

**PROGRAMA DE CAPACITACIÓN Y FORMACIÓN
PROFESIONAL EN DERECHOS HUMANOS**

Fase de formación especializada

PROGRAMA DE ESTUDIO

**El empleo de la estadística
aplicada a la investigación
en derechos humanos**

DIRECTORIO INSTITUCIONAL

PRESIDENTE

Luis González Placencia

CONSEJO

Mercedes Barquet Montané
José Antonio Caballero Juárez
José Luis Caballero Ochoa
Miguel Carbonell Sánchez
Denise Dresser Guerra
Manuel Eduardo Fuentes Muñiz
Patricia Galeana Herrera
Mónica González Contró
Nashieli Ramírez Hernández
José Woldenberg Karakowsky

VISITADURÍAS GENERALES

Primera Mario Ernesto Patrón Sánchez
Segunda Rosalinda Salinas Durán
Tercera José Antonio Guevara Bermúdez
Cuarta Guadalupe Ángela Cabrera Ramírez
Quinta Luis Jiménez Bueno

CONTRALORÍA INTERNA

Rosa María Cruz Lesbros

SECRETARÍAS

Ejecutiva José Luis Gutiérrez Espíndola
**Promoción de los Derechos Humanos
e Incidencia en Políticas Públicas** Gerardo Sauri Suárez

CONSULTORÍA GENERAL JURÍDICA

Fernando Francisco Coronado Franco

DIRECCIONES GENERALES

Quejas y Orientación Alfonso García Castillo*
Administración Irma Andrade Herrera
Comunicación por los Derechos Humanos Daniel Robles Vázquez
Educación por los Derechos Humanos Paz Echeñique Pascal

DIRECCIÓN EJECUTIVA DE SEGUIMIENTO

Montserrat Matilde Rizo Rodríguez

CENTRO DE INVESTIGACIÓN APLICADA EN DERECHOS HUMANOS

Ricardo A. Ortega Soriano

SECRETARÍA PARTICULAR DE LA PRESIDENCIA

Gabriela Gutiérrez Ruz

COORDINACIONES

Asesores María José Morales García
Interlocución Institucional y Legislativa Soledad Guadalupe López Acosta
Tecnologías de Información y Comunicación Rodolfo Torres Velázquez
Servicios Médicos y Psicológicos Sergio Rivera Cruz*
Servicio Profesional en Derechos Humanos Mónica Martínez de la Peña

* Encargado de despacho

**PROGRAMA DE CAPACITACIÓN Y FORMACIÓN
PROFESIONAL EN DERECHOS HUMANOS**

Fase de formación especializada

PROGRAMA DE ESTUDIO

**El empleo de la estadística
aplicada a la investigación
en derechos humanos**



CONTENIDOS: Marisol Luna Contreras.

COORDINACIÓN DE CONTENIDOS: Mónica Martínez de la Peña, coordinadora del Servicio Profesional en Derechos Humanos; Rossana Ramírez Dagio, subdirectora de Formación Profesional, y Héctor Rosales Zarco, jefe de Departamento de Contenidos.

COORDINACIÓN ACADÉMICA DE CONTENIDOS: Jorge Peláez Padilla, profesor-investigador de la Academia de Derecho de la Universidad Autónoma de la Ciudad de México (UACM), integrante del Colectivo en Estudios Críticos del Derecho (RADAR); Carlos María Pelayo Moller, doctorando en el Instituto de Investigaciones Jurídicas de la Universidad Nacional Autónoma de México (UNAM); Mylai Burgos Matamoros, profesora de la Facultad de Derecho (UNAM/UACM), y Guillermo E. Estrada Adán, profesor de tiempo completo en la Facultad de Derecho (UNAM) y profesor en El Colegio de México.

EDITOR RESPONSABLE: Alberto Nava Cortez. CUIDADO DE LA EDICIÓN: Bárbara Lara Ramírez. DISEÑO DE PORTADA: Maru Lucero. DISEÑO Y FORMACIÓN: Maru Lucero y Analaura Galindo. CORRECCIÓN DE ESTILO: Solar, Servicios Editoriales, S. A. de C. V. LECTURA DE PLANAS: Karina Rosalía Flores Hernández.

