 2018.

Número de alumnos	Basura orgánica generada (Kg)	Frecuencia relativa	Frecuencia absoluta acumulada	Frecuencia relativa acumulada
1	12.5	4.2 %	12.5	4.2 %
2	12.7	4.2 %	25.2	8.4 %
3	12.9	4.3 %	38.1	12.7 %
4	13.2	4.4 %	51.3	17.1 %
5	13.7	4.6 %	65	21.7 %
6	14.3	4.8 %	79.3	26.5 %
7	14.3	4.8 %	93.6	31.3 %
8	14.3	4.8 %	107.9	36.1 %
9	14.3	4.8 %	122.2	40.9 %
10	14.9	5.0 %	137.1	45.9 %
11	15.2	5.0 %	152.3	50.9 %
12	15.5	5.2 %	167.8	56.1 %
13	15.7	5.3 %	183.5	61.4 %
14	15.9	5.3 %	199.4	66.7 %
15	15.9	5.3 %	215.3	72 %
16	16.1	5.4 %	231.4	77.4 %
17	16.4	5.5 %	247.8	82.9 %
18	16.6	5.6 %	264.4	88.5 %
19	16.9	5.7 %	281.3	94.2 %
20	17.5	5.8 %	**298.8**	**100 %**
Total	**298.8**	**100 %**	-	-

Fuente: elaboración propia.

Procedimiento: sumar todos los valores y dividir tal suma entre el número de valores que componen a la serie simple.

Tabla 19. ***Media series simples***

Media $(\bar{x})$ **Series simples**	$$\bar{x} = \frac{\Sigma x}{n}$$ Donde: $\bar{x}$= Media Σ= Sumatoria x= Cada uno de los datos n= Tamaño de la muestra

Fuente: elaboración propia.

En el ejemplo: la suma de todos los datos de la muestra es de 298.8 kg. Al sustituir en la fórmula para series simples la media es de 14.94 kg.

$$\bar{x} = \frac{\Sigma x}{n} = \frac{298.8}{20} = 14.94\ Kg$$

Interpretación: Si todos los estudiantes generaran la misma cantidad de basura en casa está sería de 14.94 kg.

6.3.2. De posición

Las medidas **de posición** son valores relativos de una distribución de datos que la dividen en partes iguales. Ríus et al. (1998) señalan que los **estadísticos de posición** son "valores de la variable caracterizados por superar a cierto porcentaje de observaciones en la población (o muestra)" (p. 48).

Las medidas estadísticas aplicables a variables de tipo cuantitativo clasificadas como de posición son **centil, cuartil y decil**.

6.3.2.1. Centil (o percentil)

Definición: son los 99 valores que dividen datos clasificados en 100 partes, con aproximadamente el 1 % de los valores en cada grupo. Un centil se representa con C_r.

De acuerdo con Celis de la Rosa y Labrada (2014) "el término **percentil** deriva de *por ciento*. Cada percentil indica el porcentaje de observaciones que en una serie ordenada de menor a mayor está antes que el valor señalado" (p. 45).

Bologna (2013) define al percentil *r* como "el valor de la variable que deja el *r* por ciento de los casos por debajo de él y *(1-r)* por ciento de los casos por encima. Se indica P_r" (p. 87).

Ejemplo: un grupo de prácticas de especialización de la Escuela Nacional de Trabajo Social de la UNAM acudió a la Escuela Preparatoria Regional de Apaxco, en el Estado de México, a impartir un curso de educación ambiental. Al finalizar el curso y con objeto de crear conciencia sobre el destino de la basura; el grupo de prácticas solicitó a cada participante pesar la basura orgánica que se generaba en casa durante una semana.

Tabla 20. ***Estudiantes, según basura orgánica generada***

Apaxco Estado de México. Enero, 2018.

Orden	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Kg	12.5	12.7	12.9	13.2	13.7	14.3	14.3	14.3	14.3	14.9	15.2	15.5	15.7	15.9	15.9	16.1	16.4	16.6	16.9	17.5

Fuente: elaboración propia.

Procedimiento: ordenar la serie de valores de mayor a menor o viceversa y localizar el valor que la divide en los porcentajes complementarias buscados.

Tabla 21. ***Centil series simples***

Centil (C_r) Series simples	$$\frac{(n)(C_r)}{100}$$ Donde: n = tamaño de la muestra C_r = centil buscado *100* = constante **Define lugar donde se localiza el centil buscado.**

Fuente: elaboración propia.

En el ejemplo: para encontrar el valor del centil 30 (C_{30}) debe localizarse aquel valor que deje el 30 % de los datos con menor o igual magnitud a él y 70 % de los valores con magnitud más grande o igual a él.

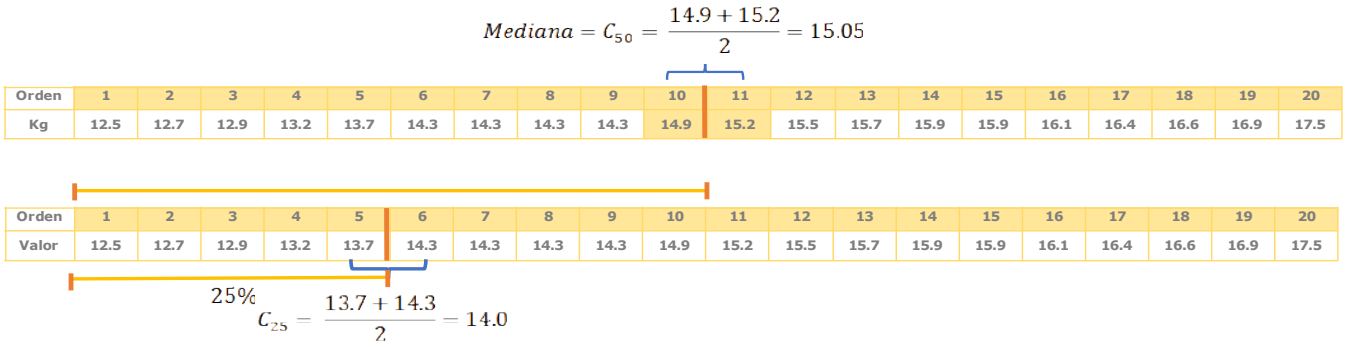
$$\frac{(n)\,(C_r)}{100} = \frac{(20)\,(30)}{100} = 6 \text{ (Lugar)}$$

En el ejemplo: el valor que ocupa el lugar número 6 es 14.3 kg.

Interpretación: válida para el centil 30 (C_{30}): el 30 % de los estudiantes generaron 14.3 kg de basura o menos y el 70 % restante produjo 14.3 kg o más.

Para el caso del centil 25 (C_{25}) es conveniente como primer paso –una vez ordenados los datos de menor a mayor o viceversa– calcular la mediana de toda la serie o centil 50 (C_{50}), lo que indica que se cuenta con dos partes de igual tamaño; con la primera parte de los datos se determina la mitad de esa mitad –en estricto sentido una mediana de ese subconjunto de datos– y con ello se conocerá el 25 % o centil 25 (C_{25}) de toda la serie datos.

Figura 18. ***Centil 25 series simples***



Fuente: elaboración propia.

Interpretación: válida para el centil 25 (C_{25}): el 25 % de los estudiantes generaron 14.0 kg de basura o menos, mientras que el 75 % restante produjo 14.0 kg o más.

Para el caso del centil 50 (C_{50}) se aplica el procedimiento de la mediana; es decir, ordenar la serie de menor a mayor o viceversa y localizar el valor que la divida en dos partes de igual tamaño, de tal manera que en una parte quede el 50 % de los datos y en la otra el 50% restante. En este sentido, **la mediana (Md) es igual al centil 50 (C50)**.

Figura 19. ***Centil 50 series simples***

$$Mediana = C_{50} = \frac{14.9 + 15.2}{2} = 15.05$$

Orden	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Kg	12.5	12.7	12.9	13.2	13.7	14.3	14.3	14.3	14.3	14.9	15.2	15.5	15.7	15.9	15.9	16.1	16.4	16.6	16.9	17.5

Fuente: elaboración propia.

Interpretación: válida para el centil 50 (C_{50}): el 50 % de los estudiantes generaron 15.05 kg de basura o menos y el otro 50% restante produjo 15.05 kg o más.

Para el caso del centil 75 (C_{75}) es conveniente como primer paso –una vez ordenados los datos de menor a mayor o viceversa– calcular la mediana de toda la serie o centil 50 (C_{50}), lo que indica que se cuenta con dos partes de igual tamaño; con la segunda parte de los datos se determina la mitad de esa mitad –en estricto sentido una mediana de ese subconjunto de datos– y con ello se conocerá el 75 % o centil 75 (C_{75}) de toda la serie datos.

Figura 20. ***Centil 75 series simples***

$$Mediana = C_{50} = \frac{14.9 + 15.2}{2} = 15.05$$

Orden	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Kg	12.5	12.7	12.9	13.2	13.7	14.3	14.3	14.3	14.3	14.9	15.2	15.5	15.7	15.9	15.9	16.1	16.4	16.6	16.9	17.5

Orden	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Valor	12.5	12.7	12.9	13.2	13.7	14.3	14.3	14.3	14.3	14.9	15.2	15.5	15.7	15.9	15.9	16.1	16.4	16.6	16.9	17.5

75%

$$C_{75} = \frac{15.9 + 16.1}{2} = 16.0$$

Fuente: elaboración propia.

Interpretación: válida para el centil 75 (C_{75}): el 75 % de los estudiantes generaron 16.0 kg de basura o menos y el 25 % restante produjo 14.0 kg o más.

Lo anterior permite mantener la congruencia con las definiciones señaladas para los centiles; así como para los cuartiles y deciles que se revisarán a continuación.

6.3.2.2. Cuartil (Q_r)

Definición: son los tres valores que dividen datos ordenados en cuatro segmentos, con aproximadamente el 25 % de los valores en cada grupo. Se expresan con la letra Q_1, Q_2 y Q_3; los cuales determinan los valores correspondientes al 25 %, 50 % y 75 % de los datos.

Triola (2009) señala que en el primer cuartil *(Q_1)* al menos el 25 % de los valores ordenados son menores o iguales que Q_1, y al menos el 75 % de los valores son mayores o iguales que Q_1. En el segundo cuartil *(Q_2)* se separa el 50 % inferior de los valores ordenados del 50 % superior, lo que es equivalente a la mediana. En el tercer cuartil *(Q_3)* al menos el 75 % de los valores ordenados son menores o iguales que Q_3, y al menos el 25 % de los valores son mayores o iguales que Q_3.

Ejemplo: un grupo de prácticas de especialización de la Escuela Nacional de Trabajo Social de la UNAM acudió a la Escuela Preparatoria Regional de Apaxco en el Estado de México, a impartir un curso de educación ambiental. Al finalizar el curso y con objeto de crear conciencia sobre el destino de la basura; el grupo de prácticas solicitó a cada participante pesar la basura orgánica que se generaba en casa durante una semana.

Tabla 22. ***Estudiantes, según basura orgánica generada***

Apaxco Estado de México. Enero, 2018.

Orden	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Kg	12.5	12.7	12.9	13.2	13.7	14.3	14.3	14.3	14.3	14.9	15.2	15.5	15.7	15.9	15.9	16.1	16.4	16.6	16.9	17.5

Fuente: elaboración propia.

Procedimiento: ordenar la serie de valores de mayor a menor o viceversa y localizar el valor del cuartil buscado.

Tabla 23. ***Cuartil series simples***

Cuartil (Q_r) Series simples	$$\frac{(n)(Q_r)}{4}$$ Donde: n= tamaño de la muestra Q_r = cuartil buscado 4= constante **Define el lugar donde se encuentra el cuartil buscado.**

Fuente: elaboración propia.

En el ejemplo: para encontrar el valor del cuartil 1 (Q_1) debe localizarse aquel valor que deje una cuarta parte de los datos con menor o igual magnitud a él y a las tres cuartas partes restantes de los valores con magnitudes más grandes o iguales a él.

$$\frac{(n)\ (\ Q_r)}{4} = \frac{(20)\ (1)}{4} = 5\ (\text{Lugar})$$

En el ejemplo: el valor que ocupa el lugar número 5 es 13.7 kg. Sin embargo, para lograr una mayor exactitud en el cálculo de los cuartiles es conveniente aplicar el siguiente procedimiento.

Para el caso del cuartil 1 (Q_1) es conveniente como primer paso –una vez ordenados los datos de menor a mayor o viceversa– calcular la mediana de toda la serie o centil 50 (C_{50}), lo que indica que se cuenta con dos partes de igual tamaño. Con la primera parte de los datos se determina la mitad de esa mitad –en estricto sentido una mediana de ese subconjunto de datos– y con ello se conocerá el 25 % o cuartil 1 (Q_1) de toda la serie datos. En este sentido, **el cuartil 1 (Q_1) es igual al centil 25 (C_{25})**.

Figura 21. ***Cuartil 1 (Q_1)***

$$Mediana = C_{50} = \frac{14.9 + 15.2}{2} = 15.05$$

Orden	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Kg	12.5	12.7	12.9	13.2	13.7	14.3	14.3	14.3	14.3	14.9	15.2	15.5	15.7	15.9	15.9	16.1	16.4	16.6	16.9	17.5

Orden	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Valor	12.5	12.7	12.9	13.2	13.7	14.3	14.3	14.3	14.3	14.9	15.2	15.5	15.7	15.9	15.9	16.1	16.4	16.6	16.9	17.5

25%

$$Q_1 = C_{25} = \frac{13.7 + 14.3}{2} = 14.0$$

Fuente: elaboración propia.

Interpretación: válida para el centil 25 (C_{25}): el 25 % de los estudiantes generaron 14.0 kg de basura o menos y el 75 % restante produjo 14.0 kg o más.

Para el caso del cuartil 2 (Q_2) se aplica el procedimiento de la mediana; es decir, ordenar la serie de menor a mayor o viceversa y localizar el valor que la divida en dos partes de igual tamaño, de tal manera que en una parte quede el 50 % de los datos y en la otra el 50 % restante. En este sentido, **la mediana (Md) es igual al cuartil 2 (Q_2) y centil 50 (C_{50})**.

Figura 22. ***Cuartil 2 (Q_2)***

$$Mediana = Q_2 = C_{50} = \frac{14.9 + 15.2}{2} = 15.05$$

Orden	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Kg	12.5	12.7	12.9	13.2	13.7	14.3	14.3	14.3	14.3	14.9	15.2	15.5	15.7	15.9	15.9	16.1	16.4	16.6	16.9	17.5

Fuente: elaboración propia.

Interpretación: válida para el cuartil 2 (Q_2): el 50 % de los estudiantes generaron 15.05 kg de basura o menos y el otro 50 % restante produjo 15.05 kg o más.

Para el caso del cuartil 3 (Q_3) es conveniente como primer paso –una vez ordenados los datos de menor a mayor o viceversa– calcular la mediana de toda la serie o centil 50 (C_{50}), lo que indica que se cuenta con dos partes de igual tamaño; con la segunda parte de los datos se determina la mitad de esa mitad –en estricto sentido una mediana de ese subconjunto de datos– y con ello se conocerá el 75 % o centil 75 (C_{75}) de toda la serie datos. En este sentido, **el cuartil 3 (Q_2) es igual al centil 75 (C_{75})**.

Figura 23. ***Cuartil 3 (Q_2)***

$$Mediana = Q_2 = C_{50} = \frac{14.9 + 15.2}{2} = 15.05$$

Orden	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Kg	12.5	12.7	12.9	13.2	13.7	14.3	14.3	14.3	14.3	14.9	15.2	15.5	15.7	15.9	15.9	16.1	16.4	16.6	16.9	17.5

Orden	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Valor	12.5	12.7	12.9	13.2	13.7	14.3	14.3	14.3	14.3	14.9	15.2	15.5	15.7	15.9	15.9	16.1	16.4	16.6	16.9	17.5

75%

$$Q_3 = C_{75} = \frac{15.9 + 16.1}{2} = 16.0$$

Fuente: elaboración propia.

Interpretación: válida para el cuartil 3 (Q_3): el 75 % de los estudiantes generaron 16.0 kg de basura o menos, y el otro 25 % restante produjo 16.0 kg o más.

6.3.2.3. Decil (D_r)

Definición: son los nueve valores que dividen la serie de datos en diez partes iguales. Se expresan con la letra ***D_r***; cada uno representa el 10 %.

Para Ruiz (2004) los **deciles** son "los valores de la variable que dividen a la distribución en las partes iguales, cada una de las cuales engloba el 10 % de los datos" (p. 14). Definición similar a la establecida por Ríus et al. (1998), quienes establecen que los **deciles** son "los valores de la variable que dividen a las observaciones en 10 grupos de igual tamaño" (p. 49).

Ejemplo: un grupo de prácticas de especialización de la Escuela Nacional de Trabajo Social de la UNAM acudió a la Escuela Preparatoria Regional de Apaxco en el Estado de México, a impartir un curso de educación ambiental. Al finalizar el curso, y con objeto de crear conciencia sobre el destino de la basura, el grupo de prácticas solicitó a cada participante pesar la basura orgánica que se generaba en casa durante una semana.

Tabla 24. ***Estudiantes, según basura orgánica generada***

Apaxco Estado de México. Enero, 2018.

Orden	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Kg	12.5	12.7	12.9	13.2	13.7	14.3	14.3	14.3	14.3	14.9	15.2	15.5	15.7	15.9	15.9	16.1	16.4	16.6	16.9	17.5

Fuente: elaboración propia.

Procedimiento: ordenar la serie de valores de mayor a menor o viceversa y localizar el valor del decil buscado.

Tabla 25. ***Decil series simples***

Decil (D_r) **Series simples**	$\frac{(n)(D_r)}{10}$ Donde: n= tamaño de la muestra D_r = decil buscado 10= constante **Define el lugar donde se encuentra el decil buscado**

Fuente: elaboración propia.

En el ejemplo: para encontrar el valor del decil 7 (D_7) debe localizarse aquel valor que deje 70 % de los datos con menor o igual magnitud a él, y el 30 % restante de los valores con magnitudes más grandes o iguales a él.

$$\frac{(n)\ (D_r)}{10} = \frac{(20)\ (7)}{10} = 14\ (\text{Lugar})$$

En el ejemplo: el valor que ocupa el lugar número 14 es 15.9 kg.

Interpretación: válida para el decil 7 (D_7): el 70 % de los estudiantes generaron 15.9 kg de basura o menos y el otro 30 % restante produjo 15.9 kg o más.

Para el caso del decil 5 (D_5) se aplica el procedimiento de la mediana; es decir, ordenar la serie de menor a mayor o viceversa y localizar el valor que la divida en dos partes de igual tamaño, de tal manera que en una parte quede el 50 % de los datos y en la otra el 50 % restante. En este sentido, **la mediana (Md) es igual al cuartil 2 (Q_2), decil 5 (D_5) y centil 50 (C_{50})**.

Figura 24. ***Decil 5 (D_5)***

$$Mediana = Q_2 = D_5 = C_{50} = \frac{14.9 + 15.2}{2} = 15.05$$

Orden	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Kg	12.5	12.7	12.9	13.2	13.7	14.3	14.3	14.3	14.3	14.9	15.2	15.5	15.7	15.9	15.9	16.1	16.4	16.6	16.9	17.5

Fuente: elaboración propia.

Interpretación: válida para el decil 5 (D_5): el 50 % de los estudiantes generaron 15.05 kg de basura o menos y el otro 50 % restante produjo 15.05 kg o más.

6.3.3. De dispersión o variabilidad

Para Hernández et al. (2014) las **medidas de la variabilidad** "indican la dispersión de los datos en la escala de medición de la variable considerada y responden a la pregunta: ¿dónde están diseminadas las puntuaciones o los valores obtenidos?" (p. 287). Gorgas et al. (2011) señalan que estas medidas "indicarán la variabilidad de los datos en torno a su valor promedio, es decir si se encuentran muy o poco esparcidos en torno a su centro" (p. 30). Es decir, como establece Celis de la Rosa y Labrada (2014),

> la dispersión de un conjunto de observaciones se refiere a la variedad que exhiben sus valores. Si todos los valores son los mismos, no existe dispersión; si no lo son, hay dispersión en los datos. La magnitud de la dispersión puede ser pequeña cuando los valores, aunque diferentes, están próximos entre sí. Si los valores están ampliamente "diseminados", la dispersión es mayor. (p. 45)

Las medidas estadísticas aplicables a variables de tipo cuantitativo clasificadas como de dispersión son **rango, rango intercuartil, desviación estándar, varianza, coeficiente de variación**.

6.3.3.1. Rango (R)

Definición: es la diferencia entre el valor mayor y el valor menor de una serie de datos. Levin y Levin (2004) definen al **rango** como "la diferencia entre el puntaje más alto y el más bajo de la distribución" (p. 56).

Hernández et al. (2014) establecen que, al **rango**, también se le llama **recorrido** y lo definen como "la diferencia entre la puntuación mayor y la puntuación menor, e indica el número de unidades en la escala de medición que se necesitan para incluir los valores máximo y mínimo" (p. 288). Además, establecen que cuanto más grande sea el rango, mayor será la dispersión de los datos de una distribución.

Ejemplo: un grupo de prácticas de especialización de la Escuela Nacional de Trabajo Social de la UNAM acudió a la Escuela Preparatoria Regional de Apaxco en el Estado de México, a impartir un curso de educación ambiental. Al finalizar el curso, y con objeto de crear conciencia sobre el destino de la basura, el grupo de prácticas solicitó a cada participante pesar la basura orgánica que se generaba en casa durante una semana.

Tabla 26. ***Estudiantes, según basura orgánica generada***

Apaxco Estado de México. Enero, 2018.

Orden	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Kg	12.5	12.7	12.9	13.2	13.7	14.3	14.3	14.3	14.3	14.9	15.2	15.5	15.7	15.9	15.9	16.1	16.4	16.6	16.9	17.5

Fuente: elaboración propia.

Procedimiento: encontrar, por sustracción o resta, la diferencia entre el valor más grande de la serie ($X_{máx}$) y el valor más pequeño ($X_{mín}$).

Tabla 27. ***Rango series simples***

Rango (R) ***Series simples***	$\text{Rango} = (X_{máx} - X_{mín})$ Donde: $X_{máx}$ = valor máximo o mayor $X_{mín}$ = valor mínimo o menor

Fuente: elaboración propia.

En el ejemplo: el valor más grande es 17.5 y el valor más pequeño es 12.5. Al sustituir los datos en la fórmula para series simples el rango es de 5 kg.

$$\text{Rango} = (17.5 - 12.5) = \ 5\ kg$$

Interpretación: la diferencia entre el estudiante que más genera basura y el que menos lo hace es de 5 kg.

6.3.3.2. Rango intercuartílico (RIQ)

Definición: es la diferencia entre el cuartil 3 (Q_3) y el cuartil 1(Q_1) en una serie de datos. Se expresa con las letras *RIQ*.

Ejemplo:

Tabla 28. ***Estudiantes, según basura orgánica generada***

Apaxco Estado de México. Enero, 2018.

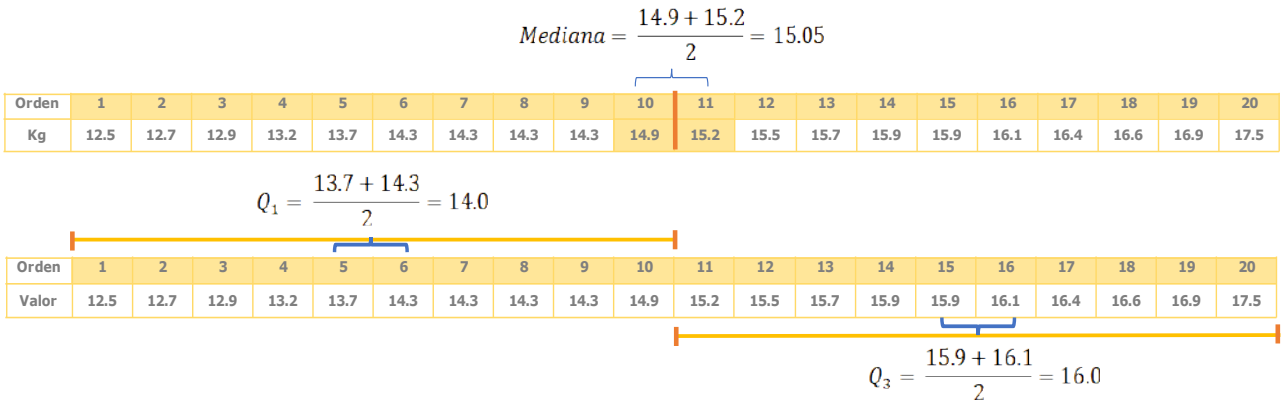
Orden	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Kg	12.5	12.7	12.9	13.2	13.7	14.3	14.3	14.3	14.3	14.9	15.2	15.5	15.7	15.9	15.9	16.1	16.4	16.6	16.9	17.5

Fuente: elaboración propia.

Procedimiento: ordenar la serie de valores de mayor a menor o viceversa; calcular la mediana de toda la serie, lo que indica que se cuenta con dos partes de igual tamaño; con la primera parte de los datos se determina la mitad de esa mitad –en estricto sentido, una mediana de ese subconjunto de datos– y con ello se conocerá el 25 % o cuartil 1 (Q_1) de toda la serie datos. Con la segunda parte de los datos se determina la mitad

de esa mitad –en estricto sentido, una mediana de ese subconjunto de datos– y con ello se conocerá el 75 % o centil 75 *(C_{75})* de toda la serie datos; con ambos resultados, calcular la diferencia entre ellos.

Figura 25. ***Rango intercuartílico (RIQ)***



Fuente: elaboración propia.

Tabla 29. ***Rango intercuartílico de series simples***

Rango intercuartílico (RIQ) Series simples	$RIQ = (Q_3 - Q_1)$

Fuente: elaboración propia.

En el ejemplo: el *RIQ* equivale a 2.0 kg.

$$RIQ = (Q_3 - Q_1) = 16.0 - 14.0 = 2.0\ Kg$$

Interpretación: la diferencia entre el 50 % central de kilogramos de basura generados por los estudiantes es de 2.0.

6.3.3.3. Desviación estándar (s)

Definición: es la raíz cuadrada del promedio de desviaciones de las puntuaciones cuadráticas con respecto a la media de una serie de datos. Es la raíz cuadrada de la varianza. Blalock (1986) define a la **desviación estándar** como "la raíz cuadrada de la media aritmética de las desviaciones cuadradas con respecto la media" (p. 93).

Levin y Levin (2004) definen a la **desviación estándar** como "la raíz cuadrada de la media de las desviaciones de la media de una distribución elevadas al cuadro" (p. 59). Por otra parte, Kelmansky (2009) señala que el **desvío estándar** representa "una distancia típica de cualquier punto del conjunto de datos a su centro (medido por la media). Es una distancia promedio de cada observación a la media (p. 131). Moncho (2015) sostiene que el objetivo es calcular un valor representativo que exprese cuánto se aleja cada dato

respecto a la media. Mientras mayor sea este valor promedio de las distancias, mayor será la separación entre los datos y la media, lo que indica un nivel más alto de dispersión o variabilidad en el conjunto observado (p. 10).

El símbolo **σ (sigma)** se utiliza para representar la desviación estándar de una población, mientras que se usa para señalar la desviación estándar de una muestra.

Ejemplo: un grupo de prácticas de especialización de la Escuela Nacional de Trabajo Social de la UNAM acudió a la Escuela Preparatoria Regional de Apaxco en el Estado de México, a impartir un curso de educación ambiental. Al finalizar el curso y con objeto de crear conciencia sobre el destino de la basura; el grupo de prácticas solicitó a cada participante pesar la basura orgánica que se generaba en casa su durante una semana.

Tabla 30. ***Estudiantes, según basura orgánica generada***

Apaxco Estado de México. Enero, 2018

Orden	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Kg	12.5	12.7	12.9	13.2	13.7	14.3	14.3	14.3	14.3	14.9	15.2	15.5	15.7	15.9	15.9	16.1	16.4	16.6	16.9	17.5

Fuente: elaboración propia.

Procedimiento:

Paso 1. Obtener la media ($\bar{x}$) de la serie de valores:

Tabla 31. ***Desviación estándar***

Series simples	$\bar{x} = \frac{\Sigma x}{n}$ Donde: $\bar{x}$ = media Σ = sumatoria x = cada uno de los datos n = tamaño de la muestra

Fuente: elaboración propia.

Paso 2. Calcular la desviación real de cada dato (x) con respecto a la media $(x - \bar{x})$

Paso 3. Elevar al cuadrado cada una de las desviaciones $(x - \bar{x})^2$

Paso 4. Obtener la sumatoria de las desviaciones cuadráticas $\Sigma(x - \bar{x})^2$

Paso 5. Dividir la sumatoria de las desviaciones cuadráticas $\Sigma(x - \bar{x})^2$entre el número de datos menos uno.

$$\frac{\Sigma(x-\bar{x})^2}{n-1}$$

Paso 6. Obtener la raíz cuadrada.

$$s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}}$$

En el ejemplo:

Paso 1. La suma de todos los datos de la muestra es de 298.8 kg. Al sustituir en la fórmula para series simples la media es de 14.94 kg.

$$\bar{x} = \frac{\Sigma x}{n} = \frac{298.8}{20} = 14.94 \text{ kg}$$

Paso 2 a 6. Calcular la desviación real de cada dato (x) con respecto a la media $(x - \bar{x})$. Elevar al cuadrado cada una de las desviaciones $(x - \bar{x})^2$. Obtener la sumatoria de las desviaciones cuadráticas $\Sigma(x - \bar{x})^2$. Dividir la sumatoria de las desviaciones cuadráticas $\Sigma(x - \bar{x})^2$entre el número de datos menos uno $\frac{\Sigma(x-\bar{x})^2}{n-1}$ y obtener la raíz cuadrada. Como se muestra a continuación:

Tabla 32. ***Estudiantes, según basura orgánica generada***

Apaxco Estado de México. Enero, 2018.

Número de alumnos	Basura orgánica generada (Kg) x	Desviación real de cada dato con respecto a la media $(x - \bar{x})$	Desviación real cuadrática de cada dato con respecto a la media $(x - \bar{x})^2$
1	12.5	-2.44	5.95
2	12.7	-2.24	5.01
3	12.9	-2.04	4.16
4	13.2	-1.74	3.03
5	13.7	-1.24	1.54
6	14.3	-0.64	0.41
7	14.3	-0.64	0.41
8	14.3	-0.64	0.41
9	14.3	-0.64	0.41
10	14.9	-0.04	0.00
11	15.2	0.26	0.07
12	15.5	0.56	0.31
13	15.7	0.76	0.58
14	15.9	0.96	0.92
15	15.9	0.96	0.92
16	16.1	1.16	1.35
17	16.4	1.46	2.13
18	16.6	1.66	2.76
19	16.9	1.96	3.84
20	17.5	2.56	6.55
Total	**298.8**		$\Sigma(x - \bar{x})^2 = 40.76$

Fuente: Apaxco Estado de México (2018).