Primera edición, 2012

D. R. © 2012, Comisión de Derechos Humanos del Distrito Federal
Av. Universidad 1449, col. Florida, pueblo de Axotla,
del. Álvaro Obregón, 01030 México, D. F.

www.cd hdf.org.mx

Ejemplar de distribución gratuita, prohibida su venta.

Se autoriza la reproducción total o parcial de la presente publicación siempre y cuando se cite la fuente.

El empleo de la estadística aplicada a la investigación en derechos humanos*

* Marisol Luna Contreras. Profesora-investigadora de la Facultad Latinoamericana de Ciencias Sociales (Flacso), sede México; actuaria por la Universidad Nacional Autónoma de México (UNAM) y maestra en Población por Flacso-México. Cuenta con la especialidad en Estadística aplicada por el Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas (IIMAS-UNAM). Sus principales temas de investigación son la mortalidad materna, el uso y la satisfacción con la salud materna y el gasto en salud y educación en los hogares. Todas las opiniones presentadas en este texto son responsabilidad exclusiva de la autora.

ÍNDICE

El empleo de la estadística aplicada a la investigación en derechos humanos

Marisol Luna Contreras

Presentación.	7
Módulo I. ¿Qué utilidad tienen las investigaciones cuantitativas de derechos humanos?	9
Módulo II. ¿Cuáles son los componentes básicos de una investigación estadística?	10
Módulo III. Algunos ejercicios de estadística descriptiva (la tabla de frecuencia).	12
Módulo IV. Algunos ejercicios de estadística descriptiva (medidas descriptivas)	14
Módulo V. Tablas de contingencia.	16
Bibliografía	18
Ejercicios	27
Autoevaluación	27
Clave de respuestas	27

PRESENTACIÓN

El Servicio Profesional en Derechos Humanos (SPDH) fue creado en 2005 con el propósito de responder a la demanda de especialización en el trabajo que desempeña la Comisión de Derechos Humanos del Distrito Federal (CDHDF) como organismo público autónomo.

A partir de la creación del SPDH, la Comisión ha realizado un esfuerzo significativo para la consolidación y el desarrollo de los procesos de ingreso y ascenso, capacitación, formación y evaluación anual del desempeño.

La presente guía ha sido elaborada a partir del trabajo conjunto de la Coordinación del Servicio Profesional en Derechos Humanos de la CDHDF con las y los coordinadores académicos de las áreas modulares del SPDH y, desde luego, las y los autores de los contenidos del presente material, quienes serán las y los responsables de la impartición de los cursos presenciales a las y los integrantes del Servicio.

La Fase de formación especializada, a la que pertenece esta guía, se integra por contenidos tendientes a satisfacer las necesidades específicas de formación de cada una de las áreas sustantivas que integran la CDHDF: defensa, promoción y estudio.

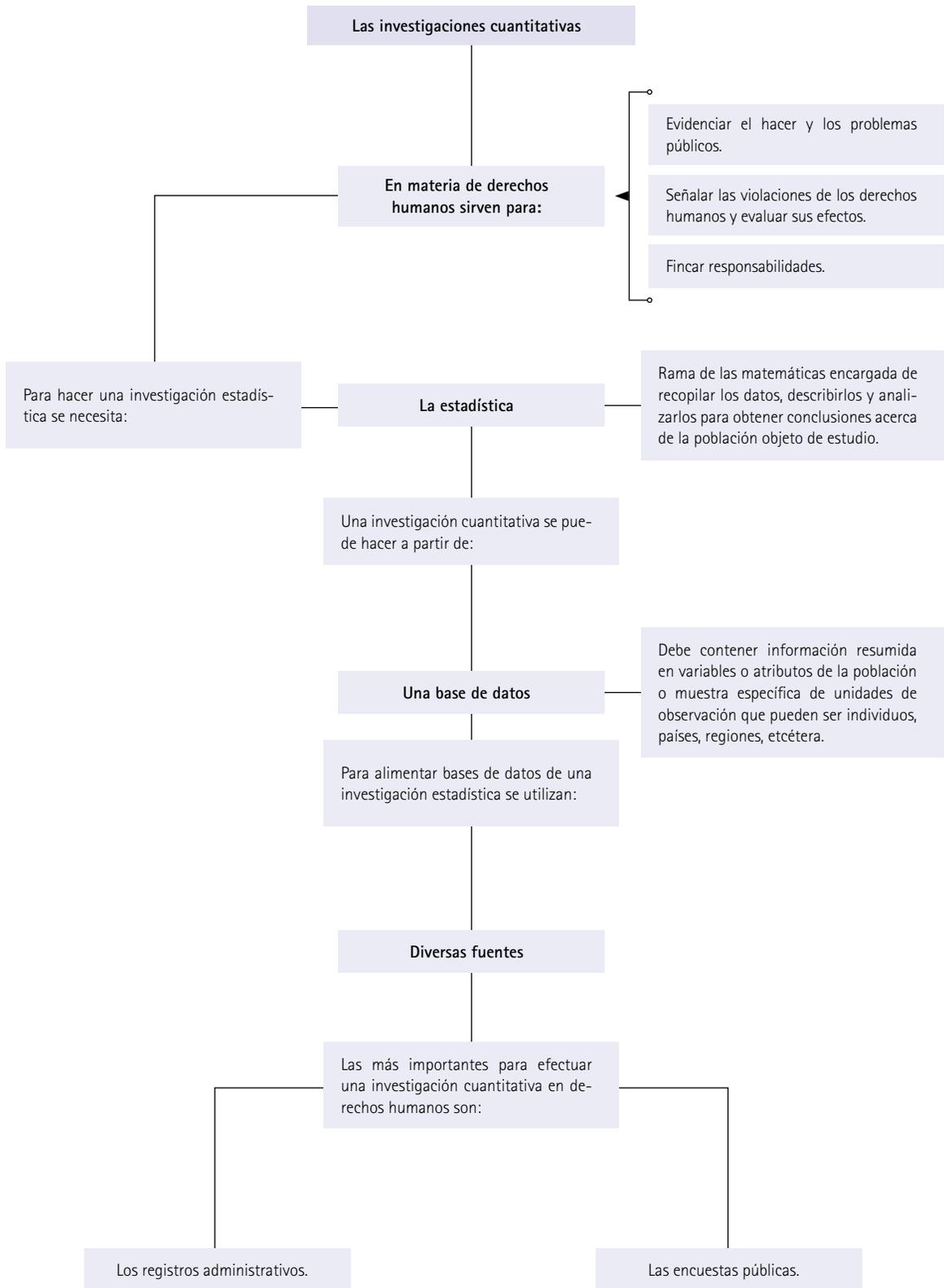
Los materiales que conforman este documento han sido contruidos con la expectativa de reflexionar sobre la aplicación práctica de los diversos aspectos que conforman la materia de los derechos humanos en el área de **estudio**, y comprende contenidos en torno a la metodología aplicada a los derechos humanos; el empleo de la estadística aplicada a la investigación de los derechos humanos y respecto de la metodología para la identificación, elaboración y evaluación de indicadores de políticas públicas en materia de derechos humanos.

Esta segunda edición de la Fase de formación especializada constituye un paso importante hacia la consolidación de una metodología *ad hoc* de enseñanza de los derechos humanos dirigida a las y los servidores públicos de los organismos públicos autónomos encargados de protegerlos, pues sin duda está encaminada a fortalecer la defensa y la promoción de los derechos humanos en nuestro país.

Coordinación del Servicio Profesional en Derechos Humanos

MÓDULO I.

**¿Qué utilidad tienen
las investigaciones cuantitativas
de derechos humanos?**



Los derechos y las garantías son el objeto más deseado para quienes habitan en un régimen democrático.

La defensa de los derechos humanos se convirtió en una necesidad y en una exigencia de la vida moderna para evitar las atrocidades del abuso del poder, en cualquiera de sus acepciones (política, económica, etc.), que han padecido millones de personas en todo el mundo y a través del tiempo.

Acontecimientos tan desastrosos como la segunda Guerra Mundial dejaron en la historia colectiva lecciones que obligan a reconsiderar el papel de los gobiernos y los ciudadanos en la construcción social.