Sustituyendo:

$$s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}} = \sqrt{\frac{40.76}{20 - 1}} = 1.46 \text{ kg}$$

Interpretación 1: la producción de basura de los estudiantes de la Escuela Preparatoria Regional de Apaxco en el Estado de México se desvía en promedio 1.46 kg de la media de grupo.

Para una interpretación más amplia y con mayor referencia sobre que nos dice la desviación estándar es necesario considerar las **propiedades de la curva normal**.

a) Es un **polígono de frecuencias** en forma de campana para el que están calculadas sus áreas en función de los diversos valores del eje horizontal o del eje de las X o abscisas.
b) En el eje de las X o abscisas se encuentran valores de tipo cuantitativo continuo, genéricamente denominados **puntuaciones Z**, cuyas magnitudes teóricamente pueden ir de cero y de izquierda a derecha desde **menos infinito (-∞) a infinito (∞)**.
c) La media ($\bar{x}$ en una muestra o μ en una población) de todos los valores Z de la abscisa equivale a cero, pues la mitad son negativos y la mitad son positivos. En el sitio de la abscisa que corresponde al cero, es decir **la media, se encuentra la parte más alta de la curva**. En este sitio **también se encuentra la mediana** de todos los valores Z de la abscisa, pues el 50 % de ellos está antes del cero y el 50 % restante se encuentra después.
d) La curva es **simétrica** alrededor de la media; esto es, hay una mitad izquierda que es reflejo de la mitad derecha. Es decir, la asimetría es cero, la mitad de la curva es exactamente igual a la otra mitad.
e) En el eje de las X o abscisas existen **segmentos unitarios de igual longitud y de tamaño 1**. Los segmentos a la izquierda de la media tienen signo negativo y los segmentos a la derecha de la media tienen signo positivo. Tales segmentos, denominados **desviaciones estándar (S),** pueden dividirse en fracciones infinitamente pequeñas y continuas.
f) La curva es **asintótica**; es decir, sus extremos teóricamente nunca tocan la abscisa. Por ello, la longitud de la abscisa podría ser infinitamente larga; sin embargo, se acostumbra la gráfica solo hasta la distancia de tres segmentos a la izquierda y a la derecha de la media.
g) **Toda el área bajo la curva equivale a 1 o 100 %.** Por lo anterior, el área a la izquierda de la media equivale a 0.5 o 50 % y el área a la derecha de la media equivale también a 0.5 o 50 %.
h) Es **unimodal;** es decir, presenta una sola moda.
i) Es una función particular entre desviaciones con respecto a la media de una distribución y la probabilidad de que éstas ocurran.
j) El área que se encuentra sobre el segmento de la abscisa que va desde la media hasta el valor Z de -1 o -1s equivale a 0.3413 o 34.13 %; por simetría, el área que se encuentra sobre el segmento que va desde la media hasta el valor *Z* de 1 o *1s* de la abscisa también equivale a 0.3413 o 34.13 %. En conjunto **el área de -1z a 1z o -1s a 1s equivale a 0.6826 o 68.27 % central de los datos.** El área que se encuentra sobre el segmento de la abscisa que va más allá del **valor Z de -1 equivale a 0.1587 o 15.87 %**; por simetría, el área que se encuentra sobre el segmento que va más allá (hacia infinito) del valor *Z* de 1 de la abscisa también equivale a 0.1587 o 15.87 %.
k) El área que se encuentra sobre el segmento de la abscisa que va desde la media hasta **el valor Z de -2 o -2s equivale a 0.4772 o 47.72 %**; por simetría, el área que se encuentra sobre el segmento que va desde la media hasta el valor *Z* de 2 o *2s* de la abscisa también equivale a 0.4772 o 47.72 %. En conjunto **el área de -2z a 2z o -2s a 2s equivale a 0.9544 o 95.44 % central de los datos.** El área que se encuentra sobre el segmento de la abscisa que va más allá del **valor Z de -2 equivale a 0.0228 o 2.28 %**; por simetría, el área que se encuentra sobre el segmento que va más allá (hacia infinito) del valor *Z* de 2 de la abscisa también equivale a 0.0228 o 2.28 %.
l) El área que se encuentra sobre el segmento de la abscisa que va desde la media hasta **el valor Z de -3o -3s equivale a 0.4987 o 49.87 %**; por simetría, el área que se encuentra sobre el segmento que va desde la media hasta el valor *Z* de 3 o *3s* de la abscisa también equivale a 0.4987 o 49.87 %. En conjunto

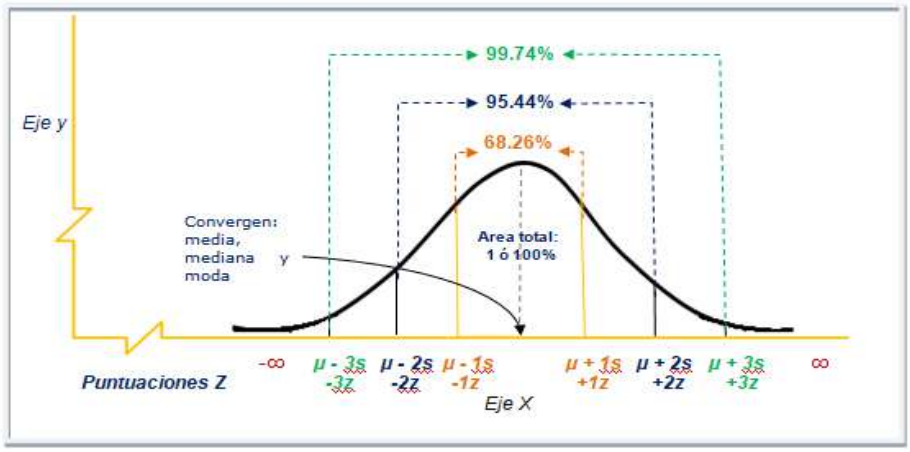
el área de -3z a 3z o -3s a 3s equivale a 0.9974 o 99.74% central de los datos. El área que se encuentra sobre el segmento de la abscisa que va más allá del **valor Z de -3 equivale a 0.0013 o 0.13 %**; por simetría, el área que se encuentra sobre el segmento que va más allá (hacia infinito) del valor *Z* de 3 de la abscisa también equivale a 0.0013 o 0.13 %.

m) Es **mesocúrtica.** El valor de su curtosis equivale a cero.

n) La **media, la mediana y la moda** coinciden en el mismo punto.

o) Para cualquier segmento de la abscisa, y aún para fracciones de segmento, se encuentran calculadas las áreas correspondientes en la tabla específicamente diseñada para tal efecto la **tabla de la distribución normal**.

Figura 26. ***Propiedades de la curva normal***



Fuente: elaboración propia.

Interpretación 2: considerando las propiedades de la curva normal y bajo el supuesto de que los datos se distribuyen de manera normal o semejante a la curva normal es posible plantear las siguientes preguntas:

¿Cuántos kilogramos de basura orgánica genera el 68.26 % central de los estudiantes de la Escuela Preparatoria Regional de Apaxco en el Estado de México?

Derivado de las propiedades de la curva normal se sabe que la distancia de una desviación estándar negativa (-1s) a una desviación estándar positiva (1s) es de 68.26 % y que el valor de la media ($\bar{x}$) se localiza justo en el centro de la curva. Entonces, a la media (14.94 kg) se le resta una desviación estándar (1.46 kg) y se obtiene el valor de: 13.48 kg:

$$\bar{x} - 1s = 14.94 - 1.46 = 13.48 \text{ kg}$$

Por otra parte, a la media (14.94 kg) se le suma una desviación estándar (1.46 kg) y se obtiene el valor de: 16.40 kg:

$$\bar{x} + 1s = 14.94 + 1.46 = 16.40 \text{ kg}$$

Respuesta: **el 68.26 % central de los estudiantes de la Escuela Preparatoria Regional de Apaxco en el Estado de México genera entre 13.48 kg y 16.40 kg**.

Asimismo, es importante conocer el comportamiento del resto de los datos, por lo cual se hacen los siguientes cuestionamientos:

¿Cuántos kilogramos de basura orgánica genera el 15.87 % superior de los estudiantes?

¿Cuántos kilogramos de basura orgánica genera el 15.87 % inferior de los estudiantes?

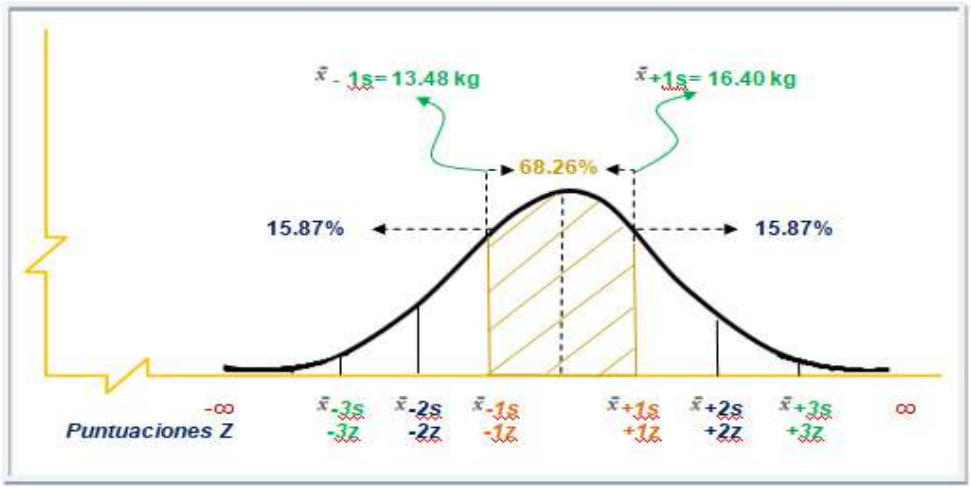
Se sabe que la distancia de una desviación estándar negativa (-1s) a una desviación estándar positiva (1s) es de 68.26 % y que el valor de la media ($\bar{x}$) y la mediana se localiza justo en el centro de la curva; por tanto, la mitad de los datos o el 50 % de los mismos se localizan por debajo de la mediana, y la otra mitad, el otro 50 %, por encima de la mediana, lo que significa que es simétrica; entonces, se deduce que el 15.87 % de los datos se refiere a cada uno de los extremos de la curva. Es decir, si asumimos que la curva es simétrica al dividir el 68.26 % entre 2 se obtiene 34.13 % del lado positivo y del lado negativo; si solo la mitad vale 50 % al restar a este valor el 34.13 % se obtiene 15.87 %, en cada extremo.

Respuesta:

El 15.87 % superior de los estudiantes genera 16.40 kg de basura orgánica o más.

El 15.87 % inferior de los estudiantes genera 13.48 kg de basura orgánica o menos.

Figura 27. *Distribución de datos*



Fuente: elaboración propia.

6.3.3.3.1. Áreas bajo la curva normal y puntuaciones Z

El uso de la desviación estándar está íntimamente relacionado con las propiedades de la curva normal y en especial con los valores continuos denominados puntuaciones Z. Para profundizar en el tema se exponen los siguientes apartados ejemplificando su uso en la descripción de datos.

a) Áreas bajo la curva normal

Supóngase que se midió el peso de 3200 alumnos de la ENTS-UNAM y una vez aplicado los procedimientos correspondientes se encontró que los datos asumían una distribución semejante a la curva normal con una desviación estándar de 5 kilogramos y una media de 75 kilogramos.

De conformidad con las propiedades de la curva normal se pueden señalar –entre otras– las siguientes aseveraciones:

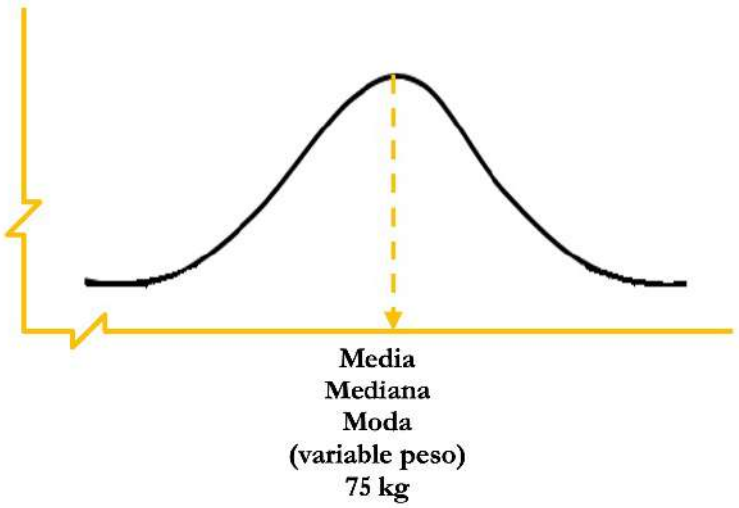
1. El **total de valores** considerados bajo la curva normal es el peso de 3200 alumnos, lo que equivale a **1 o 100 % de los casos**.

Figura 28. ***Total de valores 1 o 100 %***

Fuente: elaboración propia.

2. La **media, mediana y moda** convergen en un solo punto que se encuentra justo en el centro de la curva, y de acuerdo con los datos obtenidos equivale a 75 kilogramos.

Figura 29. ***Convergencia de media, mediana y moda***

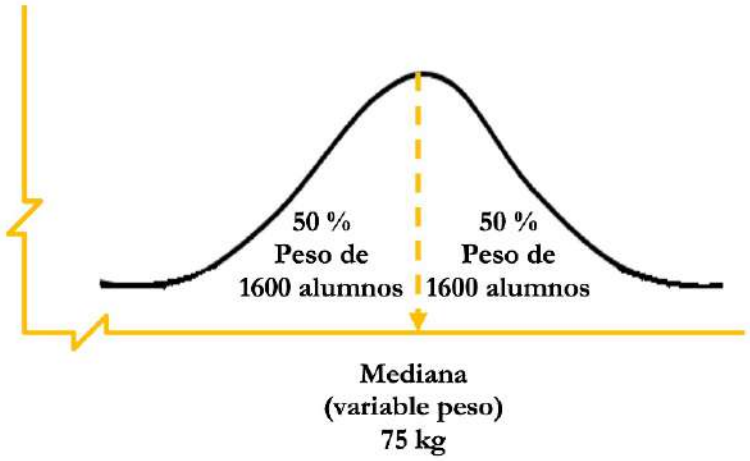


Fuente: elaboración propia.

Debido a que la curva es **simétrica** la mediana se encuentra justo en el centro de la curva y converge en el mismo punto con la media y la moda, por lo cual en este caso su valor es de a 75 kilogramos. En términos de

concentración de valores se tienen 1600 datos a la izquierda de la mediana –la mitad de la curva o el 50 % de los datos– y 1600 valores a la derecha de la mediana –la otra mitad de la curva o el otro 50% de los datos–. Recuérdese que la **mediana** es el valor que divide a una serie de datos en dos partes de igual tamaño. De tal manera que el 50 % de los alumnos pesa 75 kilogramos o menos y el otro 50 %, 75 kilogramos o más.

Figura 30. ***Curva simétrica***



Fuente: elaboración propia.

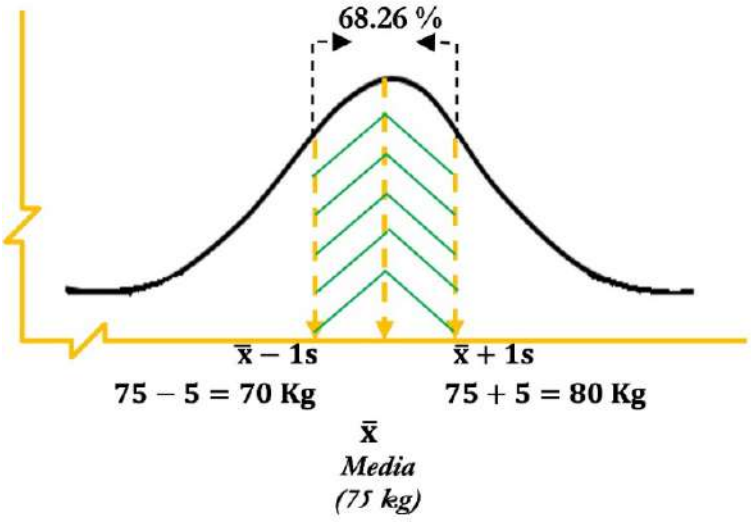
3. La distancia de **una desviación estándar negativa (-1s) a una desviación estándar positiva (1s)** es de **68.26 %,** lo cual implica en el ejemplo que:

   a) El **68.26 % central** de los alumnos tiene un peso de 70 a 80 kilogramos, debido a que:

La media menos una desviación estándar **(x̄- 1s)** es igual a 75 – 5 = **70 kilogramos**.

La media más una desviación estándar **(x̄ + 1s)** es igual a 75 + 5 = **80 kilogramos**.

Figura 31. ***Área 1s a -1s***



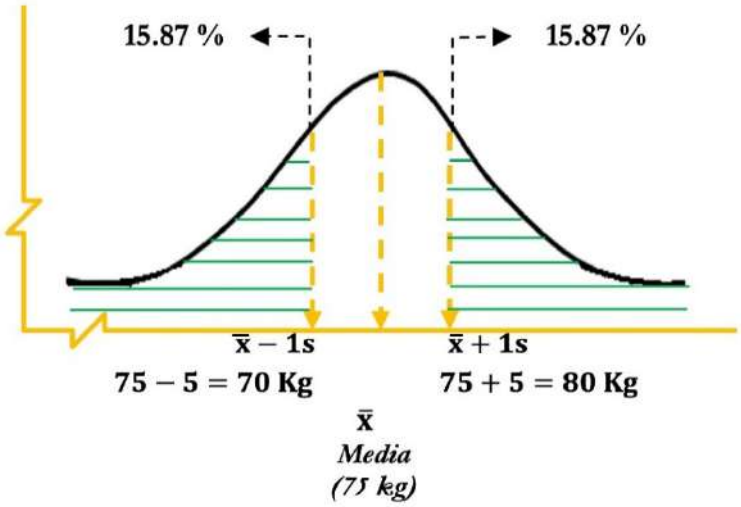
Fuente: elaboración propia.

b) Si **toda el área bajo la curva equivale a 100 %**, al restarle el 68.28 % central se obtiene el restante 31.74 % dado que la curva es **simétrica** este valor se divide entre dos y cada extremo equivale a 15.87 % por lo cual se puede establecer que:

El **15.87 % inferior** de los alumnos tiene un peso de **70 kilogramos o menos**.

El **15.87 % superior** de los alumnos tiene un peso de **80 kilogramos o más**.

Figura 32. ***Área de -1s a -∞ y de 1s a ∞***



Fuente: elaboración propia.

c) Dado que la curva es simétrica, si dividimos el 68.26 % entre dos cada extremo equivale a 34.13 %; por lo tanto:

El **34.13 % central inferior** de los alumnos tienen un peso de **70 a 75 kilogramos**.

El **34.13 % central superior** de los alumnos tienen un peso de **75 a 80 kilogramos**.

Figura 33. ***Área de la media menos 1s***

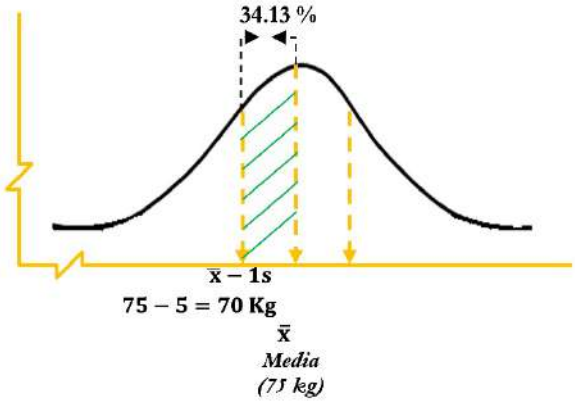
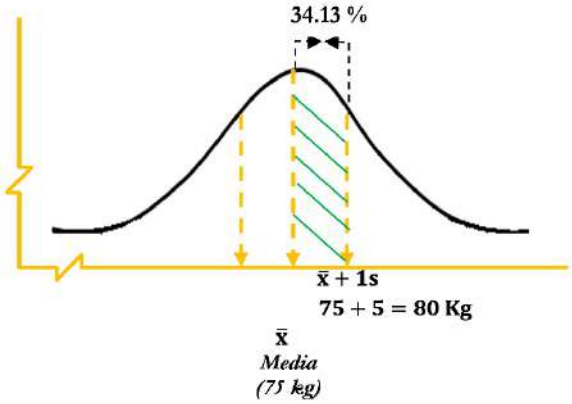


Figura 34. ***Área de la media más 1s***



Fuente: elaboración propia.

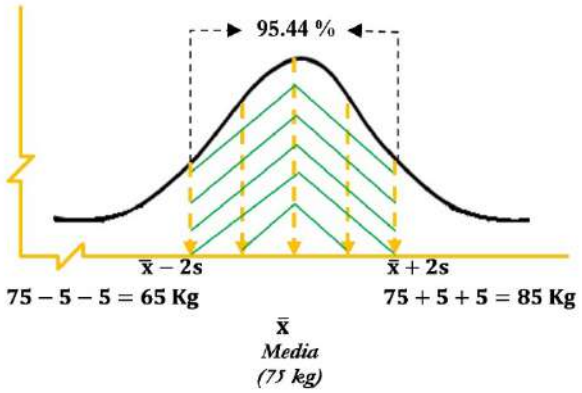
4. La distancia de **dos desviaciones estándar negativas (-2s) a dos desviaciones estándar positivas (2s)** es de **95.44 %** lo cual implica en el ejemplo que:

   d) El **95.44 % central** de los alumnos tiene un peso de 65 a 85 kilogramos, debido a que:

La media menos dos desviaciones estándar **($\bar{x}$-2s)** es igual a 75 -5 - 5= **65 kilogramos**.

La media más dos desviaciones estándar **($\bar{x}$ + 2s)** es igual a 75 + 5+ 5= **85 kilogramos**.

Figura 35. ***Área de 2s a -2s***



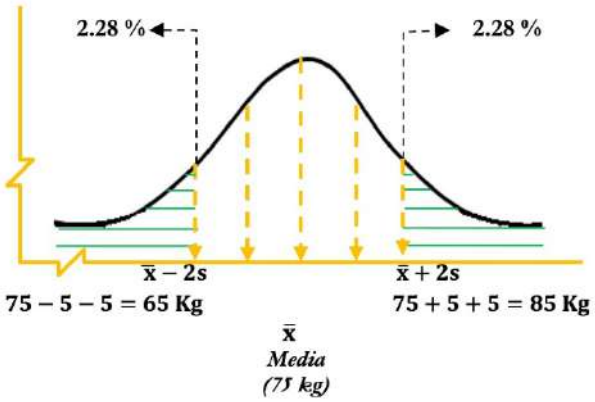
Fuente: elaboración propia.

   e) Si **toda el área bajo la curva equivale a 100 %**, al restarle el 95.44 % central se obtiene el restante 4.56 %, dado que la curva es **simétrica** este valor se divide entre dos y cada extremo equivale a 2.28 %, por lo cual se puede establecer que:

El **2.28 % inferior** de los alumnos tiene un peso de **65 kilogramos o menos**.

El **2.28 % superior** de los alumnos tiene un peso de **85 kilogramos o más**.

Figura 36. ***Área de -2s a -∞ y de 2s a ∞***



Fuente: elaboración propia.

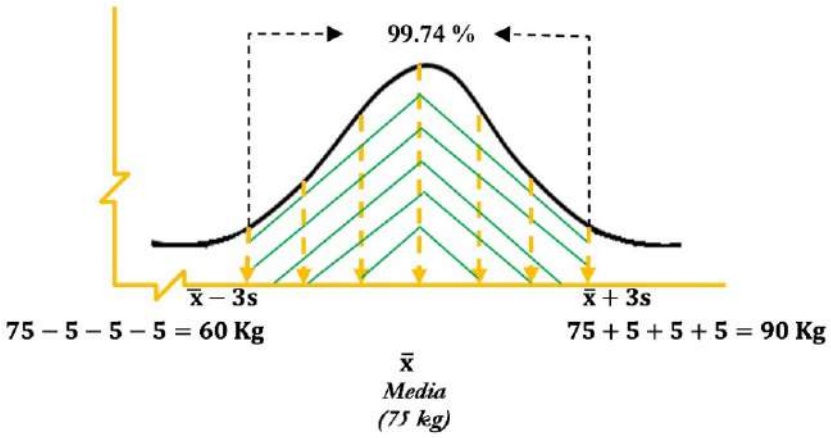
5. La distancia de **tres desviaciones estándar negativas (-3s) a tres desviaciones estándar positivas (3s)** es de **99.74 %** lo cual implica en el ejemplo que:

f) El **99.74 % central** de los alumnos tiene un peso de 60 a 90 kilogramos, debido a que:

La media menos tres desviaciones estándar **($\bar{x}$-3s)** es igual a 75 - 5 -5 – 5= **60 kilogramos**.

La media más tres desviaciones estándar **($\bar{x}$ + 3s)** es igual a 75 ++5+5+5= **90 kilogramos**.

Figura 37. ***Área de -3s a 3s***



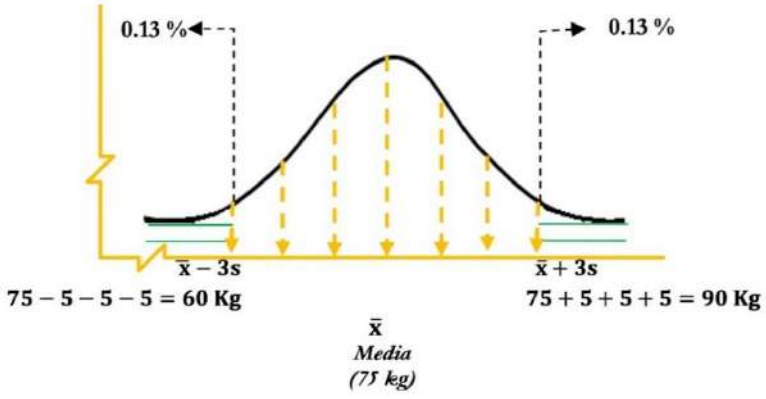
Fuente: elaboración propia.

g) Dado que **toda el área bajo la curva equivale a 100 %**, al restarle 99.74 % central se tiene un 0.26 %, ya que la curva es **simétrica** este valor se divide entre dos y cada extremo tiene un valor de 0.13 %, por lo cual se puede establecer que:

El **0.13 % inferior** de los alumnos tiene un peso de **60 kilogramos o menos**.

El **0.13 % superior** de los alumnos tiene un peso de **90 kilogramos o más**.

Figura 38. ***Área de -3s a -∞ y de 3s a ∞***



Fuente: elaboración propia.

b) Puntuaciones Z

Para conocer cualquier valor debajo de la curva es importante auxiliarse de la siguiente fórmula:

Tabla 33. *Puntuaciones Z*

Puntuaciones Z	$$z = \frac{x - \bar{x}}{s}$$ Donde: Z = puntuación Z. Valor teórico debajo de la curva x = valor buscado $\bar{x}$ = media de la serie de datos s = desviación estándar de la serie de datos

Fuente: elaboración propia.

Si se retoma el ejemplo anterior, se tiene: supóngase que se midió el peso de 3200 alumnos de la ENTS-UNAM, una vez aplicados los procedimientos correspondientes se encontró que los datos asumían una distribución semejante a la curva normal con una desviación estándar de 5 kilogramos y un promedio de 75 kilogramos.

Si se desea conocer la proporción o porcentaje de alumnos que pesan de 75 a 80.3 kilogramos, la proporción o porcentaje de alumnos que pesan 80.3 kilogramos o más, o la proporción o porcentaje de alumnos que pesan 80.3 kilogramos o más, es necesario aplicar el siguiente procedimiento:

1. Determinar la puntuación Z correspondiente al valor buscado mediante la siguiente fórmula:

$$z = \frac{x - \bar{x}}{s} = \frac{80.3 - 75}{5} = 1.06$$

Donde:
Z = puntuación Z. Valor teórico debajo de la curva
x = valor buscado = 80.3 kilogramos
$\bar{x}$ = media de la serie de datos =75 kilogramos
s = desviación estándar de la serie de datos = 5 kilogramos

2. Buscar en la *Tabla de áreas bajo la curva normal* (ver Anexo 1) Columna (1) Puntuación Z el valor de 1.06 con ello, determinar áreas específicas como se muestra a continuación:

Tabla 34. ***Tabla de áreas bajo la curva normal (segmento)***

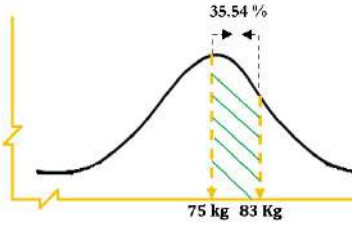
(1) Puntuación Z	(2) Distancia de Z ala media	(3) Área de la parte mayor	(4) Área de la parte menor
1.06	**0.3554**	**0.8554**	**0.1446**
	es decir,	es decir,	es decir,
	35.54 %	**85.54 %**	**14.46 %**

Fuente: elaboración propia.

Si los valores señalados se reflejan en una curva normal se encuentran las respuestas a las interrogantes antes planteadas. En el ejemplo se tiene que:

La proporción o porcentaje de alumnos que pesan de **75 a 80.3 kilogramos es de 0.3554 o 35.54 %**

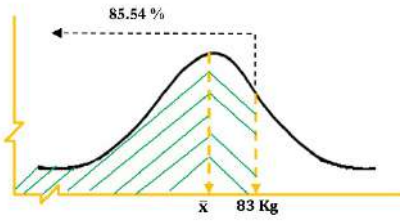
Figura 39. ***Área de la media a puntuación Z***



Es decir, si se sabe que la media es de 75 kilogramos –la cual está en el centro dado que la curva es normal– y que en una puntuación Z de 1.06 se encuentra el valor de 80.3 kilogramos; entonces **la proporción de alumnos que pesan de 75 a 80.3 kilogramos es de 0.3554 o 35.54 %**.

La proporción o porcentaje de alumnos que pesan **80.3 kilogramos o menos es de 0.8554 o 85.54 %**

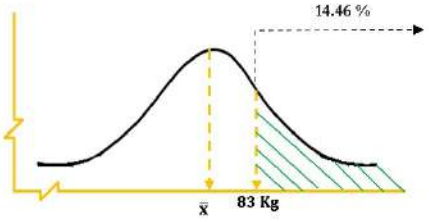
Figura 40. ***Área de la parte mayor de la distribución de datos***



Es decir, si se sabe que en una puntuación Z de 1.06 se encuentra el valor de 80.3 kg; entonces **la proporción de alumnos que pesan menos de 80.3 kg es de 0.8554 o 85.54 %.** Recuérdese que en una curva normal se encuentran puntuaciones bajas, puntuaciones alrededor de la media y puntuaciones altas, en este caso se señala de la **puntuación Z de 1.06 o de 80.3 kg hacia la izquierda dado que se trata de los pesos inferiores o menores a 80.3 kg**.