A la fecha, se han intensificado los esfuerzos globales para no repetir tan lamentables sucesos. No sólo se trata de señalar que las violaciones a los derechos ocurren, sino de evaluar sus efectos, así como de identificar patrones sistemáticos que brinden información respecto a los derechos que se vulneran, la frecuencia con que ocurren estas violaciones, quiénes son las víctimas y quiénes los victimarios, entre otros factores.

Guardar registros de la información brinda la posibilidad de mostrar evidencias para hacer públicos los problemas y fincar responsabilidades a quienes atentan contra la estabilidad social.

Durante varios años, diversos gobiernos en el mundo, en particular los latinoamericanos, han atravesado por intensos procesos de reformas institucionales que han afectado positiva y negativamente las reglas del juego democrático: "En México, delinear las reglas para un efectivo imperio de la ley ha sido un largo e inacabado camino en busca de las instituciones que, en concordancia con la creación de mayores espacios ciudadanos, constituyan restricciones a los abusos y vicios gubernamentales".¹

Muchas reformas institucionales orientadas al fortalecimiento de los regímenes democráticos han intentado ofrecer mecanismos de acción para la expresión de la ciudadanía a fin de hacer factibles los derechos y garantías o, en su caso, conformar acciones para la transformación de la vida social basadas en la organización y defensa de estos derechos por parte de los ciudadanos, organizaciones sociales —públicas y privadas—, e incluso la creación de organismos autónomos que conforman una nueva configuración del quehacer gubernamental y ciudadano.

Durante este tiempo de intensas transformaciones sociales surgen diversas interrogantes: ¿cómo evaluar que las reformas emprendidas hayan tenido los efectos esperados? y ¿cómo comprobar que los cambios puestos en marcha en materia de defensa y procuración de derechos humanos hayan sido los adecuados para el conjunto social?

Para responder estas preguntas se deben elaborar mediciones e indicadores que permitan generar datos comparables y así realizar un análisis cuantitativo de éstos. Sin embargo, ¿es posible medir conceptos como *dignidad humana*? La respuesta es sí, aunque la tarea de cuantificarlos resulte todo un reto.

La base fundamental para analizar con información cuantitativa son los datos. Gran parte se obtiene de los registros administrativos y, en algunos casos, de encuestas elaboradas para recabar información *ad hoc*.

En México, el trabajo realizado por las comisiones de derechos humanos consiste en recomendaciones sustentadas jurídicamente mediante una diversidad de razonamientos respecto de un solo tema. Aun cuando la información es extensa y difícil de delimitar, una de las bases para realizar análisis cuantitativo radica en la capacidad para ordenar los datos. Adicionalmente, el uso de encuestas es cada vez más frecuente en el ámbito

¹ Para una visión amplia respecto de los procesos de reforma institucional que han emprendido diversos países en la larga marcha hacia la consolidación de su democracia, véase Andreas Schedler, Larry Diamond y Marc Plattner, *The Self-restraining State: Power and Accountability in New Democracies*, Boulder, Lynn Rienr, 1999.

internacional de los derechos humanos,² aunque en México aún hace falta elaborar más ejercicios con este propósito.

Por otra parte, si al contexto de la defensa de los derechos humanos se añaden los cambios producidos por la construcción de una sociedad de la información, se entenderá la trascendencia del análisis de datos y la información cuantitativa, donde ésta adquiere un papel esencial para establecer procesos de relaciones sociales que van desde el uso de nuevas formas de comunicación virtual y en tiempo real, hasta la generación de derechos de información pública y la posibilidad de guardar registros de información relevante para los ciudadanos.

No se trata sólo de exigir cuentas y pedir información a los gobiernos para saber lo que tienen en sus registros y lo que se suelen reservar. El punto más relevante de la solicitud de información es su análisis y la observación de posibles patrones de regularidad en el comportamiento gubernamental, así como identificar si se actúa de manera sistemática y cuáles son los efectos, negativos o positivos, en la formación de espacios públicos con una visión amplia del respeto a las garantías y derechos de los individuos. Las herramientas estadísticas ayudan a identificar esos patrones y a tratar de encontrar soluciones adecuadas con base en las evidencias halladas.

Uno de los objetivos de esta guía es esclarecer algunos conceptos del análisis cuantitativo de datos, como muestra, media, mediana y moda, entre otros, así como sus aplicaciones prácticas al análisis relacionado con el estudio de los derechos humanos.

Cabe señalar que la *estadística* es una herramienta que va de la mano de cuestiones concretas de metodología cuantitativa. Esta guía no desarrollará temas de metodología, aunque algunos de sus conceptos están más que relacionados con ésta.

¿Cuáles son las fuentes más importantes para una investigación cuantitativa en derechos humanos?

La fuente más importante de información en México para el tema de los derechos humanos son los registros administrativos. De hecho, gran parte de las actividades de las comisiones de derechos humanos se basa en el

² Aún son pocos los libros que relacionan directamente la estadística y los derechos humanos. Para profundizar en la historia y en el contexto en que se comienzan a utilizar los datos, la información y la estadística en análisis de derechos humanos, así como en la importancia de su sistematización para introducirse en la aplicación de técnicas analíticas, véanse Jana Asher *et al.*, *Statistical Methods for Human Rights*, Nueva York, Springer, 2007; y Thomas B. Jabine, *Human Rights and Statistics, Getting the Record Straight*, Filadelfia, University of Pennsylvania Press, 1992.

desarrollo de recomendaciones sustentadas en un extenso análisis jurídico. Sintetizar y sistematizar esa información es uno de los retos más grandes de administración de la información para el análisis de los derechos humanos.³

Uno de los primeros pasos para el análisis cuantitativo es contar con la información que conformará las variables de las bases de datos, lo que ayudará a establecer características y descripciones de los objetivos fundamentales de análisis que, en general, serán las personas que acuden a los servicios de la Comisión de Derechos Humanos del Distrito Federal (CDHDF), así como los tipos de problemas que se resuelven.

Por otra parte, algunas encuestas tratan asuntos relacionados con los derechos humanos, como la Encuesta Nacional de Cultura Política, la Encuesta Nacional sobre Discriminación en México, la Encuesta Nacional de la Juventud y la Encuesta Nacional sobre Violencia contra las Mujeres, entre otras, las cuales incluyen preguntas sobre el papel del gobierno en la atención a problemáticas sociales, así como acerca de la percepción de las y los mexicanos en torno al cumplimiento de los derechos en el país. Adicionalmente, encuestas como Latinobarómetro captan información respecto al cumplimiento de libertades y garantías en distintos países latinoamericanos, incluido México.

¿Qué es la estadística?

La estadística es la rama de las matemáticas encargada de recopilar datos, describirlos y analizarlos para obtener conclusiones acerca de la población objeto de estudio.

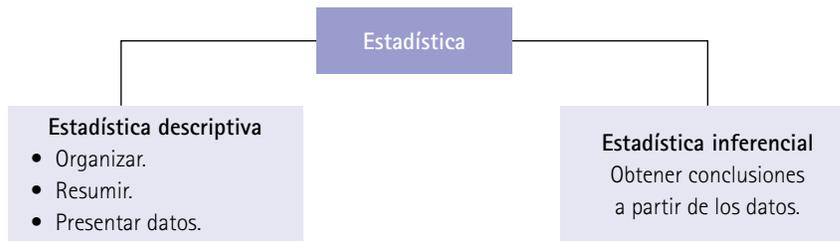
Básicamente, la estadística tiene dos ramas:

- 1) La estadística descriptiva, relacionada con la descripción y clasificación de los datos.
- 2) La estadística inferencial, que permite obtener conclusiones a partir de los datos.

Esta guía se centrará en la estadística descriptiva.

³ Véase Thomas Jabine, *op. cit.*

Figura 1. Ramas de la estadística



¿Qué características tiene la investigación estadística?

La investigación estadística descansa en el análisis cuantitativo de datos. Por ello se requiere una base de datos que contenga información resumida en variables o atributos de una población o muestra específica, o de unidades de observación, como individuos, países, regiones, etc. La información sintetizada y captada en variables permitirá elaborar asociaciones y relaciones entre los diversos atributos de los objetos de estudio, con la finalidad de explicar los patrones de lo que se denomina *comportamiento* de un fenómeno determinado.