La proporción o porcentaje de alumnos que pesan **80.3 kilogramos o más es de 0.1446 o 14.46 %**

Figura 41. ***Área de la parte menor de la distribución de datos***



Es decir, si se sabe que en una puntuación Z de 1.06 se encuentra el valor de 80.3 kilogramos; entonces la proporción de alumnos que pesan más de 80.3 kilogramos es de 0.1446 o 14.46 %. Recuérdese que en una curva normal se cuenta con puntuaciones bajas, puntuaciones alrededor del promedio y puntuaciones altas, en este caso se señala de la **puntuación Z de 1.06 o de 80.3 kg hacia la derecha dado que se trata de los pesos superiores o mayores a 80.3 kg**.

Fuente: elaboración propia.

Para conocer la **cantidad de personas u objetos contenidos en un determinado segmento** o

proporción es necesario aplicar una **regla de tres**:

Tabla 35. ***Regla de tres***

Regla de tres	Es una relación de proporcionalidad entre tres valores conocidos (a, b y c) y una incógnita (x). $\frac{a}{c} \longrightarrow \frac{b}{x}$ $\boldsymbol{a}$ es directamente proporcional a $\boldsymbol{b}$, mientras que $\boldsymbol{c}$ será directamente proporcional a $\boldsymbol{x}$. Es decir: $\frac{a}{c} = \frac{b}{x}$ entonces, $a \,.\, \boldsymbol{x} = b \,.\, c$ Al despejar $\boldsymbol{x}$ se deduce: $\boldsymbol{x} = \frac{b \,.\, c}{a}$

Fuente: elaboración propia.

En estricto sentido la regla de tres no es tan complicada como parece. Recuperemos el problema de estudio: supóngase que se midió el peso de 3200 alumnos de la ENTS-UNAM y una vez aplicados los procedimientos correspondientes se encontró que los datos asumían una distribución semejante a la curva normal con una desviación estándar de 5 kilogramos y un promedio de 75 kilogramos.

Si se conoce –como ya se revisó en el apartado anterior– que el **68.26 % central** de los alumnos tiene un peso de 70 a 80 kilogramos, que el 100 % de alumnos es de 3200 y además se desea saber a cuántos alumnos equivale el 68.28 % central de los datos, se cumplen las condiciones para aplicar una regla de tres; es decir, se tiene los tres valores conocidos $\boldsymbol{a}$ (100 %); $\boldsymbol{b}$ (3200 alumnos); $\boldsymbol{c}$ (68.26 %) y una incógnita $\boldsymbol{x}$ (¿A cuántos alumnos equivale el 68.28 % central de los datos?). Al aplicar el procedimiento se tiene:

Tabla 36. ***Regla de 3, peso de alumnos de la ENTS-UNAM***

Regla de tres	$\frac{a}{c} = \frac{b}{x}$; *es decir,* $\frac{100\ \%}{68.26\ \%} = \frac{3200}{\text{x}}$ $a\,.\,\boldsymbol{x} = c.\,b$; así, $\boldsymbol{100}$**(x)= (3200) (68.26)** Al despejar $\boldsymbol{x}$ se deduce: $\boldsymbol{x} = \frac{b\,.\,c}{a}$; *es decir;* $\boldsymbol{x} = \frac{\mathbf{(3200)\ (68.26)}}{\mathbf{100}}$ $= \mathbf{2184}\ \boldsymbol{alumnos}$

Fuente: elaboración propia.

Si se quiere conocer a cuántos alumnos equivale el **15.87 % inferior** de los datos, el procedimiento se ajusta a una regla de tres, como se muestra a continuación:

Tabla 37. ***Regla de 3, peso de alumnos de la ENTS-UNAM***

Regla de tres	$\frac{a}{c} = \frac{b}{x}$; *es decir,* $\frac{100\ \%}{15.87\ \%} = \frac{3200}{x}$ $a\,.\,\boldsymbol{x} = b.\,c$; así, **100(x)= (3200) (15.87)** Al despejar $\boldsymbol{x}$ se deduce: $\boldsymbol{x} = \frac{b\,.\,c}{a}$; *es decir;* $\boldsymbol{x} = \frac{\mathbf{(3200)\ (15.87)}}{\mathbf{100}}$ $= \mathbf{508}\ \boldsymbol{alumnos}$

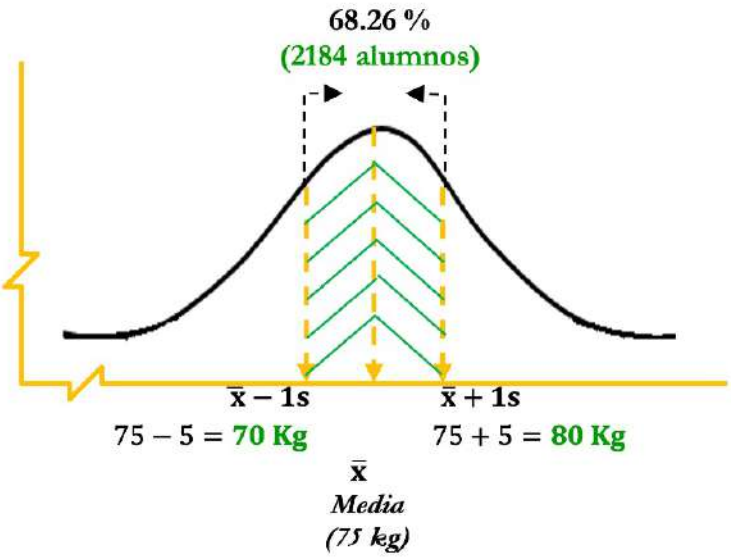
Fuente: elaboración propia.

Como la curva es simétrica, en el otro **15.87 % superior** de los datos se encuentra también 508 alumnos. Si sumamos todos los alumnos de los procedimientos realizados encontraremos que se trata de 3200 o el 100 % de los datos. Por lo anterior, se puede concluir que:

El 68.26 % central de los alumnos tiene un peso de 70 a 80 kilogramos lo que equivale a 2184 alumnos; 508 alumnos tienen un peso de 70 kilogramos o menos, es decir, el 15.87 % inferior de los datos, en tanto que el 15.87 % superior de los alumnos tiene un peso de 80 kilogramos o más; esto es, los restantes 508 alumnos.

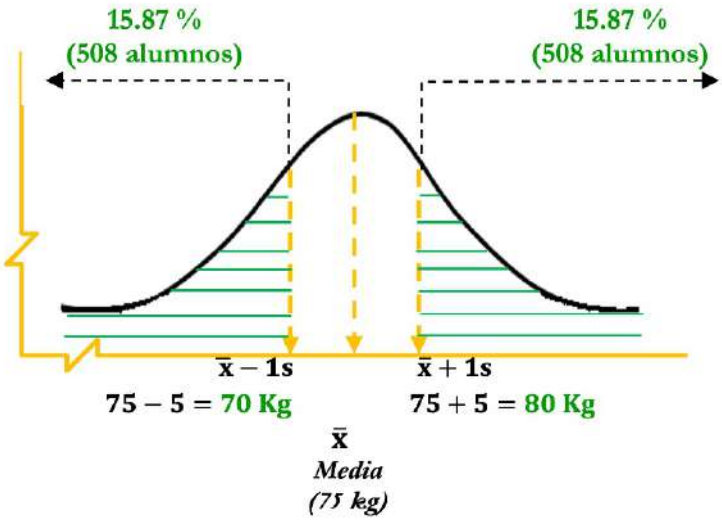
La conclusión antes señalada se refleja en las siguientes gráficas:

Figura 42. ***Distribución de datos 68.26% central (porcentaje y equivalencia)***



Fuente: elaboración propia.

Figura 43. ***Distribución de datos 15.87% inferior y superior (porcentaje y equivalencia)***



Fuente: elaboración propia.

6.3.3.4. Varianza (S^2)

Definición: es la desviación estándar elevada al cuadrado o el cuadrado del promedio de desviación de las puntuaciones cuadráticas con respecto a la media de una serie de datos. Bologna (2013) establece que se llama "**varianza** de una distribución a la suma de los cuadrados de los desvíos alrededor de la media, dividida por el total de observaciones menos uno. Se indica S^2" (p. 102).

Es una medida de dispersión que representa la variabilidad de una serie de datos respecto a su media. Mide qué tan dispersos están los datos alrededor de la media, y es igual a la desviación estándar elevada al cuadrado. Celis de la Rosa y Labrada (2014) señalan que la varianza es "una medida de dispersión que describe la separación de los valores en relación con la media" (p. 46).

Ejemplo: un grupo de prácticas de especialización de la Escuela Nacional de Trabajo Social de la UNAM acudió a la Escuela Preparatoria Regional de Apaxco en el Estado de México, a impartir un curso de educación ambiental. Al finalizar el curso, y con objeto de crear conciencia sobre el destino de la basura, el grupo de prácticas solicitó a cada participante pesar la basura orgánica que se generaba en casa su durante una semana.

Tabla 38. ***Estudiantes, según basura orgánica generada***

Apaxco Estado de México. Enero, 2018.

Orden	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Kg	12.5	12.7	12.9	13.2	13.7	14.3	14.3	14.3	14.3	14.9	15.2	15.5	15.7	15.9	15.9	16.1	16.4	16.6	16.9	17.5

Fuente: elaboración propia.

Procedimiento:

1. Obtener la media ($\bar{x}$) de la serie de valores:

Tabla 39. ***Media, series simples***

Series simples	$\bar{x} = \frac{\Sigma x}{n}$ Donde: $\bar{x}$= Media Σ= Sumatoria x= Cada uno de los datos n= Tamaño de la muestra

Fuente: elaboración propia.

2. Calcular la desviación real de cada dato (x) con respecto a la media $(x - \bar{x})$
3. Elevar al cuadrado cada una de las desviaciones $(x - \bar{x})^2$
4. Obtener la sumatoria de las desviaciones cuadráticas $\Sigma(x - \bar{x})^2$
5. Dividir la sumatoria de las desviaciones cuadráticas entre el número de datos menos uno. $S^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$

Tabla 40. ***Estudiantes, según basura orgánica generada***

Apaxco Estado de México. Enero, 2018.

Número de alumnos	Basura orgánica generada (kg) x	Desviación real de cada dato con respecto a la media $(x-\bar{x})$	Desviación real cuadrática de cada dato con respecto a la media $(x-\bar{x})^2$
1	12.5	-2.44	5.95
2	12.7	-2.24	5.01
3	12.9	-2.04	4.16
4	13.2	-1.74	3.03
5	13.7	-1.24	1.54
6	14.3	-0.64	0.41
7	14.3	-0.64	0.41
8	14.3	-0.64	0.41
9	14.3	-0.64	0.41
10	14.9	-0.04	0.00
11	15.2	0.26	0.07
12	15.5	0.56	0.31
13	15.7	0.76	0.58
14	15.9	0.96	0.92
15	15.9	0.96	0.92
16	16.1	1.16	1.35
17	16.4	1.46	2.13
18	16.6	1.66	2.76
19	16.9	1.96	3.84
20	17.5	2.56	6.55
Total	298.8		$\Sigma(x-\bar{x})^2$= 40.76

Fuente: elaboración propia.

Sustituyendo:

$$S^2 = \frac{\Sigma(x-\bar{x})^2}{n-1} = \frac{40.76}{20-1} = 2.14\ Kg^2$$

En una primera interpretación tendríamos que la producción de basura es de 2.14 kilogramos al cuadrado. Es conveniente mencionar que siempre que se calcula la varianza se tendrán unidades de medida al cuadrado, para transformar dicha medida a un valor más accesible tendremos que aplicar una raíz cuadrada; lo que equivale a la desviación estándar.

$$s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = \sqrt{\frac{40.76}{20-1}} = 1.46\ kg$$

Interpretación:

La producción de basura de los estudiantes de la Escuela Preparatoria Regional de Apaxco en el Estado de México se desvía en promedio 1.46 kg de la media de grupo.

6.3.3.5. Coeficiente de variación (CV)

Definición: es una relación o razón estadística entre la desviación estándar y la media aritmética multiplicada por 100, es decir, se expresa en términos de porcentaje. Bologna (2013) establece que el **coeficiente de variación** "expresa de manera relativa la dispersión, midiendo el peso de la desviación estándar comparado con la media. Se indica ***CV***" (p. 103).

El coeficiente de variación permite comparar la dispersión entre dos poblaciones distintas o contrastar la variación producto de dos variables distintas (que pueden provenir de una misma población). Blalock (1986) establece que es "una medida de variabilidad relativa que divide a la desviación estándar entre la media" (p. 101).

Triola (2009), por su parte, señala que "el coeficiente de variación (CV) de un conjunto de datos muestrales o poblacionales, expresado como porcentaje, describe la desviación estándar con relación a la media" (p. 103).

Ejemplo: un profesor de la ENTS-UNAM está interesado en conocer la variabilidad de las calificaciones obtenidas por dos grupos a los cuales imparte la materia de Estadística Aplicada a la Investigación Social I.

Tabla 41. ***Estudiantes, según calificación en EAIS-I***

ENTS, 2018.

Grupo 1326	8	7	6	9	8	6	9	7	8	10	8	9	7	6	8	8	9	10	9	9
Grupo 1327	7	8	6	7	7	8	9	8	8	7	9	9	8	8	7	8	8	7	8	9

Fuente: elaboración propia.

Procedimiento: obtener la media y la desviación estándar de cada uno de los grupos y aplicar la siguiente fórmula:

Tabla 42. ***Coeficiente de variación, series simples***

Coeficiente de variación (CV) **Series simples**	$$CV = \frac{s}{\bar{x}}$$ Donde: CV = coeficiente de variación $\bar{x}$ = media de los datos s = desviación estándar de los datos

Fuente: elaboración propia.

En el ejemplo:

Grupo 1326

$$\bar{x} = \frac{\Sigma x}{n} = \frac{161}{20} = 8.05$$

$$s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}} = \sqrt{\frac{28.95}{20 - 1}} = 1.23$$

$$CV = \frac{s}{\bar{x}} = \frac{1.23}{8.05} = 0.1533\ x\ 100 = 15.33\ \%$$

Grupo 1327

$$\bar{x} = \frac{\Sigma x}{n} = \frac{156}{20} = 7.8$$

$$s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}} = \sqrt{\frac{13.2}{20 - 1}} = 0.83$$

$$CV = \frac{s}{\bar{x}} = \frac{0.83}{7.80} = 0.10.68\ x\ 100 = 10.68\ \%$$

Interpretación: las calificaciones obtenidas en la materia de Estadística Aplicada a la Investigación presentan menor dispersión en el grupo 1327 con 10.68 %, que en el grupo 1326 con 15.33 %.

6.3.4. De distribución o de forma

Las **medidas de distribución o de forma** son aquellas que reflejan el grado de simetría y concentración de datos respecto a su medida central (media). Este tipo de medidas están directamente relacionadas con el concepto de curva normal y de distribución libre.

Moncho (2015) señala que "las **medidas de forma** proporcionan información sobre el comportamiento de los datos correspondientes a una variable, atendiendo a la simetría o el apuntamiento de la distribución de los mismos" (p. 13).

Las medidas estadísticas aplicables a variables de tipo cuantitativo clasificadas como de distribución o forma son **sesgo y curtosis**.

La asimetría o sesgo y la curtosis son medidas estadísticas que se utilizan para determinar cuánto se parece una distribución de datos a la distribución teórica, llamada **curva normal** o campana de Gauss; así como, dónde se concentran las puntuaciones.

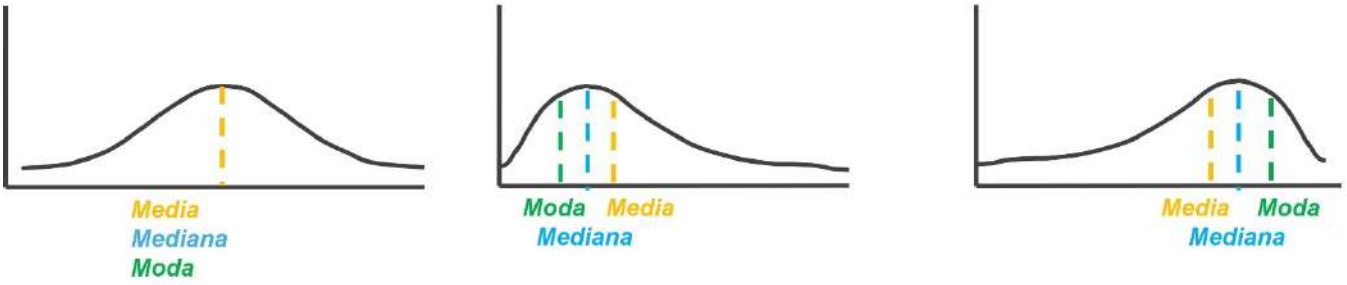
6.3.4.1. Sesgo o asimetría (a_3)

Definición: es la ausencia de simetría en una distribución; es decir, se presenta cuando la mitad de una distribución de datos reflejada en una curva no es exactamente igual o espejo de la otra mitad. Rivera y García (2005) lo definen como "el grado de asimetría de la distribución observada por el número de casos agrupados en una sola dirección" (p. 58).

Cuando una distribución está equilibrada con relación a su eje vertical, se dice que es ***simétrica***; cuando no presenta esta consideración, es ***asimétrica***.

Si una distribución de datos tiene una asimetría de cero (0) la moda, mediana y media convergen en un solo punto; es decir, son iguales y por lo tanto es simétrica; entonces se habla de una **distribución normal**. Cuando hay más valores agrupados hacia la izquierda de la curva (por debajo de la media) la ***asimetría o sesgo es positivo***. Cuando los valores tienden a agruparse hacia la derecha de la curva (por encima de la media) la **asimetría es negativa**.

Figura 44. ***Sesgo o asimetría en una distribución***



Simetría o asimetría cero. La media, mediana y moda convergen en un solo punto.

.**Asimetría o sesgo a la derecha o sesgo positivo.** La media es mayor a la mediana y la moda.

Asimetría o sesgo a la izquierda o sesgo negativo. La media es menor a la mediana y la moda.

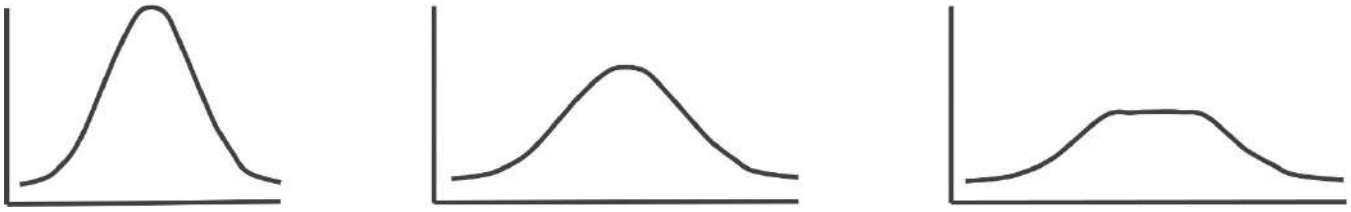
Fuente: elaboración propia.

6.3.4.2. Curtosis (a_4)

Definición: es el grado de concentración de valores de una variable alrededor de la zona central de una distribución que determina apuntamiento, normalidad o allanamiento de una curva. Elorza (2008) la define como "la agudeza que presenta el perfil de una curva unimodal" (p. 60). Mientras que en palabras de Rivera y García (2005) es "el nivel de picudez de una curva, esto es, su grado de elevación o aplanamiento" (p. 59).

Dependiendo del grado de elevación o aplanamiento de una distribución esta puede ser **leptocúrtica**, **mesocúrtica (o normal) o platicúrtica**.

Figura 45. ***Curtosis en una distribución***



Leptocúrtica: se presenta un **apuntamiento** de la curva o mayor concentración de datos alrededor de un valor central (menor dispersión).

Mesocúrtica (o normal): se presenta un **aplanamiento moderado** de la curva o concentración media de datos alrededor de un valor central.

Platicúrtica: se presenta un **aplanamiento pronunciado** de la curva o menor concentración de datos alrededor de un valor central (mayor dispersión).

Fuente: elaboración propia.

Ejemplo: un grupo de prácticas de especialización de la Escuela Nacional de Trabajo Social de la UNAM, como parte de un equipo multidisciplinario, acudió al Hospital Infantil AMM en la Ciudad de México a impartir un curso de educación nutricional.

Con objeto de hacer posteriores comparaciones de los datos y resultados obtenidos, el grupo deseaba saber si el comportamiento del peso de los participantes asumía o no una semejanza a la curva normal.

Procedimiento: de conformidad con el **método de momentos** una distribución de valores cuantitativos tiene semejanza a la curva normal:

a) Si su **sesgo (a_3)** tiene un valor de **-0.5 a 0.5**.

b) Si su **curtosis (a_4)** tiene un valor de **2 a 4**.

Tabla 43. ***Semejanza a la curva normal. Valores de referencia de acuerdo con el método de momentos***

Sesgo	$-0.5 < a_3 < 0.5$
Curtosis	$2 < a_4 < 4$

Fuente: elaboración propia.

Las fórmulas para calcular el sesgo y la curtosis, a través del **método de momentos**, son los siguientes:

Tabla 44. ***Método de momentos***

Sesgo	Curtosis
$a_3 = \frac{m_3}{(\sqrt{m_2})^3}$	$a_4 = \frac{m_4}{(m_2)^2}$

Fuente: elaboración propia.

El cálculo de cada uno de los momentos se hace con las fórmulas siguientes:

Tabla 45. ***Fórmulas para calcular momentos***

Momento 2:	*Momento 3*:	*Momento 4*
$m_2 = \frac{\Sigma(x-\bar{x})^2}{n}$	$m_3 = \frac{\Sigma(x-\bar{x})^3}{n}$	$m_4 = \frac{\Sigma(x-\bar{x})^4}{n}$

Fuente: elaboración propia.

Para obtener el promedio se aplica la siguiente fórmula:

$$\bar{x} = \frac{\Sigma x}{n}$$

Sustituyendo:

$$\bar{x} = \frac{914}{20} = 45.70\ Kg.$$

Tabla 46. *20 niños según peso*

Hospital Infantil AMM. Enero, 2018.

Peso X	Desviación respecto a la media $(x-\bar{x})$	Cuadrado de cada desviación $(x-\bar{x})^2$	Cubo de cada desviación $(x-\bar{x})^3$	Cuarta potencia de cada desviación $(x-\bar{x})^4$
40	-5.70	32.49	-185.19	1055.60
40	-5.70	32.49	-185.19	1055.60
40	-5.70	32.49	-185.19	1055.60
41	-4.70	22.09	-103.82	487.97
42	-3.70	13.69	-50.65	187.42
42	-3.70	13.69	-50.65	187.42
43	-2.70	7.29	-19.68	53.14
43	-2.70	7.29	-19.68	53.14
44	-1.70	2.89	-4.91	8.35
45	-0.70	0.49	-0.34	0.24
46	0.30	0.09	0.03	0.01
46	0.30	0.09	0.03	0.01
47	1.30	1.69	2.20	2.86
48	2.30	5.29	12.17	27.98
49	3.30	10.89	35.94	118.59
49	3.30	10.89	35.94	118.59
50	4.30	18.49	79.51	341.88
52	6.30	39.69	250.05	1575.30
52	6.30	39.69	250.05	1575.30
55	9.30	86.49	804.36	7480.52
Sumatoria	$\Sigma(x-\bar{x})=$ **0.000**	$\Sigma(x-x)^2=$ **378.200**	$\sum(x-\bar{x})^3=$ **664.920**	$\Sigma(x-\bar{x})^4=$ **15385.514**

Fuente: elaboración propia.

Al sustituir en las fórmulas para el cálculo de momentos se tiene:

Tabla 47. *Cálculo de momentos*

Momento 2:	*Momento 3*:	*Momento 4*
$m_2 = \frac{\Sigma(x-\bar{x})^2}{n}$	$m_3 = \frac{\Sigma(x-\bar{x})^3}{n}$	$m_4 = \frac{\Sigma(x-\bar{x})^4}{n}$
$m_2 = \frac{378.20}{20} = 18.91$	$m_3 = \frac{664.920}{20} = 33.25$	$m_4 = \frac{15{,}385.514}{20} = 769.28$

Fuente: elaboración propia.

Finalmente, al usar los valores calculados para los momentos y sustituyendo valores en fórmulas de sesgo y curtosis se tiene:

Tabla 48. ***Cálculo de sesgo y curtosis***

Sesgo	*Curtosis*
$a_3 = \frac{m_3}{(\sqrt{m_2})^3}$	$a_4 = \frac{m_4}{(m_2)^2}$
$a_3 = \frac{33.25}{\left(\sqrt{18.91}\right)^3} = 0.404$	$a_4 = \frac{769.28}{(18.91)^2} = 2.151$

Fuente: elaboración propia.

Interpretación: considerando que los valores obtenidos de sesgo y curtosis se encuentran dentro de los intervalos señalados por el método de momentos se concluye que:

La distribución de los pesos de los niños del Hospital Infantil AMM en la Ciudad de México se asemeja en asimetría y en grado de apuntamiento o aplanamiento a una distribución normal.

Si bien para considerar que una distribución es normal los valores del sesgo y la curtosis debieran ser de conformidad con el método de momentos cero (0) y tres (3), respectivamente. Por ello, se puede decir que si bien no son normales, sí se asemejan a una distribución normal.

Capítulo 7
El método estadístico: describir (series agrupadas)

Imagen de CLM 2019

7.1. Describir: medidas de resumen en series agrupadas

Describir implica el cálculo de medidas estadísticas que representan las características generales de un conjunto de datos basados en una muestra; implica obtener resumen para variables cuantitativas, tal es el caso de medidas de tendencia central (moda, mediana, media), de posición (centil, cuartil, decil), de dispersión (rango, rango intercuartil, desviación estándar, varianza, coeficiente de variación) y de distribución (sesgo y curtosis).

En este apartado se estudiarán las medidas estadísticas de resumen comentadas en el párrafo anterior relativas a series agrupadas, sin perder de vista que lo más importante al estudiar datos es conocer *qué nos dicen* cada uno de ellos respecto del problema de estudiar.

Como ya se ha señalado, para alcanzar el objetivo anterior, en la descripción de medida de resumen aplicaremos el *Método DPI* (definición-procedimiento-interpretación); es decir, emplearemos un proceso de entendimiento que parte de fijar con claridad el significado de una palabra (definición), pasa por el cálculo de la prueba (procedimiento) y termina cuando se da razón de ser y referencia al valor obtenido (interpretación).

7.2. Regla de Sturges

La **regla de Sturges** es un criterio utilizado para determinar el número de clases o intervalos necesarios para representar gráficamente un conjunto de datos estadísticos. Creada por el matemático alemán Herbert Sturges en 1926, quien propuso un método sencillo, basado en el número de muestras x que permitiesen encontrar el número de clases y su amplitud de rango.

Ejemplo: supóngase que se recolecta la variable peso en kilogramos de 80 estudiantes de la Escuela Secundaria Técnica No. 13 de Atitalaquia en el estado de Hidalgo. Al contar dichos pesos se encontraron valores repetidos, por lo que dibujar una gráfica con ochenta barras resultaría poco práctico, lo convencional en estos casos es agrupar la información para crear intervalos clase.

Tabla 49. ***Estudiantes, según peso, de la Escuela Secundaria Técnica 13, Atitalaquia, Hidalgo. Enero, 2019 (serie original)***

60	66	77	70	66	68	57	70	66	52	75	65	69	71	58	66	67	74	61	63
69	80	59	66	70	67	78	75	64	71	81	62	64	69	68	72	83	56	65	74
67	54	65	65	69	61	67	73	57	62	67	68	63	67	71	68	76	61	62	63
76	61	67	67	64	72	64	73	79	58	67	71	68	59	69	70	66	62	63	66

Fuente: elaboración propia.

Como primer paso se ordenan los datos de mayor a menor o viceversa, como se muestra a continuación:

Tabla 50. ***Estudiantes, según peso, de la Escuela Secundaria Técnica 13, Atitalaquia, Hidalgo. Enero, 2019(serie ordenada)***

52	54	56	57	57	58	58	59	59	60	61	61	61	61	62	62	62	62	63	63
63	63	64	64	64	64	65	65	65	65	66	66	66	66	66	66	66	67	67	67
67	67	67	67	67	67	68	68	68	68	68	69	69	69	69	69	70	70	70	70
71	71	71	71	72	72	73	73	74	74	75	75	76	76	77	78	79	80	81	**86**

Fuente: elaboración propia.

Una vez ordenados los datos es posible obtener una medida de resumen denominada **rango**; es decir, la diferencia entre el valor mayor o máximo y el valor menor o mínimo de una serie de datos.

En el ejemplo, el valor del rango es 34 kg, como se muestra a continuación:

$$\text{Rango} = (X_{máx} - X_{mín}) = 86 - 52 = 34\ kg$$

Para determinar el **número de intervalos de clase o escalas de clase *K***, se aplica la regla de Sturges de la siguiente manera:

$$K = 1 + 3.222\ Log(N)$$

Donde:

K = es el número de intervalos de clase o escalas de clase

$Log(N)$= es el logaritmo natural de N o total de datos.

Al sustituir en el ejemplo encontramos:

$$K = 1 + 3.222\ Log(80) = 7.12$$

Para calcular el **ancho de los intervalos de clase o ancho de la escala de clase (AC)** se divide el **rango** entre el valor de ***K***:

$$AC = \frac{Rango}{K}$$

Al sustituir en el ejemplo encontramos:

$$AC = \frac{Rango}{K} = \frac{34}{7.12} = 4.77 \approx 5$$

El **valor de *K*** (número de intervalo de clases) es común redondearlo al entero más próximo; en este caso, 5.

7.3. Límite inferior y límite superior

Para la construcción de cada uno de los intervalos de clase o escalas de clase se debe considerar un límite inferior y un límite superior. Se comenzará el conteo con el valor mínimo de la serie o límite inferior; para determinar el límite superior se deberá sumar al valor del límite inferior el ancho del intervalo o ancho de la escala de clase (AC) y restar 1. Este procedimiento se deberá aplicar para definir los 7 intervalos de clase o 7 escalas de clase señaladas mediante la regla de Sturges.

Así, se tienen los siguientes intervalos de clase con sus límites inferiores y superiores correspondientes, como se muestra a continuación:

Tabla 51. ***Estudiantes, según peso. Límite inferior y límite superior***

Límite inferior (LI) = Límite superior + 1	Límite superior (LS) = LI + Ancho de clase -1	Intervalos de clase LI - LS
52	52+5-1= **56**	**52 - 56**
56+1= **57**	57+5-1= **61**	**57 - 61**
61+1= **62**	62+5-1=**66**	**62 - 66**
66+1= **67**	67+5-1=**71**	**67 - 71**
71+1= **72**	71+5-1=**76**	**72 - 76**
76+1= **77**	77+5-1=**81**	**77 - 81**
81+1= **82**	82+5-1=**86**	**82- 86**

Fuente: elaboración propia.

7.4. Determinación de frecuencias

Posteriormente se determina la **frecuencia absoluta y la frecuencia relativa** que corresponden a cada intervalo de clase o escala de clase.