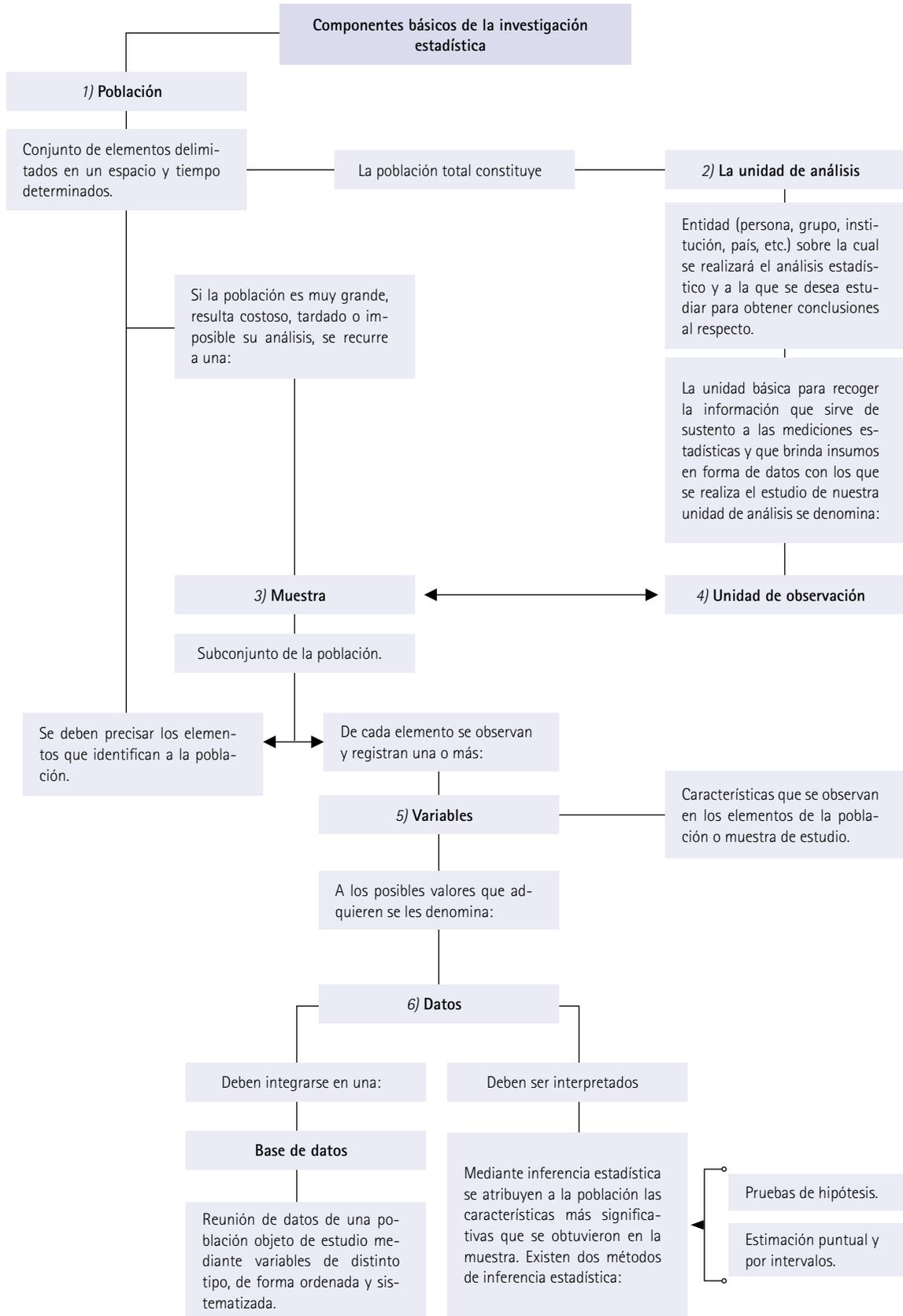
Un ejemplo muy sencillo es el siguiente: para conocer y describir algunas características de las personas que acudieron a presentar una queja ante la CDHDF durante 2011, es necesaria información básica, como su sexo y edad, lo que permitirá elaborar una breve descripción de la proporción de mujeres que asistió, respecto del total, e incluso saber si las edades varían respecto a las de los hombres que se presentaron durante el mismo año.⁴

Asimismo, una investigación de corte estadístico requiere también información sobre un grupo de población de interés en un momento determinado, que lo hace sujeto de análisis comparable. En el ejemplo, los datos estarán circunscritos a las personas que interpusieron una queja (gestión) ante la Comisión durante 2011, la cual resulta de interés para el análisis.