Tabla 52. ***Estudiantes, según peso, de la Escuela Secundaria Técnica 13, Atitalaquia, Hidalgo. Enero, 2019(serie ordenada)***

52	54	56	57	57	58	58	59	59	60	61	61	61	61	62	62	62	62	63	63
63	63	64	64	64	64	65	65	65	65	66	66	66	66	66	66	66	67	67	67
67	67	67	67	67	67	68	68	68	68	68	69	69	69	69	69	70	70	70	70
71	71	71	71	72	72	73	73	74	74	75	75	76	76	77	78	79	80	81	86

Fuente: elaboración propia.

Tabla 53. ***Estudiantes, según peso, de la Escuela Secundaria Técnica 13, Atitalaquia, Hidalgo. Enero, 2019***

Límite inferior (LI)	Límite superior (LS)	Frecuencia absoluta (FA)	Frecuencia relativa (FR)
52	- 56	3	3.8
57	- 61	11	13.7
62	- 66	23	28.7
67	- 71	27	33.7
72	- 76	10	12.5
77	- 81	5	6.3
82	- 86	1	1.3
Total		80	100.0

Fuente: elaboración propia.

La **frecuencia absoluta** se obtiene al contar el número de veces que se presenta un valor dentro del intervalo de clase establecido; en tanto que la **frecuencia relativa** se calcula mediante la división del valor de la frecuencia absoluta de cada intervalo de clase entre en total de datos por 100 para expresarla en términos de porcentaje.

Con objeto de contar con todos los elementos para calcular medidas de resumen para variables cuantitativas se determinará la **frecuencia absoluta acumulada, frecuencia relativa acumulada, centro de clase o punto medio; así como, límite inferior real y límite superior real.**

Para calcular la **frecuencia absoluta acumulada** se suman los valores de la frecuencia absoluta de una clase y de todas las clases precedentes.

En el ejemplo: el intervalo de clase o escala de clase de 52 a 56 acumula **3** frecuencias absolutas; para el intervalo de clase de 57 a 61 se cuenta con 11 frecuencias absolutas, más las tres que se tenían en el intervalo de clase anterior suman **14**. Para el intervalo de clase de 62 a 66 se tienen 23 frecuencias absolutas más 14 que se tenían en el intervalo de clase anterior suman **37**; así sucesivamente hasta obtener para el intervalo de clase de 82 a 86 un acumulado total de **80**. Como se muestra a continuación.

Tabla 54. ***Frecuencia absoluta acumulada. Estudiantes, según peso, de la Escuela Secundaria Técnica 13, Atitalaquia, Hidalgo. Enero, 2019***

Límite inferior (LI)	Límite superior (LS)	Frecuencia absoluta (FA)	Frecuencia relativa (FR)	Frecuencia absoluta acumulada (FAA)
52	56	3	3.8	3
57	61	11	13.7	14
62	66	23	28.7	37
67	71	27	33.7	64
72	76	10	12.5	74
77	81	5	6.3	79
82	86	1	1.3	80
Total		80	100.0	-

Fuente: elaboración propia.

Para determinar la **frecuencia relativa acumulada** se suman los valores de la frecuencia relativa de una clase y de todas las clases precedentes.

En el ejemplo: el intervalo de clase o escala de clase de 52 a 56 acumula **3.8** de frecuencias relativa; para el intervalo de clase de 57 a 61 se cuenta con 13.7 de frecuencias relativa, más las 3.8 que se tenían en el intervalo de clase anterior suman **17.5**. Para el intervalo de clase de 62 a 66 se tienen 28.7 de frecuencia relativa, más 17.5 que se tenían en el intervalo de clase anterior suman **46.2**; así sucesivamente hasta obtener para el intervalo de clase de 82 a 86 un acumulado total de **100**. Como se muestra a continuación.

Tabla 55. *Frecuencia relativa acumulada. Estudiantes, según peso, de la Escuela Secundaria Técnica 13, Atitalaquia, Hidalgo. Enero, 2019*

Límite inferior (LI)	Límite superior (LS)	Frecuencia absoluta (FA)	Frecuencia relativa (FR)	Frecuencia absoluta acumulada (FAA)	Frecuencia relativa acumulada (FRA)
52	56	3	3.8	3	3.8
57	61	11	13.7	14	17.5
62	66	23	28.7	37	46.2
67	71	27	33.7	64	79.9
72	76	10	12.5	74	92.4
77	81	5	6.3	79	98.7
82	86	1	1.3	80	100
Total		80	100.0	-	-

Fuente: elaboración propia.

7.5. Centro de clase o punto medio

El **centro de clase o punto medio** se obtiene sumando el límite inferior más el límite superior entre dos en cada uno de los intervalos de clase. Es decir,

$$CC = \frac{LI + LS}{2}$$

En el ejemplo: para el **primer intervalo de clase** se tiene:

$$CC = \frac{LI + LS}{2} = \frac{52 + 56}{2} = 54$$

Para el **segundo intervalo de clase** se tiene:

$$CC = \frac{LI + LS}{2} = \frac{57 + 61}{2} = 59$$

Así sucesivamente hasta obtener el **último intervalo de clase**, que para el ejemplo es **84**. Como se muestra a continuación.

Tabla 56. *Centro de clase. Estudiantes, según peso, de la Escuela Secundaria Técnica 13, Atitalaquia, Hidalgo. Enero, 2019*

Límite inferior (LI)	Límite superior (LS)	Frecuencia absoluta (FA)	Frecuencia relativa (FR)	Frecuencia absoluta acumulada (FAA)	Frecuencia relativa acumulada (FRA)	Centro de clase (CC)
52	56	3	3.8	3	3.8	54
57	61	11	13.7	14	17.5	59
62	66	23	28.7	37	46.2	64
67	71	27	33.7	64	79.9	69
72	76	10	12.5	74	92.4	74
77	81	5	6.3	79	98.7	79
82	86	1	1.3	80	100	84
Total		80	100.0	-	-	-

Fuente: elaboración propia.

7.6. Límite inferior real (LIR) y límite superior real (LSR)

El **límite inferior real (LIR)** y **límite superior real (LSR)** se determinan de la siguiente manera.

Límite inferior real (LIR): el centro de clase (CC) **menos** el ancho del intervalo de clase o ancho de la escala de clase (AC) entre dos, para en cada uno de los intervalos de clase.

Límite superior real (LSR): el centro de clase (CC) **más** el ancho del intervalo de clase o ancho de la escala de clase (AC) entre dos, para en cada uno de los intervalos de clase.

En el ejemplo: para el **primer intervalo de clase** se tiene:

Tabla 57. ***Límite inferior real y límite superior real en primer intervalo de clase***

Límite inferior real (LIR)	Límite superior real (LSR)
$\boldsymbol{LIR} = CC - \frac{AC}{2}$	$\boldsymbol{LSR} = CC + \frac{AC}{2}$
$\boldsymbol{LIR} = CC - \frac{AC}{2} = 54 - \frac{5}{2} = 51.5$	$\boldsymbol{LSR} = CC - \frac{AC}{2} = 54 + \frac{5}{2} = 56.5$

Fuente: elaboración propia.

Para el **segundo intervalo de clase** se tiene:

Tabla 58. ***Límite inferior real y límite superior real en segundo intervalo de clase***

Límite inferior real (LIR)	Límite superior real (LSR)
$\boldsymbol{LIR} = CC - \frac{AC}{2}$	$\boldsymbol{LSR} = CC + \frac{AC}{2}$
$\boldsymbol{LIR} = CC - \frac{AC}{2} = 59 - \frac{5}{2} = 56.5$	$\boldsymbol{LSR} = CC - \frac{AC}{2} = 59 + \frac{5}{2} = 61.5$

Fuente: elaboración propia.

Así sucesivamente hasta obtener el **último intervalo de clase**, que para el ejemplo tiene un límite inferior real de **81.5** y un límite superior real de **86.5**. Como se muestra a continuación.

Tabla 59. ***Estudiantes, según peso, de la Escuela Secundaria Técnica 13, Atitalaquia, Hidalgo. Enero, 2019***

Límite inferior (LI)	Límite superior (LS)	Frecuencia absoluta (FA)	Frecuencia relativa (FR)	Frecuencia absoluta acumulada (FAA)	Frecuencia relativa acumulada (FRA)	Centro de clase (CC)	Límite inferior real (LIR)	Límite superior real (LSR)
52	56	3	3.8	3	3.8	54	51.5 - 56.5	
57	61	11	13.7	14	17.5	59	56.5 - 61.5	
62	66	23	28.7	37	46.2	64	61.5 - 66.5	
67	71	27	33.7	64	79.9	69	66.5 - 71.5	
72	76	10	12.5	74	92.4	74	71.5 - 76.5	
77	81	5	6.3	79	98.7	79	76.5 - 81.5	
82	86	1	1.3	80	100	84	81.5 - 86.5	
Total		80	100.0	-	-	-	-	-

Fuente: elaboración propia.

7.7. Medidas de resumen

Las medidas estadísticas aplicables variables de tipo cuantitativo en series agrupadas se clasifican en **de tendencia central** (moda, mediana, media), **de posición** (centil, cuartil, decil), **de dispersión** (rango, rango intercuartil, desviación estándar, varianza, coeficiente de variación) y **de distribución** (sesgo y curtosis).

7.7.1. De tendencia central

7.7.1.1. Moda (Mo)

Definición: es el valor de mayor frecuencia u ocurrencia en un conjunto de datos, y es el valor que más se repite en un conjunto de datos (para profundizar en la definición ver apartado: moda de series simples).

Ejemplo: en la Escuela Secundaria Técnica 13 de Atitalaquia en el estado de Hidalgo se realizó un estudio sobre la obesidad en estudiantes de nivel básico, para ello se recolectó la variable peso en kilogramos de una muestra de 80 estudiantes.

Procedimiento: se definen límites inferiores y superiores; frecuencia absoluta y relativa; frecuencia absoluta y relativa acumulada; centro de clase o punto medio y límites inferiores y superiores reales de cada intervalo de clase, y se aplica la siguiente fórmula para el cálculo de la **moda** para datos agrupados.

$$\boldsymbol{Mo} = LIR + \left[\left(\frac{fa}{fa + fp}\right) a\right]$$

Donde:

$\boldsymbol{Mo}$= moda

LIR = límite inferior real del intervalo modal.

fa = frecuencia absoluta anterior a la frecuencia del intervalo modal.

fp= frecuencia absoluta posterior a la frecuencia del intervalo modal

a= tamaño del intervalo o ancho del intervalo de clase, donde: $\boldsymbol{a} = LSR - LIR$

En el ejemplo:

Tabla 60. ***Estudiantes, según peso, de la Escuela Secundaria Técnica 13, Atitalaquia, Hidalgo. Enero, 2019***

Límite inferior (LI)	Límite superior (LS)	Frecuencia absoluta (FA)	Frecuencia relativa (FR)	Frecuencia absoluta acumulada (FAA)	Frecuencia relativa acumulada (FRA)	Centro de clase (CC)	Límite inferior real (LIR)	Límite superior real (LSR)
52	56	3	3.8	3	3.8	54	51.5	56.5
57	61	11	13.7	14	17.5	59	56.5	61.5
62	66	23	28.7	37	46.2	64	61.5	66.5
67	**71**	**27**	**33.7**	**64**	**79.9**	**69**	**66.5**	**71.5**
72	76	10	12.5	74	92.4	74	71.5	76.5
77	81	5	6.3	79	98.7	79	76.5	81.5
82	86	1	1.3	**80**	**100**	84	81.5	86.5
Total		**80**	**100.0**	-	-	-	-	-

Fuente: elaboración propia.

$$\boldsymbol{Mo} = LIR + \left[\left(\frac{fa}{fa + fp}\right) a\right] = \mathbf{66.5} + \left[\left(\frac{\mathbf{23}}{\mathbf{23} + \mathbf{10}}\right) \mathbf{5}\right] = 69.98\ Kg.$$

$$\boldsymbol{a} = LSR - LIR = 71.5 - 66.5 = \mathbf{5}$$

Interpretación: el peso de los estudiantes que más se repite o que tiene mayor frecuencia en el grupo de estudiantes de la Escuela Secundaria Técnica 13 de Atitalaquia en el estado de Hidalgo es de 69.98 kg.

7.7.1.2. Mediana (Md)

Definición: en una serie de valores ordenados de mayor a menor o viceversa es aquel valor que divide en dos partes de igual tamaño. Es decir, del conjunto total de datos, la mitad o el 50 % de los mismos se localizan por debajo de la mediana, y la otra mitad, el otro 50 %, por encima de la mediana (para profundizar en la definición ver apartado: mediana de series simples).

Ejemplo: en la Escuela Secundaria Técnica 13 de Atitalaquia en el estado de Hidalgo se realizó un estudio sobre obesidad en estudiantes de nivel básico, para ello se recolectó la variable peso en kilogramos de una muestra de 80 estudiantes.

Procedimiento: se definen límites inferiores y superiores; frecuencia absoluta y relativa; frecuencia absoluta y relativa acumulada; centro de clase o punto medio y límites inferiores y superiores reales de cada intervalo de clase y se aplica la siguiente fórmula para el cálculo de la **mediana** para datos agrupados.

$$\boldsymbol{Md} = LIR + \left[\left(\frac{\frac{n}{2} - f_{aa}}{f}\right)a\right]$$

Donde:

$\boldsymbol{Md}$ = mediana

LIR = límite inferior real del intervalo, donde la frecuencia relativa acumulada está cerca del 50 %.

n = número de casos.

f_{aa}= frecuencia absoluta acumulada anterior al intervalo, donde cae la mediana (50 %).

f = frecuencia absoluta donde cae la mediana.

a = tamaño del intervalo o ancho del intervalo de clase, donde: $\boldsymbol{a} = LSR - LIR$.

Tabla 61. ***Estudiantes, según peso, de la Escuela Secundaria Técnica 13, Atitalaquia, Hidalgo. Enero, 2019***

Límite inferior (LI)	Límite superior (LS)	Frecuencia absoluta (FA)	Frecuencia relativa (FR)	Frecuencia absoluta acumulada (FAA)	Frecuencia relativa acumulada (FRA)	Centro de clase (CC)	Límite inferior real (LIR)	Límite superior real (LSR)
52	56	3	3.8	3	3.8	54	51.5	56.5
57	61	11	13.7	14	17.5	59	56.5	61.5
62	66	23	28.7	37	46.2	64	61.5	66.5
67	71	27	33.7	64	79.9	69	66.5	71.5
72	76	10	12.5	74	92.4	74	71.5	76.5
77	81	5	6.3	79	98.7	79	76.5	81.5
82	86	1	1.3	80	100	84	81.5	86.5
Total		80	100.0	-	-	-	-	-

Fuente: elaboración propia.

En el ejemplo:

$$\boldsymbol{Md} = LIR + \left[\left(\frac{\frac{n}{2} - f_{aa}}{f}\right)a\right] = 66.5 + \left[\left(\frac{\frac{80}{2} - 37}{27}\right)5\right] = 67.05\ kg.$$

$$\boldsymbol{a} = LSR - LIR = 71.5 - 66.5 = 5$$

Interpretación: la mitad de los estudiantes de la Escuela Secundaria Técnica 13 de Atitalaquia en el estado de Hidalgo pesa 67.05 kg o menos y la otra mitad 67.05 kg o más.

7.7.1.3. Media

Definición: es el valor único que resume a toda una serie de datos, por lo que es el valor que tendrían todos los datos si ellos fueran del mismo valor (para profundizar en la definición ver apartado: media de series simples).

Ejemplo: en la Escuela Secundaria Técnica 13 de Atitalaquia en el estado de Hidalgo se realizó un estudio sobre obesidad en estudiantes de nivel básico, para ello se recolectó la variable peso en kilogramos de una muestra de 80 estudiantes.

Procedimiento: multiplicar la frecuencia de cada una de las clases por su correspondiente centro de clase o punto medio, sumar dicho producto y, finalmente, dividirlo entre el número de datos.

La media para datos agrupados se obtiene utilizando la siguiente fórmula:

$$\bar{x} = \frac{\Sigma fx}{n}$$

Donde:
$\bar{x}$ = media
Σ = sumatoria
f = frecuencia absoluta
x = punto medio o centro de clase
n = tamaño de la muestra

Tabla 62. ***Estudiantes, según peso, de la Escuela Secundaria Técnica 13, Atitalaquia, Hidalgo. Enero, 2019***

Límite inferior (LI)	Límite superior (LS)	Frecuencia absoluta (FA)	Frecuencia relativa acumulada (FRA)	Centro de clase (CC)	Límite inferior real (LIR)	Límite superior real (LSR)	Frecuencia absoluta por punto medio o centro de clase (fx)
52	56	3	3.8	54	51.5	56.5	162
57	61	11	17.5	59	56.5	61.5	649
62	66	23	46.2	64	61.5	66.5	1472
67	71	27	79.9	69	66.5	71.5	1863
72	76	10	92.4	74	71.5	76.5	740
77	81	5	98.7	79	76.5	81.5	395
82	86	1	100	84	81.5	86.5	84
Total		80	-	-	-	-	Σfx= 5365

Fuente: elaboración propia.

En el ejemplo:

$$\bar{x} = \frac{\Sigma fx}{n} = \frac{5365}{80} = 67.06 \text{ kg}$$

Interpretación: si todos los estudiantes de la Escuela Secundaria Técnica 13 de Atitalaquia en el estado de Hidalgo pesaran lo mismo sería de 67.06 kg.

7.7.2. De posición

7.7.2.1. Centil (C_r)

Definición: son los 99 valores que dividen datos clasificados en 100 grupos, con aproximadamente el 1 % de los valores en cada grupo. Un centil se representa con C_r (para profundizar en la definición ver apartado: centil de series simples).

Ejemplo: en la Escuela Secundaria Técnica 13 de Atitalaquia en el estado de Hidalgo se realizó un estudio sobre obesidad en estudiantes de nivel básico para ello se recolectó la variable peso en kilogramos de una muestra de 80 estudiantes.

Procedimiento: se definen límites inferiores y superiores, frecuencia absoluta y relativa, frecuencia absoluta y relativa acumulada, centro de clase o punto medio, y límites inferiores y superiores reales de cada intervalo de clase; finalmente, se aplica la siguiente fórmula para el cálculo del **cetil** para datos agrupados.

$$\boldsymbol{C_r} = LIR + \left[\left(\frac{\frac{rn}{100} - \boldsymbol{f_{aa}}}{f}\right) a\right]$$

Donde:
$\boldsymbol{C_r}$ = centil buscado donde $\boldsymbol{r}$ puede ser 1, 2, 3, … , 99
LIR = límite inferior real del intervalo, donde la frecuencia relativa acumulada está cerca del centil buscado
n = número de casos
$\boldsymbol{f_{aa}}$= frecuencia absoluta acumulada anterior al intervalo, donde cae el centil buscado
f = frecuencia absoluta donde cae el centil buscado
a = tamaño del intervalo o ancho del intervalo de clase. Donde: $\boldsymbol{a} = LSR - LIR$

Tabla 63. ***Estudiantes, según peso, de la Escuela Secundaria Técnica 13, Atitalaquia, Hidalgo. Enero, 2019***

Límite inferior (LI)	Límite superior (LS)	Frecuencia absoluta (FA)	Frecuencia relativa (FR)	Frecuencia absoluta acumulada (FAA)	Frecuencia relativa acumulada (FRA)	Centro de clase (CC)	Límite inferior real (LIR)	Límite superior real (LSR)
52	56	3	3.8	3	3.8	54	51.5 - 56.5	
57	61	11	13.7	14	17.5	59	56.5 - 61.5	
62	66	23	28.7	37	46.2	64	61.5 - 66.5	
67	71	27	33.7	64	79.9	69	66.5 - 71.5	
72	76	10	12.5	74	92.4	74	71.5 - 76.5	
77	81	5	6.3	79	98.7	79	76.5 - 81.5	
82	86	1	1.3	80	100	84	81.5 - 86.5	
Total		80	100.0	-	-	-	-	-

Fuente: elaboración propia.

En el ejemplo: para encontrar el valor del **centil 70 (C_{70})** debe localizarse aquel valor que deje el 70 % de los datos con menor o igual magnitud a él y 30 % de los valores con magnitud más grande o igual a él.

$$C_r = LIR + \left[\left(\frac{\frac{rn}{100} - f_{aa}}{f}\right)a\right] = C_{70} = 66.5 + \left[\left(\frac{\frac{(70)(80)}{100} - 37}{27}\right)5\right] = 70.01 \text{ Kg.}$$

$$a = LSR - LIR = 71.5 - 66.5 = 5$$

Interpretación: válida para el centil 70 (C_{70}): el 70 % de los estudiantes de la Escuela Secundaria Técnica 13 de Atitalaquia en el estado de Hidalgo pesa 70.01 kg o menos, mientras que el 30 % restante pesa 70.01 kg o más.

7.7.2.2. Cuartil (Q)

Definición: son los tres valores que dividen datos ordenados en cuatro grupos, con aproximadamente el 25 % de los valores en cada grupo. Se expresan con las letras Q_1, Q_2 y Q_3, que determinan los valores correspondientes al 25 %, 50 % y 75 % de los datos (para profundizar en la definición ver apartado: cuartil de series simples).

Ejemplo: en la Escuela Secundaria Técnica 13 de Atitalaquia en el estado de Hidalgo se realizó un estudio sobre obesidad en estudiantes de nivel básico, para ello se recolectó la variable peso en kilogramos de una muestra de 80 estudiantes.

Procedimiento: se definen límites inferiores y superiores, frecuencia absoluta y relativa, frecuencia absoluta y relativa acumulada, centro de clase o punto medio, y límites inferiores y superiores reales de cada intervalo de clase; finalmente, se aplica la siguiente fórmula para el cálculo del **cuartil** para datos agrupados.

$$\boldsymbol{Q_r} = LIR + \left[\left(\frac{\frac{rn}{4} - f_{aa}}{f}\right)a\right]$$

Donde:
Q_r = cuartil buscado donde $\boldsymbol{r}$ puede ser 1, 2 o 3
LIR = límite inferior real del intervalo, donde la frecuencia relativa acumulada está cerca del cuartil buscado
n = número de casos
$\boldsymbol{f_{aa}}$ = frecuencia absoluta acumulada anterior al intervalo, donde cae el cuartil buscado
f = frecuencia absoluta donde cae el cuartil buscado
a = tamaño del intervalo o ancho del intervalo de clase, donde: $\boldsymbol{a = LSR - LIR}$

Tabla 64. ***Estudiantes, según peso, de la Escuela Secundaria Técnica 13, Atitalaquia, Hidalgo. Enero, 2019***

Límite inferior (LI)	Límite superior (LS)	Frecuencia absoluta (FA)	Frecuencia relativa (FR)	Frecuencia absoluta acumulada (FAA)	Frecuencia relativa acumulada (FRA)	Centro de clase (CC)	Límite inferior real (LIR)	Límite superior real (LSR)
52	56	3	3.8	3	3.8	54	51.5	56.5
57	61	11	13.7	14	17.5	59	56.5	61.5
62	66	23	28.7	37	46.2	64	61.5	66.5
67	**71**	**27**	**33.7**	**64**	**79.9**	**69**	**66.5**	**71.5**
72	76	10	12.5	74	92.4	74	71.5	76.5
77	81	5	6.3	79	98.7	79	76.5	81.5
82	86	1	1.3	**80**	**100**	84	81.5	86.5
Total		**80**	**100.0**	-	-	-	-	-

Fuente: elaboración propia.

En el ejemplo: para encontrar el valor del **cuartil 3 (Q_3)** debe localizarse aquel valor que deje el 75 % de los datos con menor o igual magnitud a él y 25 % de los valores con magnitud más grande o igual a él.

$$\boldsymbol{Q_r} = LIR + \left[\left(\frac{\frac{rn}{4} - f_{aa}}{f}\right)a\right] = \boldsymbol{Q_3} = \mathbf{66.5} + \left[\left(\frac{\frac{(3)(80)}{4} - 37}{27}\right)5\right] = 70.75\ kg$$

Interpretación: válida para el cuartil 3 (Q_3): el 75% de los estudiantes de la Escuela Secundaria Técnica 13 de Atitalaquia en el estado de Hidalgo pesa 70.75 kg o menos, mientras que el 25 % restante pesa 70.75 kg o más.

7.7.2.3. Decil (D)

Definición: son los nueve valores que dividen la serie de datos en diez partes iguales. Se expresan con la letra D_r, y cada valor representa el 10 % (para profundizar en la definición ver apartado: decil de series simples).

Ejemplo: en la Escuela Secundaria Técnica 13 de Atitalaquia en el estado de Hidalgo se realizó un estudio sobre obesidad en estudiantes de nivel básico, para ello se recolectó la variable peso en kilogramos de una muestra de 80 estudiantes.

Procedimiento: se definen límites inferiores y superiores, frecuencia absoluta y relativa, frecuencia absoluta y relativa acumulada, centro de clase o punto medio, y límites inferiores y superiores reales de cada intervalo de clase; finalmente, se aplica la siguiente fórmula para el cálculo del **decil** para datos agrupados.

$$\boldsymbol{D_r} = LIR + \left[\left(\frac{\frac{rn}{10} - f_{aa}}{f}\right)a\right]$$

Donde:
$\boldsymbol{D_r}$ = decil buscado donde $\boldsymbol{r}$ puede ser 1, 2, 3, …, 9.
LIR = límite inferior real del intervalo, donde la frecuencia relativa acumulada está cerca del decil buscado.
n = número de casos.
$\boldsymbol{f_{aa}}$ = frecuencia absoluta acumulada anterior al intervalo, donde cae el decil buscado.
f = frecuencia absoluta donde cae el decil buscado
a = tamaño del intervalo o ancho del intervalo de clase, donde: $\boldsymbol{a} = LSR - LIR$

Tabla 65. Estudiantes, según peso, de la Escuela Secundaria Técnica 13, Atitalaquia, Hidalgo. Enero, 2019

Límite inferior (LI)	Límite superior (LS)	Frecuencia absoluta (FA)	Frecuencia relativa (FR)	Frecuencia absoluta acumulada (FAA)	Frecuencia relativa acumulada (FRA)	Centro de clase (CC)	Límite inferior real (LIR)	Límite superior real (LSR)
52	56	3	3.8	3	3.8	54	51.5	56.5
57	61	11	13.7	14	17.5	59	56.5	61.5
62	66	23	28.7	37	46.2	64	61.5	66.5
67	71	27	33.7	64	79.9	69	66.5	71.5
72	76	10	12.5	74	92.4	74	71.5	76.5
77	81	5	6.3	79	98.7	79	76.5	81.5
82	86	1	1.3	80	100	84	81.5	86.5
Total		80	100.0	-	-	-	-	-

Fuente. elaboración propia.

En el ejemplo: para encontrar el valor del **decil 4 (D_4)** debe localizarse aquel valor que deje el 40 % de los datos con menor o igual magnitud a él, y el 60 % de los valores con magnitud más grande o igual a él.

$$\boldsymbol{D_r} = LIR + \left[\left(\frac{\frac{rn}{10} - f_{aa}}{f}\right)a\right] = \boldsymbol{D_4} = \mathbf{61.5} + \left[\left(\frac{\frac{(4)(80)}{10} - \mathbf{14}}{\mathbf{23}}\right)\mathbf{5}\right] = 65.41\ kg.$$

$$\boldsymbol{a} = LSR - LIR = 66.5 - 61.5 = \mathbf{5}$$

Interpretación: válida para el decil 4 (D_4): el 40 % de los estudiantes de la Escuela Secundaria Técnica 13 de Atitalaquia en el estado de Hidalgo pesa 65.41 kg o menos, mientras que el 60 % restante pesa 65.41 kg o más.

7.7.3. De dispersión o variabilidad

7.7.3.1. Rango (R)

Definición: es la diferencia entre el valor mayor y el valor menor de una serie (para profundizar en la definición ver apartado: rango de series simples).

Ejemplo: en la Escuela Secundaria Técnica 13 de Atitalaquia en el estado de Hidalgo se realizó un estudio sobre obesidad en estudiantes de nivel básico, para ello se recolectó la variable peso en kilogramos de una muestra de 80 estudiantes.

Procedimiento: encontrar, por sustracción o resta, la diferencia entre el límite superior del intervalo de clase más grande de la serie ($X_{máx}$) y el límite inferior del intervalo de clase más pequeño ($X_{mín}$), mediante la siguiente fórmula:

$$\text{Rango} = (X_{máx} - X_{mín})$$

Donde:

$X_{máx}$ = valor del límite superior del intervalo de clase máximo o mayor

$X_{mín}$ = valor del límite inferior del intervalo de clase mínimo o menor

Tabla 66. ***Estudiantes, según peso, de la Escuela Secundaria Técnica 13, Atitalaquia, Hidalgo. Enero, 2019***

Límite inferior (LI)	Límite superior (LS)	Frecuencia absoluta (FA)	Frecuencia relativa (FR)	Frecuencia absoluta acumulada (FAA)	Frecuencia relativa acumulada (FRA)	Centro de clase (CC)	Límite inferior real (LIR)	Límite superior real (LSR)
52	- 56	3	3.8	3	3.8	54	51.5	- 56.5
57	- 61	11	13.7	14	17.5	59	56.5	- 61.5
62	- 66	23	28.7	37	46.2	64	61.5	- 66.5
67	- 71	27	33.7	64	79.9	69	66.5	- 71.5
72	- 76	10	12.5	74	92.4	74	71.5	- 76.5
77	- 81	5	6.3	79	98.7	79	76.5	- 81.5
82	-**86**	1	1.3	80	100	84	81.5	- 86.5
Total		**80**	**100.0**	-	-	-	-	-

Fuente: elaboración propia.

En el ejemplo: el valor más grande del límite superior del intervalo de clase es 86 y el valor más pequeño límite inferior del intervalo de clase es 52. Al sustituir los datos en la fórmula para series agrupadas el rango es de 34 kg.

$$Rango = (X_{máx} - X_{mín}) = 86 - 52 = 34\ kg$$

Interpretación: la diferencia entre el estudiante de la Escuela Secundaria Técnica 13 de Atitalaquia en el estado de Hidalgo que más pesa y el que menos pesa es de 34 kg.

7.7.3.2. Rango intercuartílico (RIQ)

Definición: es la diferencia entre el cuartil 3 (Q_3) y el cuartil 1(Q_1) en una serie de datos. Se expresan con las letras *RIQ*.

Ejemplo: en la Escuela Secundaria Técnica 13 de Atitalaquia en el estado de Hidalgo se realizó un estudio sobre obesidad en estudiantes de nivel básico, para ello se recolectó la variable peso en kilogramos de una muestra de 80 estudiantes.

Procedimiento: obtener el **cuartil 1 (Q_1)** y el **cuartil 3 (Q_3)** por separado y encontrar la diferencia entre ambos mediante una resta. El **rango intercuartílico** para datos agrupados se obtiene utilizando la siguiente fórmula:

$$RIQ = (Q_3 - Q_1)$$

Los cuartiles para datos agrupados se obtienen utilizando la siguiente fórmula:

$$\boldsymbol{Q_r} = LIR + \left[\left(\frac{\frac{rn}{4} - f_{aa}}{f}\right)a\right]$$

Donde:
Q_r = cuartil buscado donde $\boldsymbol{r}$ puede ser 1,2 o 3
LIR = límite inferior real del intervalo, donde está la frecuencia relativa acumulada del cuartil buscado
n = número de casos
$\boldsymbol{f_{aa}}$ = frecuencia absoluta acumulada anterior al intervalo, donde cae el cuartil buscado
f = frecuencia absoluta donde cae el cuartil buscado
a = tamaño del intervalo o ancho del intervalo de clase, donde: $\boldsymbol{a} = LSR - LIR$

Tabla 67. ***Estudiantes, según peso, de la Escuela Secundaria Técnica 13, Atitalaquia, Hidalgo. Enero, 2019***

Límite inferior (LI)	Límite superior (LS)	Frecuencia absoluta (FA)	Frecuencia relativa (FR)	Frecuencia absoluta acumulada (FAA)	Frecuencia relativa acumulada (FRA)	Centro de clase (CC)	Límite inferior real (LIR)	Límite superior real (LSR)
52	56	3	3.8	3	3.8	54	51.5	56.5
57	61	11	13.7	14	17.5	59	56.5	61.5
62	66	23	28.7	37	46.2	64	61.5	66.5
67	71	27	33.7	64	79.9	69	66.5	71.5
72	76	10	12.5	74	92.4	74	71.5	76.5
77	81	5	6.3	79	98.7	79	76.5	81.5
82	86	1	1.3	80	100	84	81.5	86.5
Total		**80**	**100.0**	-	-	-	-	-

Fuente: elaboración propia.

En el ejemplo el **cuartil 1 (Q_1)** se obtiene de la siguiente manera:

$$\boldsymbol{Q_r} = LIR + \left[\left(\frac{\frac{rn}{4} - f_{aa}}{f}\right)a\right] = \boldsymbol{Q_1} = \mathbf{61.5} + \left[\left(\frac{\frac{(1)(80)}{4} - 14}{23}\right)5\right] = 62.80\ kg.$$

En el ejemplo el **cuartil 3 (Q_3)** se obtiene de la siguiente manera:

$$\boldsymbol{Q_r} = LIR + \left[\left(\frac{\frac{rn}{4} - f_{aa}}{f}\right)a\right] = \boldsymbol{Q_3} = \mathbf{66.5} + \left[\left(\frac{\frac{(3)(80)}{4} - 37}{27}\right)5\right] = 70.75\ kg.$$

Entonces el *RIQ* equivale a:

$$\boldsymbol{RIQ} = (Q_3 - Q_1) = 70.75 - 62.80 = 7.95\ kg.$$

Interpretación: la diferencia entre el 50 % central del peso de los estudiantes de la Escuela Secundaria Técnica 13 de Atitalaquia en el estado de Hidalgo es de 7.95 kg.

7.7.3.3. Desviación estándar (s)

Definición: es la raíz cuadrada del promedio de desviación de las puntuaciones cuadráticas con respecto a la media de una serie de datos. Es la raíz cuadrada de la varianza (para profundizar en la definición ver apartado: desviación estándar de series simples).

Ejemplo: en la Escuela Secundaria Técnica 13 de Atitalaquia en el estado de Hidalgo se realizó un estudio sobre obesidad en estudiantes de nivel básico, para ello se recolectó la variable peso en kilogramos de una muestra de 80 estudiantes, con los siguientes resultados.

Tabla 68. ***Estudiantes, según peso, de la Escuela Secundaria Técnica 13, Atitalaquia, Hidalgo. Enero, 2019***

Límite inferior (LI)	Límite superior (LS)	Frecuencia absoluta (FA)	Frecuencia relativa acumulada (FRA)	Centro de clase (CC)
52	56	3	3.8	54
57	61	11	17.5	59
62	66	23	46.2	64
67	71	27	79.9	69
72	76	10	92.4	74
77	81	5	98.7	79
82	86	1	100	84
Total		**80**	-	-

Fuente: elaboración propia.

Procedimiento:

Paso 1. Obtener la media ($\bar{x}$) de la serie de datos.

La media para datos agrupados se obtiene utilizando la siguiente fórmula:

$$\bar{x} = \frac{\Sigma f x}{n}$$

Donde:
$\bar{x}$ = media
Σ = sumatoria
f = frecuencia absoluta
x = punto medio o centro de clase
n = tamaño de la muestra

Paso 2. Calcular la desviación real del punto medio o centro de clase (x) con respecto a la media $(x - \bar{x})$.

Paso 3. Elevar al cuadrado cada una de las desviaciones $(x - \bar{x})^2$.

Paso 4. Multiplicar el cuadrado cada una de las desviaciones $(x - \bar{x})^2$ por la frecuencia absoluta (f) para obtener la sumatoria de las mismas $\Sigma f(x - \bar{x})^2$ y dividirla entre $n - 1$.

Paso 5. Obtener la raíz cuadrada de la $\frac{\Sigma f(x-\bar{x})^2}{n-1}$. Es decir, $s = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{n-1}}$

En el ejemplo:

Paso 1. Obtener la media ($\bar{x}$) de la serie de datos.

Tabla 69. ***Estudiantes, según peso, de la Escuela Secundaria Técnica 13, Atitalaquia, Hidalgo. Enero, 2019***

Límite inferior (LI)	Límite superior (LS)	Frecuencia absoluta (FA) *f*	Centro de clase (CC) *x*	Frecuencia absoluta por punto medio o centro de clase (*f*x)
52	56	3	54	162
57	61	11	59	649
62	66	23	64	1472
67	71	27	69	1863
72	76	10	74	740
77	81	5	79	395
82	86	1	84	84
Total		**80**	-	Σfx= **5,365**

Fuente: elaboración propia.

$$\bar{x} = \frac{\Sigma f x}{n} = \frac{5365}{80} = 67.06 \text{ kg}$$

Pasos del 2 al 5.

Tabla 70. ***Estudiantes, según peso, de la Escuela Secundaria Técnica 13, Atitalaquia, Hidalgo. Enero, 2019***

Límite inferior (LI)	Límite superior (LS)	Frecuencia absoluta (FA) f	Frecuencia relativa acumulada (FRA)	Centro de clase (CC) x	Desviación del punto medio o centro de clase con respecto a la media $(x-\bar{x})$	Desviación del punto medio o centro de clase con respecto a la media al cuadrado $(x-\bar{x})^2$	Desviación de punto medio o centro de clase con respecto a la media al cuadrado por la frecuencia absoluta $f(x-\bar{x})^2$
52	56	3	3.8	54	-13.06	170.56	511.69
57	61	11	17.5	59	-8.06	64.96	714.60
62	66	23	46.2	64	-3.06	9.36	215.36
67	71	27	79.9	69	1.94	3.76	101.62
72	76	10	92.4	74	6.94	48.16	481.64
77	81	5	98.7	79	11.94	142.56	712.82
82	86	1	100	84	16.94	286.96	286.96
Total		**80**	-	-	-	-	$\Sigma f(x-\bar{x})^2$=**3024.69**

Fuente: elaboración propia.

Sustituyendo:

$$s = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{n-1}} = \sqrt{\frac{3024.69}{80-1}} = \ 6.18 \text{ kg}$$

Interpretación 1: el peso de los estudiantes de la Escuela Secundaria Técnica 13 de Atitalaquia en el estado de Hidalgo se desvía en promedio 6.18 kg de la media de grupo.

Para una interpretación más amplia de la desviación estándar es necesario considerar las **propiedades de la curva normal**.

a) Es un **polígono de frecuencias** en forma de campana para el que están calculadas sus áreas en función de los diversos valores del eje horizontal o del eje de las X o abscisas.
b) En el eje de las X o abscisas se encuentran valores de tipo cuantitativo continuo, genéricamente denominados **puntuaciones Z**, cuyas magnitudes teóricamente pueden ir de cero, y de izquierda a derecha desde **menos infinito (-∞) a infinito (∞)**.
c) La media ($\bar{x}$ en una muestra o μ en una población) de todos los valores Z de la abscisa equivale a cero, pues la mitad son negativos y la mitad son positivos. En el sitio de la abscisa que corresponde al cero, es decir **la media, se encuentra la parte más alta de la curva**. En este sitio **también se encuentra**

la mediana de todos los valores Z de la abscisa, pues el 50 % de ellos está antes del cero y el 50 % restante se encuentra después.

d) La curva es **simétrica** alrededor de la media, hay una mitad izquierda que es reflejo de la mitad derecha. Es decir, la asimetría es cero, la mitad de la curva es exactamente igual a la otra mitad.

e) En el eje de las X o abscisas existen **segmentos unitarios de igual longitud y de tamaño 1**. Los segmentos a la izquierda de la media tienen signo negativo y los segmentos a la derecha de la media tienen signo positivo. Tales segmentos, denominados **desviaciones estándar (S)** pueden dividirse en fracciones infinitamente pequeñas y continuas.

f) La curva es **asintótica**; es decir, sus extremos teóricamente nunca tocan la abscisa. Por ello, la longitud de la abscisa podría ser infinitamente larga; sin embargo, se acostumbra graficar solo hasta la distancia de tres segmentos a la izquierda y a la derecha de la media.

g) **Toda el área bajo la curva equivale a 1 o 100 %**. Por lo anterior, el área a la izquierda de la media equivale a 0.5 o 50 % y el área a la derecha de la media equivale también a 0.5 o 50 %.

h) Es **unimodal**; es decir, presenta una sola moda

i) Es una función particular entre desviaciones con respecto a la media de una distribución y la probabilidad de que estas ocurran.

j) El área que se encuentra sobre el segmento de la abscisa, que va desde la media hasta el valor Z de -1 o -1s, equivale a 0.3413 o 34.13 %; por simetría, el área que se encuentra sobre el segmento que va desde la media hasta el valor *Z* de 1 o *1s* de la abscisa también equivale a 0.3413 o 34.13 %. En conjunto **el área de -1z a 1z o -1s a 1s equivale a 0.6826 o 68.27 % central de los datos**. El área que se encuentra sobre el segmento de la abscisa que va más allá del **valor Z de -1 equivale a 0.1587 o 15.87 %**; por simetría, el área que se encuentra sobre el segmento que va más allá (hacia infinito) del valor *Z* de 1 de la abscisa también equivale a 0.1587 o 15.87 %.

k) El área que se encuentra sobre el segmento de la abscisa que va desde la media hasta **el valor Z de -2 o -2s equivale a 0.4772 o 47.72 %**; por simetría, el área que se encuentra sobre el segmento que va desde la media hasta el valor *Z* de 2 o *2s* de la abscisa también equivale a 0.4772 o 47.72 %. En conjunto **el área de -2z a 2z o -2s a 2s equivale a 0.9544 o 95.44 % central de los datos.** El área que se encuentra sobre el segmento de la abscisa que va más allá del **valor Z de -2 equivale a 0.0228 o 2.28 %**; por simetría, el área que se encuentra sobre el segmento que va más allá (hacia infinito) del valor *Z* de 2 de la abscisa también equivale a 0.0228 o 2.28 %.

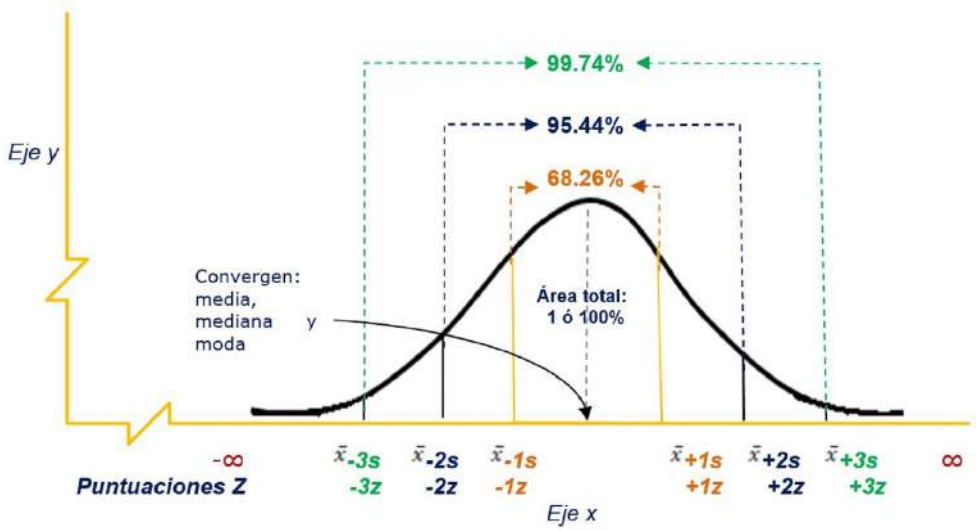
l) El área que se encuentra sobre el segmento de la abscisa que va desde la media hasta **el valor Z de -3o -3s equivale a 0.4987 o 49.87 %**; por simetría, el área que se encuentra sobre el segmento que va desde la media hasta el valor *Z* de 3 o *3s* de la abscisa también equivale a 0.4987 o 49.87 %. En conjunto **el área de -3z a 3z o -3s a 3s equivale a 0.9974 o 99.74 % central de los datos.** El área que se encuentra sobre el segmento de la abscisa que va más allá del **valor Z de -3 equivale a 0.0013 o 0.13 %**; por simetría, el área que se encuentra sobre el segmento que va más allá (hacia infinito) del valor *Z* de 3 de la abscisa también equivale a 0.0013 o 0.13 %.

m) Es **mesocúrtica**, el valor de su curtosis equivale a cero.

n) La **media, la mediana y la moda** coinciden en el mismo punto.

o) Para cualquier segmento de la abscisa, y aún para fracciones de segmento, las áreas correspondientes se encuentran calculadas en la tabla específicamente diseñada para tal efecto, la **tabla de la distribución normal**.

Figura 46. ***Curva normal (Características)***



Fuente: elaboración propia.

Interpretación 2: considerando las propiedades de la curva normal es posible plantear la siguiente pregunta: **¿Cuánto pesa el 68.26 % central de los estudiantes de la Escuela Secundaria Técnica No. 13 de Atitalaquia en el Estado de Hidalgo?**

Derivado de las propiedades de la curva normal se sabe que la distancia de una desviación estándar negativa (-1s) a una desviación estándar positiva (1s) es de 68.26 % y que el valor de la media ($\bar{x}$) se localiza justo en el centro de la curva. Entonces:

A la media (67.06 kg) se le resta una desviación estándar (6.18 kg) se obtiene el valor de 60.88 kg.

$$\bar{x} - 1s = 67.06 - 6.18 = 60.88\ kg$$

Por otra parte, a la media (67.06 kg) se le suma una desviación estándar (6.18 kg) y se obtiene el valor de 73.24 kg.

$$\bar{x} + 1s = 67.06 + 6.18 = 73.24\ kg$$

Respuesta: el 68.26 % central de los estudiantes de la Escuela Secundaria Técnica 13 de Atitalaquia en el estado de Hidalgo pesa entre 60.88 kg y 73.24 kg.

Asimismo, es importante conocer el comportamiento del resto de los datos, por lo cual se hace los siguientes cuestionamientos:

¿Cuántos kilogramos pesa el 15.87 % superior de los estudiantes?

¿Cuántos kilogramos pesa el 15.87 % inferior de los estudiantes?

Se sabe que la distancia de una desviación estándar (-1s) negativa a una desviación estándar positiva (1s) es de 68.26 %, y que el valor de la media ($\bar{x}$) y la mediana se localiza justo en el centro de la curva; por

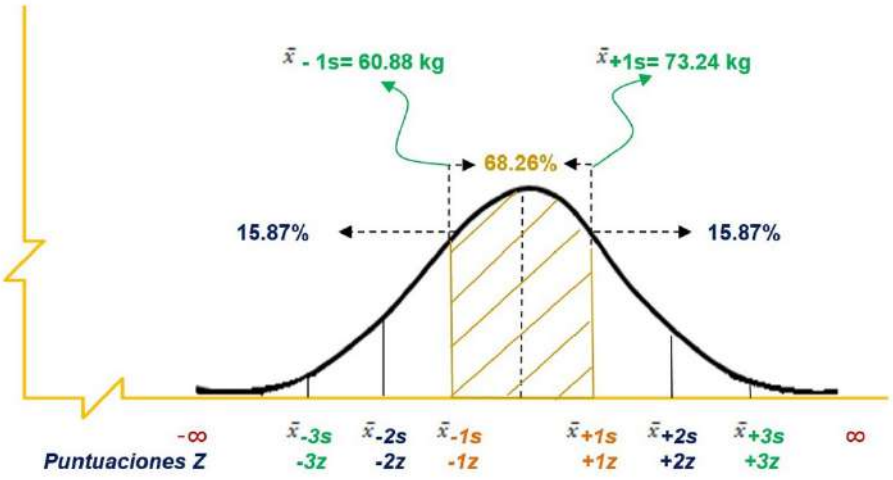
tanto, la mitad de los datos o el 50 % de los mismos se localizan por debajo de la mediana, y la otra mitad, el otro 50 %, por encima de la mediana, lo que significa que es simétrica. Entonces, se deduce que el 15.87 % de los datos se refiere a cada uno de los extremos de la curva. Es decir, si asumimos que la curva es simétrica al dividir el 68.26 % entre 2 se obtiene 34.13 % del lado positivo y del lado negativo; si solo la mitad vale 50 %, al restar a este valor el 34.13 % se obtiene 15.87 % en cada extremo.

Respuestas:

El 15.87 % superior de los estudiantes pesa 73.24 kg o más.

El 15.87 % inferior de los estudiantes pesa 60.88 kg o menos.

Figura 47. ***Distribución de datos***



Fuente: elaboración propia.

Nota importante: para ampliar el uso e interpretación de la desviación estándar no olvide consultar el apartado *6.3.3.4.1. Áreas bajo la curva normal y puntuaciones Z*.

7.7.3.4. Varianza

Definición: es una medida de dispersión que representa la variabilidad de una serie de datos respecto a su media. Mide qué tan dispersos están los datos alrededor de la media. La varianza es igual a la desviación estándar elevada al cuadrado (para profundizar en la definición ver apartado: varianza de series simples).

Ejemplo: en la Escuela Secundaria Técnica 13 de Atitalaquia en el estado de Hidalgo se realizó un estudio sobre obesidad en estudiantes de nivel básico, para ello se recolectó la variable peso en kilogramos de una muestra de 80 estudiantes, con los siguientes resultados.

Tabla 71. ***Estudiantes, según peso, de la Escuela Secundaria Técnica 13, Atitalaquia, Hidalgo. Enero, 2019***

Límite inferior (LI)	Límite superior (LS)	Frecuencia absoluta (FA)	Frecuencia relativa acumulada (FRA)	Centro de clase (CC)
52	56	3	3.8	54
57	61	11	17.5	59
62	66	23	46.2	64
67	71	27	79.9	69
72	76	10	92.4	74
77	81	5	98.7	79
82	86	1	100	84
Total		80	-	-

Fuente: elaboración propia.

Procedimiento:

Paso 1. Obtener la media ($\bar{x}$) de la serie de datos.

La media para datos agrupados se obtiene utilizando la siguiente fórmula:

$$\bar{x} = \frac{\Sigma f x}{n}$$

Donde:
$\bar{x}$= Media
Σ= Sumatoria
f= Frecuencia absoluta
x= Punto medio o centro de clase
n= Tamaño de la muestra

Paso 2. Calcular la desviación real del punto medio o centro de clase (x) con respecto a la media $(x - \bar{x})$.

Paso 3. Elevar al cuadrado cada una de las desviaciones $(x - \bar{x})^2$.

Paso 4. Multiplicar el cuadrado cada una de las desviaciones $(x - \bar{x})^2$ por la frecuencia absoluta (f) para obtener la sumatoria de las mismas $\Sigma f(x - \bar{x})^2$ y dividirla entre $n - 1$. Es decir, $S^2 = \frac{\Sigma f(x-\bar{x})^2}{n-1}$

En el ejemplo:

Paso 1. Obtener la media ($\bar{x}$) de la serie de datos.

$$\bar{x} = \frac{\Sigma f x}{n} = \frac{5{,}365}{80} = 67.06\ kg$$

Pasos del 2 al 4.

Tabla 72. ***Estudiantes, según peso, de la Escuela Secundaria Técnica 13, Atitalaquia, Hidalgo. Enero, 2019***

Límite inferior (LI)	Límite superior (LS)	Frecuencia absoluta (FA)	Frecuencia relativa acumulada (FRA)	Centro de clase (CC) x	Desviación del punto medio o centro de clase con respecto a la media $(x-\bar{x})$	Desviación del punto medio o centro de clase con respecto a la media al cuadrado $(x-\bar{x})^2$	Desviación de punto medio o centro de clase con respecto a la media al cuadrado por la frecuencia absoluta $f(x-\bar{x})^2$
52	56	3	3.8	54	-13.06	170.56	511.69
57	61	11	17.5	59	-8.06	64.96	714.60
62	66	23	46.2	64	-3.06	9.36	215.36
67	71	27	79.9	69	1.94	3.76	101.62
72	76	10	92.4	74	6.94	48.16	481.64
77	81	5	98.7	79	11.94	142.56	712.82
82	86	1	100	84	16.94	286.96	286.96
Total		80	-	-	-	-	$\Sigma f(x-\bar{x})^2$ =3024.69

Fuente: elaboración propia.

Sustituyendo:

$$S^2 = \frac{\Sigma f(x-\bar{x})^2}{n-1} = \frac{3{,}024.29}{79} = 38.28\ kg^2$$

En una primera interpretación tendríamos que *El peso de los estudiantes varia 38.28 kilogramos al cuadrado.* Es conveniente mencionar que siempre que se calcula la varianza se tendrán unidades de medida al cuadrado; para transformar dicha medida a un valor más accesible tendremos aplicar una raíz cuadrada, lo que equivale a la desviación estándar.

$$s = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{n-1}} = \sqrt{\frac{3{,}024.69}{80-1}} = \ 6.18\ kg.$$

Interpretación: el peso de los estudiantes de la Escuela Secundaria Técnica 13 de Atitalaquia en el estado de Hidalgo se desvía en promedio 6.18 kg de la media de grupo.

7.7.3.5. Coeficiente de variación (CV)

Definición: es una relación o razón estadística entre la desviación estándar y la media aritmética multiplicada por 100, es decir, se expresa en términos de porcentaje (para profundizar en la definición ver apartado: coeficiente de variación de series simples).

Ejemplo: en la Escuela Secundaria Técnica 13 de Atitalaquia en el estado de Hidalgo se realizó un estudio sobre obesidad en estudiantes de nivel básico, para ello se recolectó la variable peso en kilogramos de una muestra de 80 estudiantes, con los siguientes resultados.

Tabla 73. ***Estudiantes, según peso, de la Escuela Secundaria Técnica 13, Atitalaquia, Hidalgo. Enero, 2019***

Límite inferior (LI)	Límite superior (LS)	Frecuencia absoluta (FA)	Frecuencia relativa acumulada (FRA)	Centro de clase (CC)
52	56	3	3.8	54
57	61	11	17.5	59
62	66	23	46.2	64
67	71	27	79.9	69
72	76	10	92.4	74
77	81	5	98.7	79
82	86	1	100	84
Total		80	-	-

Fuente: elaboración propia.

Procedimiento:

Paso 1. Obtener la media de los datos.

Paso 2. Obtener la desviación estándar de la serie.

Paso 3. Aplicar la siguiente fórmula:

$$CV = \frac{s}{\bar{x}}$$

Donde:
CV= Coeficiente de variación
$\bar{x}$ = Media de los datos
s = Desviación estándar de los datos

En el ejemplo:

Paso 1. Obtener la media ($\bar{x}$) de la serie de datos.

$$\bar{x} = \frac{\Sigma fx}{n} = \frac{5365}{80} = 67.06 \text{ kg}$$

Paso 2.

$$s = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{n-1}} = \sqrt{\frac{3024.69}{80-1}} = 6.18 \text{ kg}$$

Paso 3.

$$CV = \frac{s}{\bar{x}} = \frac{6.18}{67.06} = 0.0921 \; x \; 100 = 9.21 \; \%$$

Interpretación: el peso de los estudiantes de la Escuela Secundaria Técnica 13 de Atitalaquia en el estado de Hidalgo presentan una dispersión o variación del 9.21 %.

7.7.4. Distribución o de forma

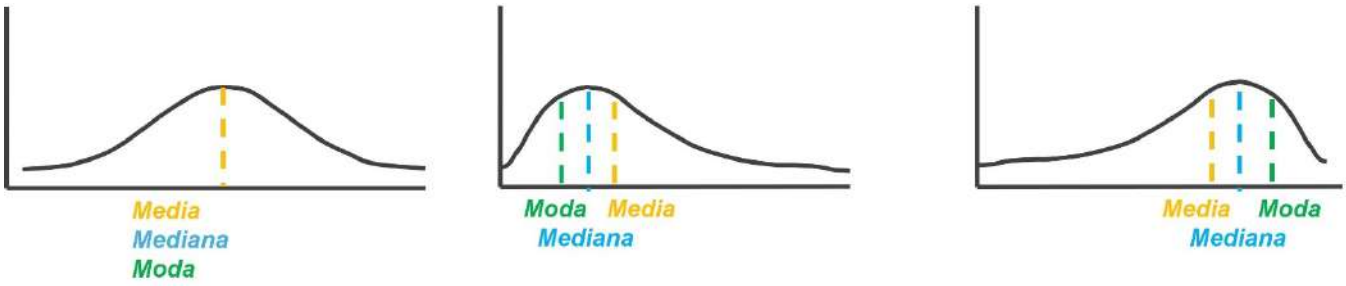
Las medidas estadísticas aplicables a variables de tipo cuantitativo clasificadas como de distribución o forma son **sesgo y curtosis**.

7.7.4.1. Sesgo o asimetría (a_3)

Definición: es la falta de simetría en una distribución. Cuando una curva está equilibrada con relación a su eje vertical, se dice que es simétrica; cuando no observa esta situación, se dice que es asimétrica (para profundizar en la definición ver apartado: sesgo o asimetría de series simples).

Si una distribución de datos tiene una asimetría de cero (0) la moda, mediana y media convergen en un solo punto; es decir, son iguales; por tanto es simétrica; entonces se habla de una **distribución normal**. Cuando hay más valores agrupados hacia la izquierda de la curva (por debajo de la media) la **asimetría o sesgo es positivo**. Cuando los valores tienden a agruparse hacia la derecha de la curva (por encima de la media) la **asimetría es negativa**.

Figura 48. ***Sesgo o asimetría en una distribución***



Simetría o asimetría cero. La media, mediana y moda convergen en un solo punto.

Asimetría o sesgo a la derecha o sesgo positivo. La media es mayor a la mediana y la moda.

Asimetría o sesgo a la izquierda o sesgo negativo. La media es menor a la mediana y la moda.

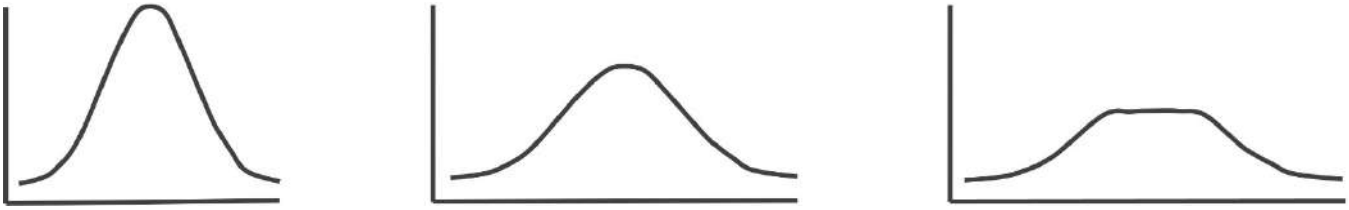
Fuente: elaboración propia.

7.7.4.2. Curtosis (a_4)

Definición: grado de concentración de valores de una variable alrededor de la zona central de una distribución (para profundizar en la definición ver apartado: curtosis de series simples).

Dependiendo del grado de elevación o aplanamiento de una distribución esta puede ser **leptocúrtica, mesocúrtica (o normal) o platicúrtica**.

Figura 49. ***Curtosis en una distribución***



Leptocúrtica: se presenta un **apuntamiento** de la curva o mayor concentración de datos alrededor de un valor central (menor dispersión).

Mesocúrtica (o normal): se presenta un **aplanamiento moderado** de la curva o concentración media de datos alrededor de un valor central.

Platicúrtica: se presenta un **aplanamiento pronunciado** de la curva o menor concentración de datos alrededor de un valor central (mayor dispersión).

Fuente: elaboración propia.

Ejemplo: en la Escuela Secundaria Técnica 13 de Atitalaquia en el estado de Hidalgo se realizó un estudio sobre obesidad en estudiantes de nivel básico, para ello se recolectó la variable peso en kilogramos de una muestra de 80 estudiantes. Con objeto de hacer posteriores comparaciones de los datos y resultados obtenidos el grupo deseaba saber si el comportamiento de los pesos de los participantes asumía o no una semejanza a la curva normal.

Procedimiento: de conformidad con el **método de momentos** una distribución de valores cuantitativos tiene semejanza a la curva normal:

a) Si su **sesgo (a_3)** tiene un valor de **-0.5 a 0.5.**

b) Si su **curtosis (a_4)** tiene un valor de **2 a 4.**

Las fórmulas para calcular el sesgo y la curtosis, a través del **método de momentos**, son los siguientes:

Tabla 74. ***Fórmulas para calcular sesgo y curtosis***

Sesgo	Curtosis
$a_3 = \frac{m_3}{(\sqrt{m_2})^3}$	$a_4 = \frac{m_4}{(m_2)^2}$

Fuente: elaboración propia.

El cálculo de cada uno de los momentos se hace con las fórmulas siguientes:

Tabla 75. ***Fórmulas para calcular momentos***

Momento 2:	*Momento 3*:	*Momento 4*
$m_2 = \frac{\Sigma f(x-\bar{x})^2}{n}$	$m_3 = \frac{\Sigma f(x-\bar{x})^3}{n}$	$m_4 = \frac{\Sigma f(x-\bar{x})^4}{n}$

Fuente: elaboración propia.

Para obtener el promedio se aplica la siguiente fórmula:

$$\bar{x} = \frac{\Sigma f x}{n}$$

Sustituyendo:

$$\bar{x} = \frac{\Sigma f x}{n} = \frac{5365}{80} = 67.06 \text{ kg}$$

Tabla 76. ***Estudiantes, según peso, de la Escuela Secundaria Técnica 13, Atitalaquia, Hidalgo. Enero, 2019***

Límite inferior (LI)	Límite superior (LS)	Frecuencia absoluta (FA)	Centro de clase (CC)	Desviación del punto medio o centro de clase con respecto a la media $(x-\bar{x})$	Desviación del punto medio o centro de clase con respecto a la media al cuadrado $(x-\bar{x})^2$	Desviación de punto medio o centro de clase con respecto a la media al cuadrado por la frecuencia absoluta $f(x-\bar{x})^2$	Desviación de punto medio o centro de clase con respecto a la media elevada a la tercera potencia por la frecuencia absoluta $f(x-\bar{x})^3$	Desviación de punto medio o centro de clase con respecto a la media elevada a la cuarta potencia por la frecuencia absoluta $f(x-\bar{x})^4$
52	56	3	54	-13.06	170.56	511.69	-6682.68	87 275.82
57	61	11	59	-8.06	64.96	714.60	-5759.67	46 422.96
62	66	23	64	-3.06	9.36	215.36	-659.01	2016.57
67	71	27	69	1.94	3.76	101.62	197.14	382.45
72	76	10	74	6.94	48.16	481.64	3342.55	23 197.32
77	81	5	79	11.94	142.56	712.82	8511.05	101 621.90
82	86	1	84	16.94	286.96	286.96	4861.16	82 348.11
Total		**80**	**-**	**-**	**-**	$\Sigma f(x-\bar{x})^2 =$ **3024.69**	$\Sigma f(x-\bar{x})^3 =$ **3810.54**	$\Sigma f(x-\bar{x})^4 =$ **343 265.14**

Fuente: elaboración propia.