⁴ Para el caso de la CDHDF, la gestión de casos puede ser de oficio o iniciada por algún individuo o grupo, por lo que el sexo de las personas que intervienen en el proceso no necesariamente será hombre o mujer. Por ello se aclara que en los ejemplos en que se usa esta característica, se recurre a todas las categorías que considera la Comisión.

MÓDULO II.

¿Cuáles son los componentes básicos de una investigación estadística?



Conceptos de población y muestra

La investigación que utilice la estadística como herramienta para el análisis de datos debe distinguir e identificar la población (objeto de la investigación), la muestra, la unidad de análisis y medición, las variables que se incluirán en el análisis y el tipo de escala de dichas variables.

Población: conjunto de elementos delimitados en un espacio y tiempo determinados. En una investigación, la población se define al especificar los elementos de nuestro interés y las características que deben tener para formar parte del estudio.

Muestra: subconjunto de la población que debe contener un número determinado de elementos para ser representativo de ésta.

Para la investigación estadística es fundamental especificar claramente la población objeto de estudio, de manera que se distingan los elementos que forman parte de ella y los que no lo son. Ejemplos:

Ejemplo 1

Se trabajará con las gestiones recibidas durante 2010 en la oficina central de la CDHDF y que hayan concluido. En este caso, la población de interés debe cumplir tres características: *a)* ser gestiones recibidas en la oficina central; *b)* haber sido recibidas durante 2010; *c)* tener estatus de gestión concluida. Obsérvese en el cuadro 1 cómo se eligen los elementos de la población.

Cuadro 1. Ejemplo de delimitación de la población objeto de estudio

Unidad donde fue recibida	Año de recepción	Estatus de la gestión	¿Es parte de la población?	Explicación
Unidad Norte	2010	Concluida	✘	No es parte de la población, no cumple con dos de las tres características: fue recibida en otra sede y en otro año.
Oficina central	2009	No concluida	✘	Aunque fue recibida en la sede de interés, no es parte de la población, pues se recibió en otro año y no ha sido concluida. Por tanto, no cumple con las tres características.
Oficina central	2009	Concluida	✘	Aunque fue recibida en la sede de interés y ha concluido, la gestión es de 2009 y sólo interesan las recibidas durante 2010. Por tanto, no es parte de la población.
Oficina central	2010	Concluida	✓	Sí es parte de la población, ya que cumple con los tres criterios: fue recibida en la oficina central en 2010 y ha concluido.

Fuente: elaboración propia.

Ejemplo 2

En este caso interesan las gestiones recibidas vía telefónica durante 2009 para orientación verbal. La población de interés serán todas aquellas gestiones que cumplan con tres características: *a)* haber sido recibidas en 2009; *b)* haber sido atendidas por vía telefónica; *c)* el servicio debió ser de orientación verbal. Por lo tanto, no interesan ni la sede de recepción ni su estatus de gestión.

Cuadro 2. Ejemplo de delimitación de la población objeto de estudio

Unidad donde fue recibida	Año de recepción	Estatus de la gestión	Vía de entrada	Servicio	¿Es parte de la población?	Explicación
Unidad Norte	2010	Concluida	Correo electrónico	Orientación verbal		No es parte de la población; no cumple con dos de las tres características: fue recibida en otro año y llegó por correo electrónico.
Oficina central	2009	Concluida	Telefónica	Suplencia de queja		Aunque fue recibida en 2009 por vía telefónica, el servicio requerido fue de suplencia de queja, por lo tanto, no es parte de la población.

Unidad donde fue recibida	Año de recepción	Estatus de la gestión	Vía de entrada	Servicio	¿Es parte de la población?	Explicación
Unidad Poniente	2009	Concluida	Telefónica	Orientación verbal		Si es parte de la población, cumple con los tres criterios: fue recibida en 2009 por vía telefónica y fue para orientación verbal. Aquí no es importante la oficina de recepción ni su estatus de gestión.
Oficina central	2009	Concluida	Telefónica	Orientación verbal		Si es parte de la población, cumple con los tres criterios: fue recibida en 2009 por vía telefónica y fue para orientación verbal. Aquí no interesa la oficina de recepción ni su estatus de gestión.