Al sustituir en las fórmulas para el cálculo de momentos se tiene:

Tabla 77. ***Cálculo de momentos***

Momento 2	Momento 3	Momento 4
$m_2 = \frac{\Sigma f(x-\bar{x})^2}{n}$	$m_3 = \frac{\Sigma f(x-\bar{x})^3}{n}$	$m_4 = \frac{\Sigma f(x-\bar{x})^4}{n}$
$m_2 = \frac{3024.69}{80} = 37.80$	$m_3 = \frac{3810.54}{80} = 47.63$	$m_4 = \frac{343\ 265.14}{80} = 4290.81$

Fuente: elaboración propia.

Finalmente, usando los valores calculados para los momentos y sustituyendo para las fórmulas de sesgo y curtosis se tiene:

Tabla 78. ***Cálculo de sesgo y curtosis***

Sesgo	Curtosis
$a_3 = \frac{m_3}{(\sqrt{m_2})^3}$	$a_4 = \frac{m_4}{(m_2)^2}$
$a_3 = \frac{47.63}{\left(\sqrt{37.80}\right)^3} = 0.204$	$a_4 = \frac{4290.81}{(37.80)^2} = 3.00$

Fuente: elaboración propia.

Interpretación: considerando que los valores obtenidos de sesgo y la curtosis se encuentran dentro de los intervalos señalados por el método de momentos se concluye que:

La distribución de los pesos de los estudiantes de la Escuela Secundaria Técnica 13 de Atitalaquia en el estado de Hidalgo se asemeja en asimetría y en grado de apuntamiento o aplanamiento a una distribución normal.

Si bien para considerar que una distribución es normal los valores del sesgo y la curtosis debieran ser de conformidad con el método de momentos cero (0) y tres (3) respectivamente, se puede decir que no son normales, pero se asemejan o parecen a una distribución normal.

Capítulo 8
Población, muestra y muestreo

Imagen de CLM 2019.

8.1. Población, muestra y muestreo

Examinar todos los elementos de una determinada población puede significar el empleo de demasiados recursos y de un tiempo desmedido, por lo que se recurre al análisis de una pequeña parte, es decir, a obtener un tamaño de muestra, realizar un muestreo y estudiar datos con el objetivo de establecer generalizaciones con respecto a un grupo total.

La parte del grupo, fracción o subconjunto de elementos de la población que se examina recibe el nombre de **muestra**, y el grupo total a partir del cual se seleccionó la muestra se conoce como **población** o universo. Al conjunto de procedimientos aplicados para extraer la muestra recibe el nombre de **muestreo**.

Tabla 79. ***Conceptos básicos***

Población	Muestreo	Muestra
Es la totalidad de un conjunto de elementos, seres u objetos que se pretende investigar.	Conjunto de procedimientos que se ejecutan para seleccionar una muestra.	Es un subconjunto de la totalidad de un conjunto de elementos, seres u objetos que se pretende investigar. Es una fracción o subgrupo de una población.

Fuente: elaboración propia.

Bologna (2013) señala al respecto que el **muestreo**

> es un conjunto de procedimientos mediante los cuales se selecciona de un universo determinado, llamado **población**, un subconjunto que recibe el nombre de **muestra**, con el objetivo de llegar al conocimiento de determinadas características de los elementos de la población a través de la observación y generalización de esas características presentes en los elementos de la muestra. (p. 269)

Velasco et al. (2002) establecen por su parte los siguientes conceptos:

> a) **Población**: es todo conjunto de objetos, situaciones o sujetos con un rasgo común; es un conjunto global de casos que satisface una serie predeterminada de criterios.
>
> b) **Elementos o unidades muestrales**: es la unidad básica alrededor de la cual se recaba la información; es el elemento que da origen al valor de las variables (un expediente, una radiografía, un paciente, un animal de laboratorio, etc.).
>
> c) **Muestra**: es el subconjunto de la población integrado por las unidades muestrales seleccionadas.
>
> d) **Muestreo**: es la selección de un número de unidades de estudio a partir de una población definida (p. 15).

Hernández et al. (2014) afirman que, al iniciar el diseño de una investigación, es inevitable abordar el tema de la selección muestral, lo cual conlleva dos decisiones fundamentales: en primer lugar, la definición de la **unidad de análisis**, que puede referirse a personas, instituciones, objetos, grupos, situaciones o

acontecimientos, es decir, aquello sobre lo que se obtendrán los datos; y en segundo lugar, la **delimitación de la población**, entendida como el establecimiento de los criterios específicos que debe reunir la unidad de análisis en términos de tres dimensiones clave: el contenido (qué se va a medir), el espacio (dónde se va a medir) y el tiempo (cuándo se realizará la medición).

Elorza (2008), por su parte, hace referencia a población objetivo y marco muestral; entiende por **población objetivo** a la conformada por "los elementos que cumplan con determinadas características en un tiempo y espacio" (p. 183). Por otra parte, "el listado que contiene todos los elementos que integran la población objetivo" se le conoce como **marco de muestreo** (p. 184).

Al obtener una parte o fracción de una población, el problema principal consiste en asegurar que el subconjunto sea representativo de la misma, de tal manera que permita generalizar los resultados obtenidos, de ello depende la forma en que sea extraída la muestra; es decir, del muestreo.

8.1.1. Cualidades de una buena muestra

Ander-Egg (1995) señala que una muestra tiene validez técnico-estadística si cumple con los siguientes requisitos:

1) Ser representativa o reflejo general del conjunto o universo estudiado, reproduciendo lo más exactamente las características del mismo.
2) Que todos los elementos o unidades el conjunto o universo tenga la misma posibilidad de ser escogidos.
3) Que su tamaño sea estadísticamente proporcionado a la magnitud del universo.
4) Que el error muestral se mantenga dentro de los límites adoptados como permitidos.
5) Que sea posible calcular los errores de muestreo y el intervalo de confianza en que se mueva cada una de las estimaciones. (p. 359)

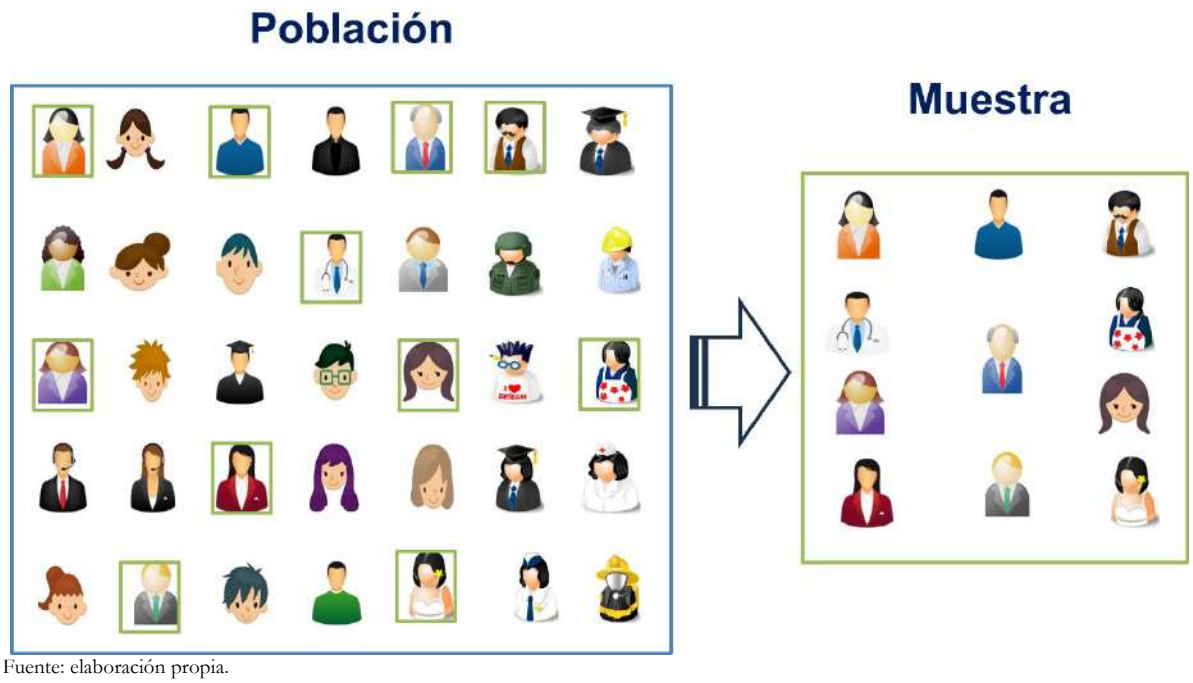
Velasco et al. (2002) señalan que para efectuar un muestreo se tiene que responder a tres preguntas:

1. ¿Cuál es la población en estudio? Debe ser representativa. Implica tener todas las características importantes de la población de la que se tomó la muestra, en proporciones similares.

2. ¿Cuántas personas se requieren en la muestra? Debe ser adecuada. Se refiere a su tamaño, se calcula con diversas fórmulas establecidas de acuerdo con el tipo de estudio a realizar.

3. ¿Cómo seleccionar la muestra? Seleccionar un método de muestreo: probabilístico o no probabilístico. (p. 16)

Figura 50. ***Relación población-muestreo-muestra***



Fuente: elaboración propia.

En este sentido, el esquema de relación población-muestreo-muestra recupera la esencia del concepto de muestra, ya que no se trata solo de obtener un subconjunto de la población, sino de que tal fracción sea la suma de diversos elementos contenidos en el total de aspectos o características de la población. Ritchey (2004) establece que una muestra es representativa "es aquella en la que todos los segmentos de la población están incluidos en la muestra, en sus proporciones correctas respecto a población" (p. 41).

8.2. Tamaño de muestra

Cuando un investigador decide analizar un fenómeno social generalmente se encuentra ante la disyuntiva de determinar cuántos sujetos deben conformar su muestra, para ello hay una extensa gama de fórmulas y consideraciones que tienen como referencia distintos parámetros. En este apartado se abordará un esquema accesible y generalmente utilizado para obtener el subconjunto de la población llamado muestra particularmente en las ciencias sociales; es decir, a través de proporciones.

Para calcular el tamaño de la muestra es necesario considerar los siguientes factores:

- **Error máximo admisible (*d*) o error de precisión** de los resultados. Es la máxima diferencia que se puede tolerar entre el valor de la variable obtenido en la muestra y el verdadero valor de esta en el universo o la población.

El valor de (*d*) lo establece el investigador al definir qué tan preciso quiere ser en los resultados de su estudio; es decir, qué tan exacto quiere conocer el valor de una variable. Por ejemplo, si se tiene por objetivo determinar el porcentaje de trabajadores sociales egresados de la ENTS-UNAM que pertenecen a una asociación gremial y se establece un **error de precisión** de 5 %; entonces, si en la muestra resulta

que el porcentaje de trabajadores sociales que pertenecen a una asociación gremial es de 20 %, se está aceptando realmente que el verdadero valor esté entre 20 % ± 5 %, es decir, entre 15 % y 25 %. El error máximo admisible está inversamente relacionado con el tamaño de la muestra (a mayor error, menor tamaño de muestra, a menor error mayor tamaño de la muestra).

- **Coeficiente de confianza de la estimación o nivel de confianza**: es la medida probabilística de que el intervalo fijado con *d* contenga el valor poblacional; es decir, determina la probabilidad de que los resultados sean ciertos dentro de determinados límites porcentuales fijados por el error de precisión.

Supongamos que en el ejemplo de los trabajadores sociales egresados de la ENTS-UNAM que pertenecen a una asociación gremial se considera un nivel de confianza de 95 %, entonces se puede afirmar con un 95 % de certeza y 5 % de probabilidad de equivocarnos que el porcentaje de trabajadores sociales que pertenecen a una asociación gremial está entre 15 % y 25 %. El tamaño de la muestra se relaciona directamente con nivel de confianza (a mayor nivel confianza mayor tamaño de muestra, a menor nivel confianza menor tamaño de muestra).

En Ciencias Sociales generalmente se acepta un nivel de confianza de 0.95 o mayor y lo define de manera directa el investigador.

Para localizar cualquier valor del nivel de confianza expresado en puntuación Z se recurre a la *Tabla de la distribución normal* (ver Anexo 1), si se recuerda las propiedades de la curva normal está es simétrica (existe la misma área en extremo derecho que en el extremo izquierdo; es decir, la mitad es espejo de la otra mitad) y el punto de referencia es la media que se encuentra justo en el centro de la distribución. De tal manera que para tener un 95 % o 0.95 de confianza, la mitad de la curva partiendo de la media tiene un valor de 47.50 % o 0.4750 y la otra mitad 47.50 % o 0.4750, ambos suman 95 % o 0.95; si se localiza el valor de 0.4750 en la *Tabla de la distribución normal* se encontrará en la *Columna (2) Distancia de Z a la media* lo cual implica una puntuación Z de 1.96 (*Columna 1).*

Tabla 80. ***Áreas bajo la curva normal (segmento)***

(1) Puntuación Z	(2) Distancia de Z ala media	(3) Área de la parte mayor	(4) Área de la parte menor
…	…	…	…
1.96	**0.4750**	0.9750	0.0250
…	…	…	…

Fuente: elaboración propia.

Para un 99 % o 0.99 de confianza, la mitad de la curva partiendo de la media tiene un valor de 49.50 % o 0.4950 y la otra mitad 49.50 % o 0.4950, ambos suman 99 % o 0.99; si se localiza el valor de 0.4950 en la *Tabla de la distribución normal* se encontrará en la *Columna (2) Distancia de Z a la media* que hay dos valores cercanos 2.57 con 0.4949 y 2.58 con 0.4951, para determinar el valor exacto se debe promediar las distancias de Z a la media (0.4949+0.4951)/2 = 0.4950 y por tanto las puntuaciones Z(2.57+2.58)/2=2.575.

Tabla 81. *Áreas bajo la curva normal (segmento)*

(1) Puntuación Z	(2) Distancia de Z ala media	(3) Área de la parte mayor	(4) Área de la parte menor
…	…	…	…
2.57	**0.4949**	0.9949	0.0051
2.58	**0.4951**	0.9951	0.0049
…	…	…	…

Fuente: elaboración propia.

- **Homogeneidad de la población o variabilidad**: es la presencia (éxito) y ausencia (fracaso) de una variable dentro de una población expresada en porcentaje o proporción, está definida por p = probabilidad de éxito y q = probabilidad de fracaso. Cuando no se dispone de parámetros basados en estudios previos se recomienda aceptar la máxima variabilidad, es decir, 50 % o 0.5 para p y 50 % o 0.5 para q.

Población infinita o desconocida

$$n = \frac{Z_a^{\,2} \times p \times q}{d^2}$$

Donde:
$\boldsymbol{n}$ = tamaño de la muestra.
$\boldsymbol{Z_\alpha}$ = nivel de confianza expresado en puntuación Z definido mediante la distribución normal o de Gauss. (Ver *Tabla de la distribución normal.* Anexo 1). El nivel de confianza generalmente aceptado es de 95 % que le corresponde una puntuación Z de 1.96. Para un nivel de seguridad del 99 % la puntuación Z es igual a 2.575.
$\boldsymbol{p}$ = variabilidad positiva.
$\boldsymbol{q}$ = variabilidad negativa.
$\boldsymbol{d}$ = precisión o error que se prevé cometer.

Población finita o conocida

$$n = \frac{N \times Z_a^{\,2} \times p \times q}{d^2 \times (N-1) + Z_a^{\,2} \times p \times q}$$

Donde:
$\boldsymbol{n}$ = tamaño de la muestra.
$\boldsymbol{N}$ = tamaño de la población.
$\boldsymbol{Z_\alpha}$ = nivel de confianza expresado en puntuación Z definido mediante la distribución de Gauss (Ver *Tabla de la distribución normal.* Anexo 1). El nivel de confianza generalmente aceptado es de 95 % que le corresponde una puntuación Z de 1.96. Para un nivel de seguridad del 99 % la puntuación Z es igual a 2.575.
$\boldsymbol{p}$ = variabilidad positiva.
$\boldsymbol{q}$ = variabilidad negativa.
$\boldsymbol{d}$ = precisión o error que se prevé cometer.

8.2.1. Tamaño de la muestra para una población infinita o desconocida

Un grupo interdisciplinario quiere hacer un estudio sobre violencia familiar hacia la mujer en la colonia Antonia M. M. Establece como población a todas aquellas mujeres mayores de 18 años, solteras, casadas, en unión libre y viudas que pertenezcan a dicha colonia. ¿Cuál es el número de mujeres *n* que se tienen que entrevistar, si se asume un nivel de confianza de 95 % o 0.95 con un error de precisión de 5 % o 0.05? No se tiene antecedentes sobre estudios realizados por lo que se asume la máxima variabilidad; es decir, *p* = 50 % o 0.5 y *q* = 50 % o 0.5.

Fórmula:

$$n = \frac{Z_a^{\,2} \times p \times q}{d^2}$$

Identificando valores:

***n* =** tamaño de la muestra: valor a calcular.

Zα = nivel de confianza: 95 % o 0.95. Valor correspondiente en la distribución normal, Zα= 1.96.

***p* =** variabilidad positiva: 50 % o 0.5.

***q* =** variabilidad negativa: 50 % o 0.5.

***d* =** precisión o error que se prevé cometer: 5 % o 0.05

Sustituyendo:

$$n = \frac{(1.96)^2\ (50)\ (50)}{(5)^2} = 384$$

¿Cuál es el número de mujeres *n* que se tienen que entrevistar, si se asume un nivel de confianza de 95 % o 0.95 con un error de muestreo de 5 % 0 0.05?

Respuesta: **384 mujeres**

Conclusión: si se encuesta a 384 mujeres, el 95 % de las veces la variable a medir estará en el intervalo ± 5 % respecto al dato que se observe en la encuesta. Es decir, si el 80 % de las encuestadas contestó que sufre violencia familiar, dado el error de precisión asumido (± 5 %) entonces entre el 75 % y el 85 % sufre violencia familiar y es válido para el 95 % de la población.

8.2.2. Tamaño de la muestra para la población finita o conocida

Un grupo interdisciplinario quiere hacer un estudio sobre violencia familiar hacia la mujer en la colonia Antonia M. M. Establece como población a todas aquellas mujeres mayores de 18 años, solteras, casadas, en unión libre y viudas que pertenezcan a dicha colonia. El número total de mujeres que cumplen tales características es de *N*= 950. ¿Cuál es el número de mujeres *n* que se tienen que entrevistar, si se asume un nivel de confianza de 95 % o 0.95 con un error de precisión de 5 % o 0.05? No se tienen antecedentes sobre estudios realizados, por lo que se asume la máxima variabilidad, es decir, *p* = 50 % o 0.5 y *q* = 50 % o 0.5.

Fórmula:

$$n = \frac{N \times Z_a^{\,2} \times p \times q}{d^2 \times (N-1) + Z_a^{\,2} \times p \times q}$$

Identificando valores:

***n* =** tamaño de la muestra: valor a calcular.

***N* =** tamaño de la población.

Zα = nivel de confianza: 95 % o 0.95. Valor correspondiente en la distribución normal, Zα = 1.96.

***p* =** variabilidad positiva: 50 % o 0.5.

***q* =** variabilidad negativa: 50 % o 0.5.

***d* =** precisión o error que se prevé cometer: 5 % o 0.05

Sustituyendo

$$n = \frac{(950)\ (1.96)^2\ (50)\ (50)}{(5)^2\ (949) + (1.96)^2\ (50)\ (50)} = 274$$

¿Cuál es el número de mujeres *n* que se tienen que entrevistar, si se asume un nivel de confianza de 95 % o 0.95 con un error de muestreo de 5 % 0 0.05?

Respuesta: **274 mujeres**

Si encuestas a 274 mujeres, el 95 % de las veces la variable a medir estará en el intervalo ± 5 % respecto al dato que se observe en la encuesta. Es decir, si el 80 % de las encuestadas contestó que sufre violencia familiar, dado el error de precisión asumido (± 5 %) entonces entre el 75 % a 85 % sufre violencia familiar y es válido para el 95 % de la población.

8.3. Tipos de muestreo

Una vez que se obtiene el tamaño de la muestra es preciso determinar cómo seleccionar los sujetos o elementos que deberán integrar ese subgrupo de la población; es decir, el tipo de muestreo a aplicar o el conjunto de procedimientos para extraer la muestra.

Considerando la estructura y los procedimientos de selección de sujetos, los muestreos se clasifican en **probabilísticos y no probabilísticos**.

Bologna (2013) señala que en los **muestreos probabilísticos**

> las muestras obtenidas por estos procedimientos permiten generalizar los resultados obtenidos en ellas a toda la población de referencia. El requisito para que una muestra sea probabilística es que sus elementos hayan sido elegidos al azar (aleatoriamente), sin la participación voluntaria que decida a quién incluir y a quién excluir de la muestra. (p. 277)

En tanto que en los **muestreos no probabilísticos** las muestras "no cumplen el requisito de aleatoriedad en la selección de los elementos que la componen. Los resultados no se pueden generalizar de manera probabilística más allá de los casos observados" (p. 277).

Tabla 82. ***Tipos de muestreo***

Muestreo probabilístico	Aleatorio simple Aleatorio sistemático Aleatorio estratificado Aleatorio estratificado y por racimos
Muestreo no probabilístico	De juicio o criterio De diseño de bola de nieve Accidental o por accidente Por conveniencia Por cuota Por sujetos-tipo Por sujetos voluntarios Por expertos

Fuente: elaboración propia.

8.3.1. Muestreo probabilístico

En este tipo de muestreo cada unidad muestral de la población en estudio tendrá la misma probabilidad de ser incluida en la investigación. Permite generalizar (con determinado margen de error) en la población total los resultados obtenidos. Selltiz et al. (1980) establecen que "la característica esencial del ***muestreo de probabilidad*** es que puede especificarse para cada elemento de la población la probabilidad de que irá incluido en la muestra" (p. 687).

Velasco et al. (2002) establecen que un ***muestreo es probabilístico*** cuando el método de selección de la muestra permite que todos los elementos de la población tengan la misma probabilidad de ser seleccionados en la muestra. Utiliza procedimientos de selección aleatoria para asegurar que cada unidad de la muestra se seleccione por probabilidad (es factible si se conoce el marco muestral, es decir, se cuenta con un listado completo de todas las unidades que componen la población) (pp. 17-18).

Una de las ventajas que presenta este tipo de muestreo, señala Elorza (2008), es que cuenta con modelos para calcular el tamaño de la muestra con una confiabilidad y un grado de error. Establece que en el muestreo probabilístico es viable calcular parámetros insesgados referentes al universo en estudio a partir de poca información sobre el mismo.

Selltiz et al. (1980) y Pick y López (1998) consideran que este tipo de muestreo es el más adecuado e incluso la única modalidad que hace posibles planes de muestreo representativo, puesto que reduce al máximo los prejuicios de selección que el investigador pueda tener.

Entre las principales desventajas que presenta este tipo de muestreo se encuentra su alto costo y la exigencia de mayor tiempo; a algunos investigadores –dice Elorza (2008)– les resulta complejo, ya que no cuentan con conocimientos de estadística inferencial.

Muestreo probabilístico

Es el procedimiento en el que todos los elementos de una población tienen la misma probabilidad de ser elegidos.

8.3.1.1. Aleatorio simple

Bologna (2013) señala que un **muestreo aleatorio simple** "es una técnica que asigna igual probabilidad de pertenecer a la muestra a todos los individuos de la población" (p. 277). Para su realización se requiere contar con una lista de los elementos de la población. Este listado es lo que se denomina el marco de la muestra. Dicho marco debe cumplir con los requisitos de exhaustividad (es decir, que sea completo) y debe cuidarse que no existan duplicaciones.

Pagano (1999) señala que una **muestra aleatoria**

> es elegida de una población mediante un proceso que asegura: (1) cada muestra aleatoria de un tamaño dado tenga una misma probabilidad de ser elegida; (2) todos los miembros de la población tengan la misma probabilidad de quedar en la muestra". (p. 156).

Por parte, Ritchey (2004) señala que una **muestra aleatoria** "es aquella en la cual cada persona (u objeto) en la población tiene la misma oportunidad de ser seleccionada (o) para la muestra" (p. 42).

Gonick y Smith (1999) establecen al respecto que un **muestreo aleatorio simple** presenta dos propiedades que lo convierten en un estándar frente al que se miden todos los otros métodos de muestreo:

> **Representativa**: cada unidad tiene las mismas posibilidades de ser escogida. Dicho en términos estadísticos, es representativa dado que presenta ausencia de sesgo.
> **Independencia**: la selección de una unidad no influye en la selección de otras unidades. (p. 93)

Un método para obtener este tipo de muestra es colocar en un recipiente todas las unidades de análisis, ya sean fichas o papeles con nombres o números que correspondan a cada una de ellas, y extraer aleatoriamente hasta completar el tamaño de la muestra. Para este tipo de muestreo es necesario conocer la población de estudio.

Muestreo aleatorio simple

Es el procedimiento en el que un tamaño de muestra n es seleccionado de una población de tamaño N, de tal manera que cada muestra posible de tamaño n tiene la misma posibilidad de ser seleccionada.

Ejemplo: supóngase que se tiene una población de 45 alumnos y se desea obtener una muestra de $n = 10$ para conocer su opinión acerca de las enfermedades de transmisión sexual.

Primer paso. Elaborar una lista de los alumnos enumerándolos del 1 al 45.

Segundo paso. De modo arbitrario se selecciona una parte de la *Tabla de números aleatorios*, una columna y un renglón.

Supóngase que en la *Tabla de números aleatorios* se seleccionó la columna 3 y el renglón 2, lo que presenta que se iniciará a seleccionar desde el número 27 (se puede comenzar a buscar valores que integraran la muestra, ya sea, hacia abajo, hacia arriba, al lado izquierdo o derecho e incluso de manera diagonal).

En este caso, se consideran los valores que se encuentran hacia abajo; es decir, 27, 47, 04, 36, 27, 18, 24, 87, 41, 37, 41, 09, 65, 68, 62, 55, 93, 06, 37, 75, y 08.

De esta serie de números se omiten los valores mayores a la población, es decir, mayores de 45; así como los valores repetidos.

Tercer paso. Obtener las unidades muestrales resultantes del paso anterior, es decir, hay que entrevistar a los alumnos que ocupan los lugares: 27, 04, 36, 18, 24, 41, 37, 09, 06 y 08, que conforman el total de la muestra $n = 10$.

Tabla 83. ***Tabla de números aleatorios (segmento)***

Renglón	Columnas				
	1	**2**	**3**	**4**	**5**
1	53	95	67	80	79
2	62	12	27	41	05
3	90	16	~~47~~	72	20
4	10	59	04	76	80
5	32	17	36	64	08
6	54	71	~~27~~	89	41
7	10	60	18	77	34
8	42	20	24	36	78
9	73	55	~~87~~	48	49
10	21	56	41	23	58
11	09	60	37	99	06
12	63	26	~~41~~	08	21
13	98	72	09	45	69
14	87	89	~~65~~	22	98
15	05	91	~~68~~	44	67
16	75	93	~~62~~	49	95
17	76	15	~~55~~	38	29
18	26	76	~~93~~	84	08
19	08	35	06	83	76
20	59	73	~~37~~	06	26
21	87	94	~~75~~	45	72
22	05	74	08	91	37
23	49	82	39	40	51
24	02	25	92	97	41
25	59	41	49	100	13

Fuente: elaboración propia.

8.3.1.2. Aleatorio sistemático

Kelmansky (2009) señala que el **muestreo sistemático** "comienza con una unidad elegida al azar, a partir de allí continúa cada *k* unidades. Si *n* es el tamaño muestral y *N* es el tamaño de la población, entonces *k* es aproximadamente *N*/*n*" (p. 35).

Velasco et al. (2002) señalan que en un **muestreo sistemático** "todos los individuos se seleccionan a intervalos regulares, cada *K* elemento (el tercero, quinto, décimo). Se selecciona dividiendo el total de población entre el número de elementos deseados, lo que nos dará el intervalo de cada cuántos se eligen" (p. 20); por ejemplo, en una población de 300 elementos y un tamaño de muestra requerido de 60, 300/60 = 5, se escogerá cada quinto elemento.

El **muestreo aleatorio sistemático** consiste en la selección a intervalos de los individuos. Una vez que se tiene el tamaño de la población *(N)* y el tamaño de la muestra *(n)* el siguiente paso es determinar, mediante una división de los dos valores anteriores, el número de selección sistemática (*késimo, k*), es decir, el tamaño del salto o paso a considerar para obtener el tamaño de muestra que se desea.

Muestreo aleatorio sistemático

Es el procedimiento de selección de elementos a través de un *késimo*, después de seleccionar aleatoriamente un punto de inicio.

Para la obtención del *késimo* (k) en el muestreo sistemático se aplica la siguiente fórmula:

$$k = \frac{Tamaño\ de\ la\ Población}{Tamaño\ de\ la\ muestra} = \frac{N}{n}$$

Una vez que se determina el valor de *k* (tamaño del paso o tamaño del salto), en un listado de casos el investigador determinará un punto de inicio a partir del cual se elegirán los elementos de la muestra al dar saltos de tantos datos como lo establezca el *késimo*.

Ejemplo: en la Escuela Nacional de Trabajo Social se quiere hacer un estudio a los alumnos de tercer semestre respecto a sus necesidades de formación extracurricular. Se sabe que se tiene una población *(N)* de 720 y al calcular la muestra *(n)* con 95% de confianza, 5 de margen error y 50 de variabilidad positiva resultó igual a 251. De tal manera que al calcular k se tiene:

$$k = \frac{Tamaño\ de\ la\ Población}{Tamaño\ de\ la\ muestra} = \frac{N}{n} = \frac{720}{251} = 2.86 = 3$$

De la lista numerada de los 720 alumnos se elige un punto inicial y cuenta **tres** (dado que es el tamaño del *késimo* o salto) y se considera al estudiante que ocupa ese lugar para aplicarle un instrumento de

investigación; a partir de ese elemento seleccionado se cuenta nuevamente **tres** y se tendrá al segundo elemento seleccionado, se contará tres así sucesivamente hasta obtener los 251 alumnos que se consideran para la muestra.

8.3.1.3. Aleatorio estratificado

Velasco et al. (2002) señalan que en el **muestro estratificado**

> se divide primero a la población en estratos pertinentes (subgrupos) y luego de cada estrato se selecciona la muestra aleatoria, es decir, las extracciones de la muestra deben hacerse independientemente en los diferentes estratos (es una muestra aleatoria simple en cada estrato). Es posible solo cuando se conoce la proporción de la población en estudio que pertenece a cada grupo de interés. Las subpoblaciones no deben traspalarse (deben ser mutuamente excluyentes) y en su conjunto corresponden a toda la población ($n_1 + n_2 + n_3 + n_i = N$). (p. 20)

Arias (2012) señala que un **muestreo estratificado** "consiste en dividir la población en subconjuntos cuyos elementos posean características comunes, es decir, estratos homogéneos en su interior. Posteriormente se hace la escogencia al azar en cada estrato" (p. 84).

Cuando se conocen las propiedades de una población y el objetivo es mejorar la representatividad de una muestra es posible aplicar un muestreo estratificado, es decir, generar estratos o categorías que se presenten en la población y que sean relevantes para los objetivos del estudio; el número de estratos no debe ser muy grande y si se necesitan estimaciones de las subdivisiones, entonces se requiere un mayor número de estratos.

Muestreo aleatorio estratificado

Es el procedimiento de selección de unidades de muestra a partir de dividir en subgrupos o estratos a la población de estudio, para extraer proporcionalmente de cada uno de ellos los elementos que integrarán la muestra.

Ejemplo: un grupo interdisciplinario quiere hacer un estudio sobre violencia familiar hacia la mujer en la colonia Antonia M. M. Se establece como población a todas aquellas mujeres mayores de 18 años, solteras, casadas, en unión libre y viudas que pertenezcan a dicha colonia. El número total de mujeres que cumplen tales características es de $N = 950$. ¿Cuál es el número de mujeres n que se tienen que entrevistar si se asume un nivel de confianza de 95 % o 0.95 con un error de precisión de 5 % o 0.05? No se tienen antecedentes sobre estudios realizados, por lo que se asume la máxima variabilidad; es decir, $p = 50$ % o 0.5 y $q = 50$ % o 0.5.

Primer paso. Obtener el tamaño de muestra, como se conoce el tamaño de la población *(N)* la fórmula a aplicar es el siguiente:

$$n = \frac{N \times Z_a^{\,2} \times p \times q}{d^2 \times (N-1) + Z_a^{\,2} \times p \times q}$$

Identificando valores:

***n* =** tamaño de la muestra, valor a calcular.
***N* =** tamaño de la población.
***Za* =** nivel de confianza: 95 % o 0.95, correspondiente en la distribución normal, *Za* = 1.96.
***p* =** variabilidad positiva: 50 % o 0 0.5.
***q* =** variabilidad negativa: 50 % o 0 0.5.
***d* =** precisión o error que se prevé cometer: 5 % o 0.05

Sustituyendo:

$$n = \frac{(950)\,(1.96)^2\,(50)\,(50)}{(5)^2\,(949) + (1.96)^2\,(50)\,(50)} = 274$$

¿Cuál es el número de mujeres *n* que se tienen que entrevistar si se asume un nivel de confianza de 95 % o 0.95 con un error de muestreo de 5 % 0 0.05?

Respuesta: **274 mujeres**.

Una vez que se determina el tamaño de muestra *n* = 274, el grupo interdisciplinario decide aplicar un muestreo estratificado, partiendo de la siguiente tabla de datos previamente recabada por ellos mismos:

Tabla 84. ***Tabla de referencia de datos***

No. de estrato	Nombre del estrato, según estado civil de las mujeres	Población total por estrato	Tamaño de muestra por estrato
1	Solteras	420	?
2	Casadas	242	?
3	Unión libre	183	?
4	Viudas	105	?
		***N* = 950**	***n* = 274**

Fuente: elaboración propia.

Segundo paso. Obtener la **proporción (*p*)** a extraer de cada estrato definido por el tamaño de la muestra (*n*) entre el total de la población (*N*), mediante la siguiente fórmula:

$$p = \frac{Tamaño\ de\ la\ muestra}{Tamaño\ de\ la\ población} = \frac{n}{N} = \frac{274}{950} = 0.288$$

Tercer paso. Obtener el **tamaño de muestra de cada estrato**, multiplicando la proporción obtenida por el total de la población de cada estrato, mediante la siguiente fórmula:

$$Tamaño\ de\ la\ muestra\ por\ estrato = \ Población\ de\ cada\ estrato\ \ X\ \ Proporción\ (p)$$

En este sentido, para obtener el tamaño de muestra del primer estrato (solteras), con una población de 420, por la proporción de 0.288 se obtendrá un tamaño de muestra de este estrato de 121. Para el segundo estrato (casadas), con una población de 242 por la proporción de 0.288, se obtendrá un tamaño de muestra del estrato de 70; para los estratos de unión libre y viudas se aplicará el mismo procedimiento como se muestra en la siguiente tabla:

Tabla 85. ***Resultados***

No. de estrato	Nombre del estrato, según estado civil de las mujeres	Población total por estrato por proporción	Tamaño de muestra por estrato
1	Solteras	420 x 0.288	121
2	Casadas	242 x 0.288	70
3	Unión libre	183 x 0.288	53
4	Viudas	105 x 0.288	30
		N = 950	***n* = 274**

Fuente: elaboración propia.

La suma del tamaño de muestra obtenido de cada estrato debe coincidir con el tamaño de muestra total calculado en el primer paso, es este caso 274 mujeres.

Para definir a los sujetos a quienes se les tiene que aplicar un instrumento se debe recurrir a un esquema de muestreo aleatorio simple en cada tamaño de muestra por estrato, es decir, aplicar un primer muestreo aleatorio simple a 420 mujeres solteras para obtener una muestra de 121, un segundo muestreo aleatorio simple a 242 mujeres casadas para obtener a 70 mujeres, así sucesivamente.

8.3.1.4. Aleatorio estratificado y por racimos

Hernández et al.(2014) señalan que cuando el investigador se ve limitado por recursos financieros, por tiempo, distancias geográficas o por una combinación de estos y otros obstáculos, se recurre al muestreo por racimos.

Muestreo aleatorio estratificado y por racimos

Es el procedimiento de selección de sujetos o elementos muestrales inmersos en conjuntos, grupos o estratos definidos física o geográficamente.

Entre las ventajas que se pueden atribuir al muestreo estratificado y por racimos se encuentran su bajo costo y su reducido empleo de tiempo y esfuerzo.

Este tipo de muestreo basa su desarrollo en la extracción de unidades de análisis que se encuentran encapsuladas, agrupadas o encerradas en determinados lugares físicos o geográficos denominados racimos.

El muestreo estratificado y por racimos –señalan Hernández et al. (2014)– implica diferenciar entre la unidad de análisis (es decir, quiénes van a ser medidos o a quién se le va aplicar el instrumento de medición) y la unidad muestral (que se refiere al racimo a través del cual se logra el acceso a la unidad de análisis).

Ejemplo: un grupo interdisciplinario quiere hacer un estudio en la colonia Antonia M. M. sobre el manejo de la basura que se genera en casa. Se establece como población a todos aquellos hogares que pertenezcan a dicha colonia.

Dadas las condiciones a las que se hace mención, el grupo interdisciplinario recurre a la selección de una muestra estratificada y por racimos, para ello, se hace llegar de un mapa de la colonia Antonia M. M., el cual indica que hay 145 cuadras. En este caso, las cuadras se utilizan como racimos, es decir, como unidades muestrales a partir de las cuales se obtienen las unidades de análisis o sujetos de estudio.

Primer paso. Determinar el tamaño de la muestra a partir del número de racimos (en este caso cuadras).

¿Cuántas cuadras se necesitan tomar como muestra, de una población total de 145, si se asume un nivel de confianza de 95 % o 0.95 con un error de precisión de 5 % o 0.05? No se tienen antecedentes sobre estudios realizados, por lo que se asume la máxima variabilidad; es decir, p = 50 % o 0.5 y q = 50 % o 0.5.

Fórmula:

$$n = \frac{N \times Z_a^{\,2} \times p \times q}{d^2 \times (N-1) + Z_a^{\,2} \times p \times q}$$

Identificando valores

n = tamaño de la muestra, valor a calcular.
N = tamaño de la población.
Za = nivel de confianza: 95 % o 0.95, correspondiente en la distribución normal, Za = 1.96.
p = variabilidad positiva: 50 % o 0.5.
q = variabilidad negativa: 50 % o 0.5.
d = precisión o error que se prevé cometer: 5 % o 0.05.

Sustituyendo

$$n = \frac{(145)\ (1.96)^2\ (50)\ (50)}{(5)^2(145\text{-}1) + (1.96)^2\ (50)\ (50)} = 105$$

¿Cuántas cuadras se necesitan tomar como muestra, de una población total de 145, si se asume un nivel de confianza de 95 % o 0.95 con un error de muestreo de 5 % o 0.05?

Respuesta: **105 cuadras**.

Segundo paso. Seleccionar los estratos y aplicar muestreo estratificado.

El grupo interdisciplinario divide a la colonia Antonia M. M. en cuatro estratos por punto cardinal (norte, sur, este, oeste). La relación estrato-número de cuadras se presenta a continuación:

Tabla 86. ***Relación estrato-número de cuadras***

Número de estratos	**Número de cuadras o racimos por estrato**
1	56
2	42
3	29
4	18
Total	**145**

Fuente: elaboración propia.

Obtener la **proporción (*p*)** a extraer de cada estrato definido por el tamaño de la muestra (n = 105) entre el total de la población (N=145); mediante la siguiente fórmula:

$$k = \frac{Tamaño\ de\ la\ Población}{Tamaño\ de\ la\ muestra} = \frac{N}{n} = \frac{105}{145} = 0.724$$

Tercer paso. Dentro de los racimos seleccionados, elegir las unidades de análisis que van a medirse.

a) En primera instancia, se determina el número de hogares-sujetos de estudio por cada racimo o cuadra. Este valor se obtiene por medio de la numeración y conteo de los hogares-sujetos de estudio por cuadra o racimo, de acuerdo con el mapa de la colonia Antonia M. M.

b) Se multiplica por el tamaño de la muestra por estrato y racimo para obtener el total de hogares o unidades de análisis por estrato a quienes se les aplicará el instrumento.

Tabla 87. ***Resultados***

Número de estratos	Número de cuadras o racimos por estrato	Proporción	Tamaño de la muestra por estrato y racimo	Número de hogares-sujetos de estudio por cuadra o racimo	Total de hogares por estrato
1	56	x 0.724 =	41	x 23 =	943
2	42	x 0.724 =	30	x 15 =	450
3	29	x 0.724 =	21	x 19 =	399
4	18	x 0.724 =	13	x 25 =	325
Total	***N* = 145**		***n* = 105**	**82**	**2117**

Fuente: elaboración propia.

En este sentido, para obtener el tamaño de muestra o total de hogares por estrato del primer subconjunto Norte con 56 cuadras o racimos, se debe multiplicar por la proporción 0.724 para obtiene 41 cuadras que será la muestra estratificada y multiplicarlas por 23 que es el número de hogares en esa cuadra para obtener 943 hogares a entrevistar. Para obtener el tamaño de muestra o total de hogares por estrato del segundo subconjunto Sur, con 42 cuadras o racimos, se debe multiplicar por la proporción 0.724 para obtener 30 cuadras, que será la muestra estratificada y multiplicarlas, por 15 que es el número de hogares en esa cuadra para obtener 450 hogares a entrevistar. Para los estratos de Este y Oeste se aplicará el mismo procedimiento como se muestra en tabla anterior.

Las unidades de análisis u hogares por estrato se deben seleccionar aleatoriamente para asegurar que cada uno de ellos tenga la misma probabilidad de ser elegido. Es decir, se enumeran los hogares del primer estrato y se aplica un muestreo aleatorio simple, los demás estratos se deben someter a la misma situación .

8.3.2. Muestreo no probabilístico

Velasco et al. (2002) establecen que el **muestreo no probabilístico**

> se presenta si la muestra es escogida por medio de un proceso subjetivo o arbitrario de modo que la probabilidad de selección de cada unidad de la población no es conocida (se utiliza con frecuencia cuando no se conoce el marco muestral). (p. 17)

Muestreo no probabilístico

Es el procedimiento en el que la elección de los elementos no depende de la probabilidad, sino de las características determinadas por el investigador o por quien hace la muestra.

Selltiz et al. (1980) establecen que en el muestreo de no probabilidad, no probabilístico o determinístico "no existe forma de estimar la probabilidad que cada elemento tiene de ser incluido en la muestra, y no hay tampoco seguridad de que cada elemento tenga alguna posibilidad de ser incluido" (p. 687).

Hernández et al. (2014) comparten tal aseveración al afirmar que, al ser muestras no probabilísticas, no es posible calcular con precisión el error estándar, es decir, no se puede calcular con qué nivel de confianza se hace una estimación.

El muestreo de tipo no probabilístico presenta como principal desventaja el tener un valor limitado y relativo para la muestra en sí, y nulo para la población. Es decir, no existe sustento estadístico para poder hacer generalizaciones a una población, dado que en el **muestreo no probabilístico** los sujetos de una población determinada no tienen la misma probabilidad de ser elegidos, sino que dependen de las características o condiciones que establezca el investigador o grupo de encuestadores, para generar una muestra.

Una de las ventajas que presenta este tipo de muestreo es su bajo costo y escaso empleo de tiempo. Por otra parte, en este tipo de muestreo el investigador no requiere conocimientos avanzados de estadística.

De tal modo, Hernández et al. (2014) establecen que una muestra no probabilística "es útil para determinado diseño de estudio que requiere no tanto de una representatividad de elementos de una población, sino una cuidadosa y controlada elección de los sujetos con ciertas características según se estipule en el planteamiento del problema" (pág. 327).

Por otra parte, dado que la finalidad no radica en la generalización estadística de los hallazgos, las muestras no probabilísticas adquieren relevancia, ya que permiten —si se realiza una selección cuidadosa y tras una inmersión profunda en el campo— identificar casos específicos (personas, contextos o situaciones) que son pertinentes para los objetivos del investigador y que ofrecen un potencial para la recolección y el análisis cualitativo de la información.

8.3.2.1. De juicio o criterio

En este tipo de muestreo tanto el tamaño de la muestra como la elección de los elementos que la integran están sujetos al juicio del investigador, basado generalmente en su experiencia.

Arias F. (2012) lo denomina **muestreo intencional u opinático**, "en este caso los elementos son escogidos con base en criterios o juicios preestablecidos por el investigador" (p. 85). Por su parte, Elorza (2008) señala que este tipo de muestreo es útil y aún aconsejable cuando el muestreo probabilístico no es factible o resulta costoso.

Muestreo de juicio o criterio

Es el procedimiento en el que la conformación de la muestra depende del criterio, juicio o discernimiento del investigador quien se apoya generalmente en su praxis profesional.

Ejemplo: un estudiante de la ENTS desea conocer el tipo de adicción a drogas, tanto legales como ilegales, que presentan los niños de la calle. Resulta difícil obtener una lista de niños de la calle y plantear una muestra aleatoria simple, por ello es conveniente una muestra de juicio o criterio. De acuerdo con su praxis profesional y conocimiento del tema decide acudir a aplicar un instrumento a los niños de la calle que se reúnen en la estación del metro Observatorio en la Ciudad de México.

8.3.2.2. De diseño de bola de nieve

Este tipo de diseño resulta conveniente cuando se necesita alcanzar poblaciones pequeñas o datos especializados.

Babbie (2000) establece que para llevar a cabo este tipo de muestreo se reúnen los datos de los pocos miembros que se conozcan de una población objetivo que se pueda localizar y se le pide la información necesaria para ubicar a otros miembros que conozcan de esa población.

Muestreo de diseño de bola de nieve

Es el procedimiento en el que la configuración de la muestra se construye a partir de la referencia de un sujeto a otro; es decir, de la acumulación de elementos a través de la remisión del investigador de un individuo a otro y de este a otro.

La denominación de bola de nieve se refiere a la acumulación que resulta de que cada sujeto localizado proponga a otro u otros. Debido a que este procedimiento da por resultado muestras de representatividad cuestionable, se usa sobre todo con fines exploratorios.

Ejemplo: un grupo de estudiantes de Trabajo Social de la UNAM desea realizar un estudio sobre las condiciones sociolaborales de los inmigrantes mexicanos a Estados Unidos que pertenecen al municipio de Atotonilco de Tula en el estado de Hidalgo, México, una vez que han regresado a su lugar de origen.

Como es difícil contar con una lista de personas que cumpla con tales características y aplicar un procedimiento de muestreo de tipo probabilístico se recurre a un muestreo no probabilístico de carácter bola de nieve; es decir, se localiza a un inmigrante que cumpla con las características para aplicarle un instrumento, al mismo se le pregunta de otra persona que cumpla con el perfil y a este último se le cuestiona sobre otro sujeto de estudio, y así sucesivamente. En este sentido, la bola de nieve crecería en la medida que los encuestados señalen a otra persona y esta a otra.

8.3.2.3. Accidental o por accidente

Castañeda et al. (2002) señalan que este tipo de muestreo consiste en seleccionar de manera arbitraria a los individuos que conformarán la muestra; radica en aprovechar o utilizar para el estudio a las personas disponibles en un momento dado según lo que interese estudiar.

Arias (2012) lo denomina **muestreo casual o accidental**, el cual "es un procedimiento que permite elegir arbitrariamente los elementos sin un juicio o criterio preestablecido" (p. 85). Mientras que Selltiz et al. (1980) señalan que, en este tipo de muestra, "el investigador solo puede desear que la equivocación no sea demasiado grande" (p. 689).

Muestreo accidental o por accidente

Es el procedimiento en el que el investigador decide por voluntad propia qué elementos integran la muestra, aprovechando los sujetos de que dispone.

Ejemplo: un alumno de la ENTS-UNAM desea conocer cuál es la opinión de sus compañeros respecto a la forma de impartir la clase de sus profesores de asignatura. En este sentido, aplica un instrumento a los compañeros que asistieron en un día determinado y obtiene resultados.

8.3.2.4. Por conveniencia

Elorza (2008) señala que cuando una muestra está conformada únicamente por elementos disponibles o con los más dispuestos, entonces se trata de una **muestra por conveniencia**.

> Para Velasco et al. (2002) en un **muestreo por conveniencia**
>
> se seleccionan a las unidades de estudio que se encuentren disponibles al momento de la recolección de datos. Su ventaja es que es más fácil, económico y accesible y puede dar una visión inicial buena. Se usa en estudios exploratorios. Su desventaja es que puede ser poco representativo, algunas unidades estarán subrepresentadas y otras sobrerrepresentadas. (p. 18)

Muestreo por conveniencia

Es el procedimiento en el que el investigador decide de acuerdo con los objetivos del estudio los elementos integrarán la muestra.

Ejemplo: un profesor universitario desea introducir un nuevo método de enseñanza de la asignatura de Estadística. Con base en su experiencia decide aplicar un instrumento a estudiantes de las licenciaturas en Trabajo Social, Psicología y Medicina. Todo ello a conveniencia del investigador.

8.3.2.5. Por cuota

En el muestreo por cuota se determina una cantidad (cuota) de individuos de una población para que sean miembros de la muestra. El investigador selecciona una muestra considerando algunos fenómenos o variables a estudiar como sexo, raza, religión, etcétera.

> Velasco et al. (2002) establecen que en un **muestreo por cuotas**
>
> se seleccionan unidades de estudio de cada uno de los subgrupos que componen la población en una cuota predeterminada. Asegura que un determinado número de unidades de muestreo de diferentes categorías aparezcan en la muestra de modo que todos queden representados. Es útil para balancear las unidades de estudio, pero no se obtiene la representatividad de la población. (p. 19)

Pick y López (1998) señalan que para utilizar el muestreo de cuota primero se tiene que conocer la población que se pretende estudiar y hacer una clasificación por estratos de acuerdo con los objetivos del estudio. Consideran que no es necesario que se tome una proporción representativa de cada estrato, aunque se tiene que recoger un porcentaje mínimo de cada estrato. Una vez que se deciden los estratos, el investigador elige arbitrariamente los sujetos de estudio que integrarán cada uno de los estratos.

Muestreo por cuota

Es el procedimiento en el que el investigador establece una cantidad denominada cuota que obtiene de una categoría o estrato para conformar la muestra.

Por lo tanto, –dicen Pick y López (1998)– el muestreo de cuota constituye un método de muestreo estratificado en el cual la selección dentro de los estratos no es al azar, sino accidental. Lo anterior –establecen– representa su principal desventaja.

Ejemplo: un profesor de la ENTS-UNAM desea investigar cuáles son las necesidades de formación extracurricular de los estudiantes de Trabajo Social de la UNAM. Decide generar estratos por semestre par y considerar un 25 % de cada uno de ellos; por lo que obtiene los siguientes resultados:

Tabla 88. ***Resultados***

Semestre	Población	Porcentaje a extraer	Muestra
Segundo	100	25	25
Cuarto	96	25	24
Sexto	80	25	20
Octavo	76	25	19
Total	352	-	88

Fuente: elaboración propia.

El total de estudiantes que integran la muestra es de **88**.

8.3.2.6. Por sujetos-tipo

Hernández et al. (2014) apuntan que el muestreo por sujetos-tipo es aquél cuyo objetivo es la riqueza, profundidad y calidad de la información, no la cantidad ni la estandarización. Se utiliza en estudios de tipo exploratorio, así como en investigaciones de carácter cualitativo como es el caso de los estudios con perspectiva fenomenológica.

Muestreo por sujetos-tipo

Es procedimiento en el que el investigador elige elementos de muestra que comparten las mismas características (políticas, económicas, sociales, académicas, etc.) para generar información de tipo cualitativo más que cuantitativo.

Ejemplo: un investigador está interesado en conocer de manera directa cuál es la percepción de los estudiantes respecto al desempeño de los profesores de la ENTS-UNAM en el salón de clase; para ello, forma quince grupos de diez alumnos, al mismo tiempo programa una sesión por cada grupo, con el objetivo de que expresen sus puntos de vista respecto a la variable a medir bajo indicadores de actitudes, motivación, inconformidad, liderazgo, etcétera. El investigador sistematiza la información y obtiene resultados.

8.3.2.7. Por sujetos voluntarios

Se trata de muestras ocasionales donde los sujetos voluntariamente acceden a participar en un estudio. Se procura que los sujetos sean homogéneos en variables tales como edad, sexo, ocupación, de manera que los resultados o efectos no obedezcan a diferencias individuales, sino a condiciones a las que fueron sometidos.

Muestreo por sujetos voluntarios

Es el procedimiento en el que las unidades muestrales o elementos a considerar deciden libremente formar parte o no de la muestra.

Ejemplo: un profesor desea probar dos métodos de enseñanza en estadística. Para ello requiere integrar dos grupos, por lo que el académico decide convocar de manera abierta a estudiantes universitarios. Quienes estén interesados y acudan al llamado formarán parte de la muestra. En este caso, la elección de los individuos sujetos de estudio depende de circunstancias fortuitas.

8.3.2.8. Por expertos

Se trata de estudios basados en la opinión de sujetos expertos en un tema determinado. Estas muestras son frecuentes en estudios de corte cualitativo y exploratorio, con el objetivo de generar hipótesis más precisas o como base para la construcción de instrumentos de medición.

Muestreo por expertos

Es el procedimiento de selección de elementos muestrales que depende de la experiencia, reconocimiento o prestigio que una persona pueda tener respecto a un tema determinado.

Ejemplo: un investigador de la Escuela Nacional de Trabajo Social desarrolla un estudio sobre pobreza en México y recurre a una muestra de n =15 expertos. Todos son sujetos idóneos y de reconocido prestigio para abordar y discutir sobre tipología, índices, factores, organismos que la combaten, etcétera.

Glosario

Analizar o inferir. Etapa del método estadístico que proporciona los métodos para estimar las características de un grupo total (población), basándose en datos de un conjunto pequeño (muestra) de observaciones.

Barras separadas. Es un gráfico para variables cualitativas que se compone de una serie de barras verticales u horizontales, donde la extensión de la barra representa la frecuencia absoluta, proporcional o relativa de una categoría.

Barras segmentadas o apiladas 100 %. Es un gráfico para mostrar de manera simultánea dos variables cualitativas donde se establecen el porcentaje del total de cada grupo, lo que permite comparar diferencias relativas entre los mismos.

Caja y bigotes. Es un gráfico para variables cuantitativas que muestra una distribución de datos a través de cuartiles (Q), así como de puntuaciones mínimas y máximas.

Centil. Son los 99 valores que dividen datos clasificados en 100 partes, con aproximadamente el 1 % de los valores en cada grupo.

Circular. Es un gráfico para variables cualitativas en el que un círculo que se divide o segmenta desde su punto central, donde cada espacio representa la frecuencia proporcional -o relativa- de determinada categoría.

Coeficiente de confianza de la estimación o nivel de confianza. Es la medida probabilística de que el intervalo fijado con d (error de precisión) contenga el valor poblacional.

Coeficiente de variación. Es una relación o razón estadística entre la desviación estándar y la media aritmética multiplicada por 100.

Contar. Etapa del método estadístico que consiste en determinar las frecuencias que tiene cada una de las modalidades o clases de las variables de estudio.

Continuas. Nivel de medición de las variables cuantitativas que admiten cualquier valor dentro de un rango numérico determinado. Son el resultado de medir, por lo que asumen valores numéricos fraccionados o decimales.

Cuartil. Son los tres valores que dividen datos ordenados en cuatro segmentos, con aproximadamente el 25 % de los valores en cada grupo.

Curtosis. Grado de concentración de valores de una variable alrededor de la zona central de una distribución que determina apuntamiento, normalidad o allanamiento de una curva.

Curva normal. Distribución teórica y simétrica con forma de campana. El eje horizontal representa todos los posibles valores de una variable y el eje vertical la probabilidad de que ocurran dichos valores.

Describir. Etapa del método estadístico que consiste en el cálculo de medidas estadísticas de resumen que representan las características generales de un conjunto de datos; implica obtener medidas de tendencia central, posición, dispersión y de distribución.

Diagrama de puntos o correlación. Gráfico para presentar de manera simultánea dos variables de tipo cuantitativo para mostrar la asociación que existe entre las mismas.

Decil. Son los nueve valores que dividen la serie de datos en diez partes iguales, con aproximadamente el 10 % de los valores en cada grupo.

Discretas. Nivel de medición de las variables cuantitativas que no admiten todos los valores intermedios en un rango. Son el resultado de contar, por lo que asumen valores numéricos enteros.

Desviación estándar. Es la raíz cuadrada del promedio de desviaciones de las puntuaciones cuadráticas con respecto a la media de una serie de datos.

Estadística. Rama de las matemáticas que recolecta, cuenta, presenta, describe y analiza un conjunto de datos.

Estadística descriptiva. Rama de la estadística que recolecta, cuenta, presenta y establece las características generales de un subconjunto de datos denominado muestra.

Estadística inferencial o analítica. Rama de la estadística que establece los métodos y procedimientos para estimar las características de un conjunto total (denominado población), basándose en datos de un subconjunto de observaciones (llamado muestra).

Error máximo admisible (d) o error de precisión. Es la máxima diferencia que se puede tolerar entre el valor de la variable obtenido en la muestra y el verdadero valor de esta en el universo o la población

Frecuencia absoluta. Es el número de veces que se repite un dato.

Frecuencia relativa. Es el porciento de veces en que ocurre un dato. La expresión en porcentaje de la frecuencia absoluta.

Frecuencia absoluta acumulada. Es la frecuencia absoluta de cada uno de los datos más la suma de los valores anteriores a dicha suma.

Frecuencia relativa acumulada. Es la frecuencia relativa de cada uno de los datos más la suma de los valores anteriores a dicha suma.

Gráfico. Es una representación visual de los datos a través de formas, objetos, figuras, líneas, puntos, barras o círculos.

Homogeneidad de la población o variabilidad. Es la presencia (éxito) y ausencia (fracaso) de una variable dentro de una población expresada en porcentaje o proporción, está definida por p = probabilidad de éxito y q = probabilidad de fracaso.

Histograma. Es un gráfico para variables cuantitativas que se expresa mediante un diagrama de 90 grados, que muestra puntuaciones cuantitativas a lo largo del eje horizontal y la frecuencia de cada puntuación, en una columna al eje vertical.

Investigación aplicada. Propósito de la investigación cuyo objetivo es la resolución de un problema práctico, real, inmediato o concreto.

Investigación básica. Propósito de la investigación cuyo objetivo es desarrollar teorías por medio del descubrimiento de generalizaciones o principios.

Investigación. Conjunto de procesos sistemáticos, críticos y empíricos que se aplican al estudio de un fenómeno o problema de carácter natural o social.

Investigación social. Proceso sistemático, controlado, empírico y crítico de aseveraciones hipotéticas para el estudio de las interrelaciones que presentan los sujetos en lo individual o en lo colectivo dentro de un contexto denominado sociedad.

Lo social. Conjunto de interrelaciones e interacciones que presentan las personas en lo individual o en lo colectivo dentro de un contexto denominado sociedad.

Media. Es el valor único que resume a toda una serie de datos.

Mediana. En una serie de valores ordenados de mayor a menor o viceversa, es aquel valor que divide de manera exacta a una serie de datos en dos partes de igual tamaño.

Medidas de dispersión. Indican la dispersión o variabilidad de los datos de una variable cuantitativa.

Medidas de distribución o de forma. Son aquellas que reflejan el grado de simetría y concentración de datos respecto a su medida central (media).

Medidas de posición. Son valores relativos de una distribución de datos que la dividen en partes iguales.

Medidas de tendencia central. Son valores, o punto de referencia, de una distribución alrededor del cual se distribuyen los datos.

Método estadístico. Secuencia sistemática de procedimientos para el estudio de los datos cualitativos y cuantitativos, derivados de la investigación con el propósito de rechazar o no rechazar una hipótesis estadística.

Moda. Es el valor de mayor frecuencia u ocurrencia en un conjunto de datos.

Muestra. Es un subconjunto, fracción o subgrupo de la totalidad de un conjunto de elementos, seres u objetos seleccionados para un estudio.

Muestreo. Es el conjunto de procedimientos que se ejecutan para seleccionar un subconjunto, fracción o subgrupo de elementos.

Muestreo accidental o por accidente. Es aquel en el que el investigador decide por voluntad propia qué elementos integran la muestra aprovechando los sujetos de que dispone.

Muestreo aleatorio estratificado. Es la selección de unidades de muestra a partir de dividir en subgrupos o estratos a la población de estudio y extraer proporcionalmente de cada uno de ellos los elementos que integrarán la muestra.

Muestreo aleatorio estratificado y por racimos. Es la elección de sujetos o elementos muestrales inmersos en conjuntos, grupos o estratos definidos física o geográficamente.

Muestreo aleatorio simple. Es una técnica que asigna la misma probabilidad de pertenecer a la muestra a todos los elementos de la población.

Muestreo aleatorio sistemático. Es la elección de elementos para una muestra mediante un *késimo* o intervalo fijo después de seleccionar un punto de inicio.

Muestreo de diseño de bola de nieve. Es aquel en el que la configuración de la muestra se construye a partir de la referencia de un sujeto a otro; es decir, de la acumulación de elementos a través de la remisión del investigador de un individuo a otro y de este a otro.

Muestreo de juicio o criterio. Es aquel en el que la conformación de la muestra depende del criterio, juicio o discernimiento del investigador, quien se apoya generalmente en su praxis profesional.

Muestreo no probabilístico. Es aquel en la que la elección de los elementos no depende de la probabilidad, sino de las características determinadas por el investigador o por quien hace la muestra.

Muestreo por conveniencia. Es aquel en el que el investigador decide de acuerdo con los objetivos del estudio los elementos integrarán la muestra.

Muestreo por cuota. Es aquel en el que el investigador establece una cantidad denominada cuota que obtiene de una categoría o estrato para conformar la muestra.

Muestreo por expertos. Es aquel en el que la selección de los elementos muestrales depende de la experiencia, reconocimiento o prestigio que una persona pueda tener respecto a un tema determinado.

Muestreo por sujetos-tipo. Es aquel en el que el investigador elige elementos muestras que comparten las mismas características (políticas, económicas, sociales, académicas, etc.) para generar información de tipo cualitativo más que cuantitativo.

Muestreo por sujetos voluntarios. Es aquel en el que las unidades muestrales o elementos a considerar deciden libremente formar parte o no de la muestra.

Muestreo probabilístico. Es aquel en el que todos los elementos de una población tienen la misma probabilidad de ser elegidos.

Nominal. Nivel de medición de las variables cualitativas en las que los datos se ajustan por categorías que no mantienen una relación de orden entre sí.

Ordinal. Nivel de medición de las variables cualitativas en las que hay un orden o jerarquía entre las categorías.

Población. Es la totalidad de un conjunto de elementos, seres u objetos.

Pictograma. Es un gráfico para variables cualitativas de figuras, formas u objetos, donde el número o el tamaño de los objetos representen la frecuencia absoluta, proporcional o relativa de una categoría.

Polígono de frecuencias. Es un gráfico para variables cuantitativas que se expresa mediante un diagrama de 90 grados con puntuaciones de nivel de medición continua –incluso discreta– señaladas en el eje horizontal; las frecuencias de las puntuaciones están representadas por las alturas de puntos localizados sobre las puntuaciones y conectados mediante líneas rectas.

Polígono de frecuencias acumuladas. Es un gráfico para variables cuantitativas que presenta frecuencias

relativas acumuladas o frecuencias porcentuales acumuladas de una variable.

Porcentaje. Es la importancia de un subconjunto con respecto al conjunto al que pertenece multiplicado por 100.

Proporción. Es la importancia de un subconjunto con respecto al conjunto al que pertenece.

Presentar. Etapa del método estadístico que consiste en mostrar de manera numérica o visual la frecuencia de los datos a través de tablas y gráficos dependiendo del nivel de medición de la o las variables a exponer.

Rango. Es la diferencia entre el valor mayor y el valor menor de una serie de datos.

Rango intercuartílico. Es la diferencia entre el cuartil 3 (Q_3) y el cuartil 1(Q_1) en una serie de datos.

Razón. Es la relación o existencia de un elemento con respecto a otro.

Recolectar. Etapa del método estadístico que consiste en el acopio de la información cualitativa y cuantitativa a través de variables estableciendo su nivel de medición.

Sesgo. Es la ausencia de simetría en una distribución; es decir, se presenta cuando la mitad de una distribución de datos reflejada en una curva no es exactamente igual o espejo de la otra mitad.

Tabla. Es un cuadro de organización, jerarquización y comprensión numérica de datos.

Tabla de doble entrada. Es una matriz o cuadro que permite organizar y sistematizar información a partir de columnas horizontales y verticales que concentran y relacionan una serie de datos.

Tabla de lista simple de datos. Es una lista de datos sobre la relación de dos conjuntos de valores o variables respecto a un sujeto u objeto de medición.

Tasa. Es la magnitud o probabilidad de ocurrencia de un evento dentro de una población determinada.

Variable. Es una característica o valor que puede variar y cuya variación es susceptible de medirse.

Variables cualitativas. Son aquellas que representan una cualidad o atributo que clasifica a cada caso en una de varias categorías.

Variables cuantitativas. Son las variables que pueden medirse, cuantificarse o expresarse numéricamente.

Variable independiente. Es la variable de un experimento que es controlada en forma sistemática por el investigador. Es la causa, origen o razón que establece o controla el investigador.

Variable dependiente. Es la variable de un experimento, medida por el investigador, para determinar el efecto de una variable independiente. Es la consecuencia, efecto o resultado que mide el investigador.

Varianza. Es la desviación estándar elevada al cuadrado o el cuadrado del promedio de desviación de las puntuaciones cuadráticas con respecto a la media de una serie de datos.

Bibliografía

Ander-Egg, E. (2003). *Métodos y Técnicas de Investigación Social IV. Técnicas para la recogida de información*. Lumen.

Ander-Egg, E. (1995). *Técnicas de investigación social*. Lumen.

Ander-Egg, E. (2011). *Aprender a investigar. Nociones básicas para la investigación social*. Brujas.

Ángel, M. (Dir.). (2006). *Génesis y evolución histórica de los conceptos de probabilidad y estadística como herramienta metodológica*. Universidad Nacional de la Matanza. https://economicas.unlam.edu.ar/descargas/5_b107.pdf

Arias, F. (2012). *El proyecto de investigación. Introducción a la metodología científica*. Episteme.

Aris, R. y Martínez, F. (2013). Estadística para aterrorizados: recomendaciones para describir sus datos. *Medwave*, *13*(10). https://www.medwave.cl/link.cgi/Medwave/Series/MBEyEpi/5826

Barreto-Villanueva, A. (2012). El progreso de la Estadística y su utilidad en la evaluación del desarrollo. *Papeles de Población*, *18*(73), 1-31. https://www.redalyc.org/pdf/112/11224638010.pdf

Blalock, H. (1986). *Estadística social*. Fondo de Cultura Económica.

Bologna, E. (2013). *Estadística para psicología y educación*. Brujas.

Babbie, E. (2000). *Fundamentos de la investigación social*. Thomson Editores.

Bueno, C. y Escudero, T. (2006-2007). *Apuntes de estadística para profesores*. Universidad de Zaragoza. Instituto de Ciencias de la Educación.

Calduch Cervera, R. (2014). *Métodos y técnicas de investigación internacional*. Universidad Complutense de Madrid.

Campos, A. L. (2008). Una aproximación al concepto de "lo social" desde trabajo social. *Revista Tendencias & Retos*, (13), 55-70. http://www.ts.ucr.ac.cr/binarios/revistas/co/rev-co-tendencias-0013-05.pdf

Carballeda, A. (2004). *La intervención en lo social. Exclusión e integración en los nuevos escenarios sociales*. Paidós.

Castañeda, J., De la Torre, M., Morán, J. y Lara, L. (2002). *Metodología de la investigación*. McGraw-Hill.

Celis de la Rosa, J. A. y Labrada, V. (2014). *Bioestadística*. Manual Moderno.

Chou, Y. L. (1990). *Análisis Estadístico*. Mc. Graw-Hill.

Córdova, V. y Cortés, A. (2010). *Estadística y probabilidad*. Dirección Académica del Colegio de Bachilleres del Estado de Sonora.

Cozby, P. (2004). *Métodos de investigación del comportamiento*. Mc Graw Hill.

Elorza, H. (2008). *Estadística para las ciencias sociales y del comportamiento*. Oxford University Press.

Estuardo, A. (2012). *Estadística y probabilidad*. Universidad Católica de la Santísima Concepción: Instituto Profesional Virgilio Gómez de la Universidad de la Concepción.

Flores, R. y Lozano, H. (1998). *Estadística aplicada para administración*. Iberoamérica.

Gonick, L. y Smith, W. (1989). *La estadística en cómic*. Zendrera Zariquiey.

Gorgas, J., Cardiel, N. y Zamorano, J. (2011). *Estadística básica para estudiantes de ciencias*. Departamento de

Astrofísica y Ciencias de la Atmósfera. Facultad de Ciencias Físicas. Universidad complutense de Madrid.

Hernández, R., Fernández, C. y Baptista, P. (2014). *Metodología de la investigación*. McGraw-Hill.

Holguín, F. (1981). *Estadística descriptiva aplicada a las ciencias sociales*. Facultad de Ciencias Políticas y Sociales-UNAM.

Kelmansky, D. (2009) *Estadística para todos. Estrategias de pensamiento y herramientas para la solución de problemas*. Ministerio de Educación. Instituto Nacional de Educación Tecnológica.

Kerlinger, F. N. (1981). *Investigación del comportamiento. Técnicas y metodología*. Interamericana

Levin, J. y Levin, W. (2004). *Fundamentos de estadística en la investigación social*. Oxford-Alfaomega.

López, C. (2004). *Muestreo: Tamaño y tipología*. ENTS-UNAM.

López, C. (2003). *Estadística aplicada a la investigación social I*. ENTS-UNAM.

Martínez, C. (2019). *Estadística básica aplicada*. ECODE Ediciones.

Mendenhall, W., Beaver, R. y Beaver, B. (2010). *Introducción a la probabilidad y estadística*. CENGAGE Learning.

Miller, J. E. (2004). *The Chicago guide to writing about numbers*. University of Chicago Press.

Moncho, J. (2015). *Estadística aplicada a las ciencias de la salud*. Elsevier.

Morán, G. y Alvarado, D. (2010). *Métodos de investigación*. Pearson.

Moreno, I. D. (2017). La investigación social, un acercamiento a lo cotidiano. *Revista Electrónica de Investigación Educativa*, *19*(4), 1-3. http://redie.uabc. mx/redie/article/view/1872

Muñoz J. (2012, junio). *Investigación social*. Contribuciones a las Ciencias Sociales. http://www.eumed.net/rev/cccss/20/jlmc6.html

Niño, V. (2011). *Metodología de la investigación*. Ediciones de la U.

ONU-CEE. (2009a). *Cómo hacer comprensibles los datos. Parte 1. Una guía para escribir sobre números*. ONU.

ONU-CEE. (2009b). *Cómo hacer comprensibles los datos. Parte 2. Una guía para presentar estadísticas*. ONU.

Pagano, R. (1999). *Estadística para las ciencias del comportamiento*. Thomson Internacional Editores.

Pick, S. y López, A. (1998). *Cómo investigar en ciencias sociales*. Trillas.

Popper, K. R. (1974). *Conocimiento objetivo*. Tecnos.

Reynaga, O., J. De Garay, G. B. y García, R. J. (1996). *Módulo preparatorio*. Unidad de Bioestadística, Depto. de Medicina Social, Preventiva y Salud Pública. Facultad de Medicina UNAM.

Ríus, F., Barón, F. J., Sánchez, E. y Parras, L. (1998). *Bioestadística: Métodos y aplicaciones*. Universidad de Málaga.

Ritchey, F. (2004). *Estadística para las ciencias sociales. El potencial de la imaginación estadística*. Mc Graw Hill.

Ruiz, D. (2004). *Manual de estadística*. Universidad de Pablo de Olavide.

Rustom, A. (2012). *Estadística descriptiva, probabilidad e inferencia*. Facultad de Ciencias Agronómicas. Universidad de Chile.

Salazar, C. y Del Castillo, S. (2018). *Fundamentos básicos de estadística*. Sin editorial

Selltiz, C., Wrightsman, L. S. y Cook, S. W. (1980). *Métodos de investigación en ciencias sociales*. RIALP.

Sierra B., R. (1991). *Diccionario práctico de estadística*. Paraninfo.

Tamayo, M. (2004). *El proceso de investigación científica: incluye evaluación y administración de proyectos de investigación*.

Limusa.

Triola, M. (2009). *Estadística*. Pearson Educación de México.

Velasco, V. M., Martínez, V. A., Roiz, J., Huazano, F. y Nieves, A. (2002). *Muestreo y tamaño de muestra. Una guía práctica para personal de salud que realiza investigación*. e-libro.net.

Anexo

Tabla de áreas bajo la curva normal

(1) *Puntuación "Z"*	(2) *Distancia de "Z" a la media*	(3) *Área de la parte mayor*	(4) *Área de la parte menor*
0.00	.0000	.5000	.5000
0.01	.0040	.5040	.4960
0.02	.0080	.5080	.4920
0.03	.0120	.5120	.4880
0.04	.0160	.5160	.4840
0.05	.0199	.5199	.4801
0.06	.0239	.5239	.4761
0.07	.0279	.5279	.4721
0.08	.0319	.5319	.4681
0.09	.0359	.5359	.4641
0.10	.0398	.5398	.4602
0.11	.0438	.5438	.4562
0.12	.0478	.5478	.4522
0.13	.0517	.5517	.4483
0.14	.0557	.5557	.4443
0.15	.0596	.5596	.4404
0.16	.0636	.5636	.4364
0.17	.0675	.5675	.4325
0.18	.0714	.5714	.4286
0.19	.0753	.5753	.4247
0.20	.0793	.5793	.4207
0.21	.0832	.5832	.4168
0.22	.0871	.5871	.4129
0.23	.0910	.5910	.4090
0.24	.0948	.5948	.4052
0.25	.0987	.5987	.4013
0.26	.1026	.6026	.3974
0.27	.1064	.6064	.3936
0.28	.1103	.6103	.3897
0.29	.1141	.6141	.3859
0.30	.1179	.6179	.3821
0.31	.1217	.6217	.3783
0.32	.1255	.6255	.3745
0.33	.1293	.6293	.3707
0.34	.1331	.6331	.3669
0.35	.1368	.6368	.3632
0.36	.1406	.6406	.3594
0.37	.1443	.6443	.3557
0.38	.1480	.6480	.3520
0.39	.1517	.6517	.3483
0.40	.1554	.6554	.3446
0.41	.1591	.6591	.3409
0.42	.1628	.6628	.3372
0.43	.1664	.6664	.3336

0.44	.1700	.6700	.3300
0.45	.1736	.6736	.3264
0.46	.1772	.6772	.3228
0.47	.1808	.6808	.3192
0.48	.1844	.6844	.3156
0.49	.1879	.6879	.3121
0.50	.1915	.6915	.3085
0.51	.1950	.6950	.3050
0.52	.1985	.6985	.3015
0.53	.2019	.7019	.2981
0.54	.2054	.7054	.2946
0.55	.2088	.7088	.2912
0.56	.2123	.7123	.2877
0.57	.2157	.7157	.2843
0.58	.2190	.7190	.2810
0.59	.2224	.7224	.2776
0.60	.2257	.7257	.2743
0.61	.2291	.7291	.2709
0.62	.2324	.7324	.2676
0.63	.2357	.7357	.2643
0.64	.2389	.7389	.2611
0.65	.2422	.7422	.2578
0.66	.2454	.7454	.2546
0.67	.2486	.7486	.2514
0.68	.2517	.7517	.2483
0.69	.2549	.7549	.2451
0.70	.2580	.7580	.2420
0.71	.2611	.7611	.2389
0.72	.2642	.7642	.2358
0.73	.2673	.7673	.2327
0.74	.2704	.7704	.2296
0.75	.2734	.7734	.2266
0.76	.2764	.7764	.2236
0.77	.2794	.7794	.2206
0.78	.2823	.7823	.2177
0.79	.2852	.7852	.2148
0.80	.2881	.7881	.2119
0.81	.2910	.7910	.2090
0.82	.2939	.7939	.2061
0.83	.2967	.7967	.2033
0.84	.2995	.7995	.2005
0.85	.3023	.8023	.1977
0.86	.3051	.8051	.1949
0.87	.3078	.8078	.1922
0.88	.3106	.8106	.1894
0.89	.3133	.8133	.1867

0.90	.3159	.8159	.1841
0.91	.3186	.8186	.1814
0.92	.3212	.8212	.1788
0.93	.3228	.8238	.1762
0.94	.3264	.8264	.1736
0.95	.3289	.8289	.1711
0.96	.3315	.8315	.1685
0.97	.3340	.8340	.1660
0.98	.3365	.8365	.1635
0.99	.3389	.8389	.1611
1.00	.3413	.8413	.1587
1.01	.3438	.8438	.1562
1.02	.3461	.8461	.1539
1.03	.3485	.8485	.1515
1.04	.3508	.8508	.1492
1.05	.3531	.8531	.1469
1.06	.3554	.8554	.1446
1.07	.3577	.8577	.1423
1.08	.3599	.8599	.1401
1.09	.3621	.8621	.1379
1.10	.3643	.8643	.1357
1.11	.3665	.8665	.1335
1.12	.3686	.8686	.1314
1.13	.3708	.8708	.1292
1.14	.3729	.8729	.1271
1.15	.3749	.8749	.1251
1.16	.3770	.8770	.1230
1.17	.3790	.8790	.1210
1.18	.3810	.8810	.1190
1.19	.3830	.8830	.1170
1.20	.3849	.8849	.1151
1.21	.3869	.8869	.1131
1.22	.3888	.8888	.1112
1.23	.3907	.8907	.1093
1.24	.3925	.8925	.1075
1.25	.3944	.8944	.1056
1.26	.3962	.8962	.1038
1.27	.3980	.8980	.1020
1.28	.3997	.8997	.1003
1.29	.4015	.9015	.0985
1.30	.4032	.9032	.0968
1.31	.4049	.9049	.0951
1.32	.4066	.9066	.0934
1.33	.4082	.9082	.0918
1.34	.4099	.9099	.0901

1.35	.4115	.9115	.0885
1.36	.4131	.9131	.0869
1.37	.4147	.9147	.0853
1.38	.4162	.9162	.0838
1.39	.4177	.9177	.0823
1.40	.4192	.9192	.0808
1.41	.4207	.9207	.0793
1.42	.4222	.9222	.0778
1.43	.4236	.9236	.0764
1.44	.4251	.9251	.0749
1.45	.4265	.9265	.0735
1.46	.4279	.9279	.0721
1.47	.4292	.9292	.0708
1.48	.4306	.9306	.0694
1.49	.4319	.9319	.0681
1.50	.4332	.9332	.0668
1.51	.4345	.9345	.0655
1.52	.4357	.9357	.0643
1.53	.4370	.9370	.0630
1.54	.4382	.9382	.0618
1.55	.4394	.9394	.0606
1.56	.4406	.9406	.0594
1.57	.4418	.9418	.0582
1.58	.4429	.9429	.0571
1.59	.4441	.9441	.0559
1.60	.4452	.9452	.0548
1.61	.4463	.9463	.0537
1.62	.4474	.9474	.0526
1.63	.4484	.9484	.0516
1.64	.4495	.9495	.0505
1.65	.4505	.9505	.0495
1.66	.4515	.9515	.0485
1.67	.4525	.9525	.0475
1.68	.4535	.9535	.0465
1.69	.4545	.9545	.0455
1.70	.4554	.9554	.0446
1.71	.4564	.9564	.0436
1.72	.4573	.9573	.0427
1.73	.4582	.9582	.0418
1.74	.4591	.9591	.0409
1.75	.4599	.9599	.0401
1.76	.4608	.9608	.0392
1.77	.4616	.9616	.0384
1.78	.4625	.9625	.0375
1.79	.4633	.9633	.0367

1.80	.4641	.9641	.0359
1.81	.4649	.9649	.0351
1.82	.4656	.9656	.0344
1.83	.4664	.9664	.0336
1.84	.4671	.9671	.0329
1.85	.4678	.9678	.0322
1.86	.4686	.9686	.0314
1.87	.4693	.9693	.0307
1.88	.4699	.9699	.0301
1.89	.4706	.9706	.0294
1.90	.4713	.9713	.0287
1.91	.4719	.9719	.0281
1.92	.4726	.9726	.0274
1.93	.4732	.9732	.0268
1.94	.4738	.9738	.0262
1.95	.4744	.9744	.0256
1.96	.4750	.9750	.0250
1.97	.4756	.9756	.0244
1.98	.4761	.9761	.0239
1.99	.4767	.9767	.0233
2.00	.4772	.9772	.0228
2.01	.4778	.9778	.0222
2.02	.4783	.9783	.0217
2.03	.4788	.9788	.0212
2.04	.4793	.9793	.0207
2.05	.4798	.9798	.0202
2.06	.4803	.9803	.0197
2.07	.4808	.9808	.0192
2.08	.4812	.9812	.0188
2.09	.4817	.9817	.0183
2.10	.4821	.9821	.0179
2.11	.4826	.9826	.0174
2.12	.4830	.9830	.0170
2.13	.4834	.9834	.0166
2.14	.4838	.9838	.0162
2.15	.4842	.9842	.0158
2.16	.4846	.9846	.0154
2.17	.4850	.9850	.0150
2.18	.4854	.9854	.0146
2.19	.4857	.9857	.0143
2.20	.4861	.9861	.0139
2.21	.4864	.9864	.0136
2.22	.4868	.9868	.0132
2.23	.4871	.9871	.0129
2.24	.4875	.9875	.0125

2.25	.4878	.9878	.0122
2.26	.4881	.9881	.0119
2.27	.4884	.9884	.0116
2.28	.4887	.9887	.0113
2.29	.4890	.9890	.0110
2.30	.4893	.9893	.0107
2.31	.4896	.9896	.0104
2.32	.4898	.9898	.0102
2.33	.4901	.9901	.0099
2.34	.4904	.9904	.0096
2.35	.4906	.9906	.0094
2.36	.4909	.9909	.0091
2.37	.4911	.9911	.0089
2.38	.4913	.9913	.0087
2.39	.4916	.9916	.0084
2.40	.4918	.9918	.0082
2.41	.4920	.9920	.0080
2.42	.4922	.9922	.0078
2.43	.4925	.9925	.0075
2.44	.4927	.9927	.0073
2.45	.4929	.9929	.0071
2.46	.4931	.9931	.0069
2.47	.4932	.9932	.0068
2.48	.4934	.9934	.0066
2.49	.4936	.9936	.0064
2.50	.4938	.9938	.0062
2.51	.4940	.9940	.0060
2.52	.4941	.9941	.0059
2.53	.4943	.9943	.0057
2.54	.4945	.9945	.0055
2.55	.4946	.9946	.0054
2.56	.4948	.9948	.0052
2.57	.4949	.9949	.0051
2.58	.4951	.9951	.0049
2.59	.4952	.9952	.0048
2.60	.4953	.9953	.0047
2.61	.4955	.9955	.0045
2.62	.4956	.9956	.0044
2.63	.4957	.9957	.0043
2.64	.4959	.9959	.0041
2.65	.4960	.9960	.0040
2.66	.4961	.9961	.0039
2.67	.4962	.9962	.0038
2.68	.4963	.9963	.0037
2.69	.4964	.9964	.0036

2.70	.4965	.9965	.0035
2.71	.4966	.9966	.0034
2.72	.4967	.9967	.0033
2.73	.4968	.9968	.0032
2.74	.4969	.9969	.0031
2.75	.4970	.9970	.0030
2.76	.4971	.9971	.0029
2.77	.4972	.9972	.0028
2.78	.4973	.9973	.0027
2.79	.4974	.9974	.0026
2.80	.4974	.9974	.0026
2.81	.4975	.9975	.0025
2.82	.4976	.9976	.0024
2.83	.4977	.9977	.0023
2.84	.4977	.9977	.0023
2.85	.4978	.9978	.0022
2.86	.4979	.9979	.0021
2.87	.4979	.9979	.0021
2.88	.4980	.9980	.0020
2.89	.4981	.9981	.0019
2.90	.4981	.9981	.0019
2.91	.4982	.9982	.0018
2.92	.4982	.9982	.0018
2.93	.4983	.9983	.0017
2.94	.4984	.9984	.0016
2.95	.4984	.9984	.0016
2.96	.4985	.9985	.0015
2.97	.4985	.9985	.0015
2.98	.4986	.9986	.0014
2.99	.4986	.9986	.0014
3.00	.4987	.9987	.0013
3.01	.4987	.9987	.0013
3.02	.4987	.9987	.0013
3.03	.4988	.9988	.0012
3.04	.4988	.9988	.0012
3.05	.4989	.9989	.0011
3.06	.4989	.9989	.0011
3.07	.4989	.9989	.0011
3.08	.4990	.9990	.0010
3.09	.4990	.9990	.0010
3.10	.4990	.9990	.0010
3.11	.4991	.9991	.0009
3.12	.4991	.9991	.0009
3.13	.4991	.9991	.0009
3.14	.4992	.9992	.0008

3.15	.4992	.9992	.0008
3.16	.4992	.9992	.0008
3.17	.4992	.9992	.0008
3.18	.4993	.9993	.0007
3.19	.4993	.9993	.0007
3.20	.4993	.9993	.0007
3.21	.4993	.9993	.0007
3.22	.4994	.9994	.0006
3.23	.4994	.9994	.0006
3.24	.4994	.9994	.0006
3.30	.4995	.9995	.0005
3.40	.4997	.9997	.0003
3.50	.4998	.9998	.0002
3.60	.4998	.9998	.0002
3.70	.4999	.9999	.0001

Acerca del autor

Nació el 08 de julio de 1973, en el municipio de Atotonilco de Tula, Hidalgo, México. En 1992, una vez finalizada la educación media superior, presentó su examen de admisión en la Universidad Nacional Autónoma de México donde fue aceptado en la Licenciatura en Trabajo Social en la Escuela Nacional de Trabajo Social (ENTS). Durante su formación obtuvo reconocimientos y distinciones por estar entre los tres primeros lugares de la carrera; posteriormente estudió la Especialización de Trabajo Social en el Sector Salud en la ENTS-UNAM, y la Maestría en Administración de Organizaciones en la Facultad de Contaduría y Administración (FCA-UNAM). En 1997 fue invitado a trabajar en la División de Estudios de Posgrado de la ENTS como jefe de Sección Académica donde fue responsable del Programa Único de Especializaciones en Trabajo Social en Modelos de Intervención con Jóvenes, Mujeres y Adultos Mayores, además de formar parte del equipo académico para la creación del primer programa de Maestría en Trabajo Social (PMTS), que a la postre sería aprobado por el Consejo Universitario de la máxima casa de estudios de México el 12 de noviembre de 2004.

Como parte del proceso de actualización y búsqueda de nuevas formas de impartir la docencia cursó los diplomados en Estadística con SPSS; Aplicaciones de las Tecnologías de Información y Comunicación (TIC) para la Enseñanza; Diseño y Evaluación de Proyectos Sociales y Herramientas de Cómputo para Educación a Distancia. Además, cursó las certificaciones en Global Coaching; Coaching profesional con PNL y en Programación Neurolingüística.

En 2001, gracias al trabajo realizado en el área de investigación de la ENTS, fue invitado a formar parte de la planta académica de la ENTS-UNAM para impartir las materias de Estadística Aplicada a la Investigación Social I y Estadística Aplicada a la Investigación Social II, labor que ha desarrollado por más de 20 años formando a alumnos de licenciatura y posgrado, tanto en el sistema presencial como de educación abierta y a distancia. Además, ha impartido diversos cursos a profesionales de las ciencias sociales y de la salud en estadística descriptiva, estadística inferencial, estadística con SPSS, investigación y planeación, tanto en la UNAM como en universidades del interior de la República.

En 2008 recibió la invitación del Hospital Infantil de México Federico Gómez –primer Instituto Nacional de Salud de México, fundado el 30 de abril de 1943– para incorporarse como Jefe de Servicio en el Departamento de Bioestadística y Archivo Clínico, donde destacó por su sentido humano en el manejo de personal y en el análisis de información. Actualmente es Jefe de Servicio en la Subdirección de Seguimiento Programático y Diseño Organizacional de la Dirección de Planeación del mismo instituto, donde desarrolla actividades de planeación y programación, así como, creación y seguimiento de matrices de indicadores de resultados.

Es profesor titular de los módulos de estadística descriptiva e inferencial de los diplomados: Administración en Sistemas de Salud de la Secretaría de Salud en el Hospital General de México; Metodología de la Investigación Social de la Secretaría de Salud de la Ciudad de México en el Hospital de Especialidades

Belisario Domínguez; y en el diplomado de Administración en Servicios de Salud, Gerencia y Calidad de la Atención en el Colegio Nacional de Trabajadores Sociales (CONATS).

Su contribución a la docencia la realiza desde diversos ámbitos, mediante la publicación de material relacionado con el trabajo social y para el estudio de este, como fue el caso del cuadernillo *Muestreo: tamaño y tipología* y los libros de apoyo titulados *Estadística Aplicada a la Investigación Social I* y *Estadística Aplicada a la Investigación Social II*, ambos auspiciados por la ENTS-UNAM. Además, desde 2003 ha participado como dictaminador de proyectos de la ENTS-UNAM en el Instituto Nacional de Desarrollo Social, Programa de Coinversión Social, Secretaría de Desarrollo Social (hoy Secretaría del Bienestar).

El 21 de agosto de 2019 recibió por parte del Gobierno de México y de la Secretaría de Salud el Reconocimiento Nacional de Trabajo Social en el Sistema Nacional de Salud 2019 en la categoría de Docencia en Salud.

Estadística descriptiva. Dar sentido a los datos,
Ciro López Mendoza, publicado por Ediciones Comunicación Científica, S. A. de C. V.,
se terminó de imprimir en julio de 2025, Litográfica Ingramex, S. A. de C. V., Centeno 162-1,
Granjas Esmeralda, 09810, Ciudad de México. El tiraje fue de 25 ejemplares impresos y en versión digital
para acceso abierto en los formatos PDF, Epub y HTML.

Este libro tiene como objetivo fortalecer la formación de las personas profesionales de las ciencias sociales, en particular de quienes integran el universo del trabajo social. Su método es simple y lógico, se llama DPI (definición-procedimiento-interpretación) y su aporte principal es dar sentido a los datos. El *método* parte de generar un proceso de entendimiento y significado de una palabra (Definición), pasa por el cálculo de una prueba (Procedimiento) y termina cuando se da razón de ser y referencia a un valor obtenido (Interpretación). Tiene un enfoque constructivista alejado de tecnicismos y basado en experiencia docente, el cual ha llevado al alumno de la desesperación, angustia y la frustración del "no entiendo" a encontrar un camino de razonamiento propio, cuestionamiento del para qué sirve y, sobre todo, del cómo lo puede aplicar.

La presente obra comienza con la revisión de dos áreas primordiales, investigación y estadística; continúa con el estudio del método estadístico para profundizar en sus cinco etapas: recolectar, contar, presentar, describir y analizar. Finalmente aborda los elementos indispensables para el cálculo del tamaño de muestra, así como los tipos de muestreo generalmente utilizados.

Ciro López Mendoza nació en Atotonilco de Tula, Hidalgo, México. Trabajador social, especialista en el sector salud, cursó la Maestría en Administración de Organizaciones. En 2019 el Gobierno de México le otorga el Reconocimiento Nacional de Trabajo Social, categoría de Docencia en Salud.

DOI.ORG/10.52501/CC.283

COMUNICACIÓN CIENTÍFICA PUBLICACIONES ARBITRADAS
HUMANIDADES, SOCIALES Y CIENCIAS
www.comunicacion-cientifica.com

ISBN 978-607-2628-61-8