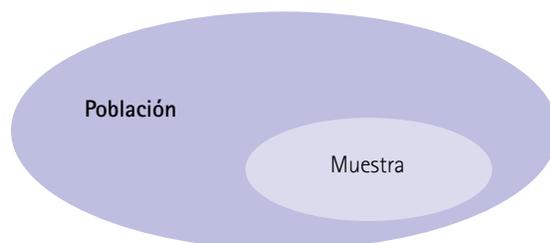
Fuente: elaboración propia.

Ejemplo 3

Se analizará el tipo de derecho que, presuntamente, se violó con mayor frecuencia en todas las gestiones concluidas y recibidas en las cinco unidades de atención de la CDHDF durante 2011. En este caso deben tomarse en cuenta los 47 238 registros, es decir, toda la información de la que se dispone. Como se observa, esta población es demasiado grande.

Cuando esto ocurre, examinar cada elemento sería demasiado costoso, tomaría demasiado tiempo y, en ocasiones, resultaría imposible. Para ahorrar tiempo y recursos económicos, así como para estudiar y comprender mejor algún fenómeno de interés, se recomienda elegir sólo algunos elementos de la población, los cuales formarán un subconjunto. A este subconjunto le llamaremos *muestra*. Obsérvese en la figura 2 cómo todos los elementos de la muestra son también elementos de la población.

Figura 2. Relación entre la población y la muestra



Fuente: elaboración propia.

Básicamente, las muestras se toman con dos tipos de métodos: muestreos probabilísticos y no probabilísticos. En el muestreo probabilístico cada elemento es susceptible de ser elegido, y su probabilidad conocida de ser seleccionado es distinta de cero. Si lo anterior no se cumple, entonces se trata de un muestreo no probabilístico.

En cada tipo de muestreo hay diferentes métodos para seleccionar los casos. La elección de un método u otro depende de los propósitos de la investigación, de la experiencia y el conocimiento previo del investigador acerca del problema por analizar y de los recursos económicos, de infraestructura y humanos con los que se cuenta. Los métodos para cada tipo de muestreo se exponen en el cuadro 3.⁵

Cuadro 3. Tipos de muestreo, métodos y características

Muestreo	Métodos	Características
Probabilístico	<ul style="list-style-type: none"> • Aleatorio. • Aleatorio estratificado. • Por conglomerados. • Combinaciones de los anteriores. 	<ul style="list-style-type: none"> • Cada elemento de la población objeto de estudio tiene una probabilidad conocida y distinta de cero. • La selección de la muestra utiliza la teoría de la probabilidad. • Permite una evaluación objetivo de los resultados y estimar su grado de confianza y de precisión.
No probabilístico	<ul style="list-style-type: none"> • De juicio. • Por conveniencia. • Por cuotas. • Bola de nieve. 	<ul style="list-style-type: none"> • La selección de los elementos de la muestra depende de los criterios del investigador. • No hay manera de estimar el grado de confianza y de precisión de los resultados.

Fuente: elaboración propia.

La unidad de análisis de la investigación estadística es la entidad (persona, grupo, institución, país, etc.) acerca de la cual se realizará el análisis estadístico y a la que se estudiará para sacar conclusiones al respecto. Debe establecerse en los objetivos y en la definición del problema de investigación. La unidad de análisis sintetiza e identifica la población en función de la cual se definirá el tamaño de la muestra. Por su parte, la unidad de observación es la que se toma como base para reunir la información que sustenta las mediciones estadísticas y brinda datos con los que se realizará el estudio de la unidad de análisis.⁶

Ejemplo 4

Se investigará si los derechos humanos de las mujeres que se encuentran en reclusión en el Distrito Federal están garantizados. En este caso, la unidad de análisis (población) son todas las mujeres recluidas tanto en el Centro Femenil de Readaptación Social de Santa Martha Acatitla como en el de Tepepan. Debido al tamaño de esa población, entrevistar a todas las reclusas sería costoso y tomaría mucho tiempo, por lo que se obtiene una muestra aleatoria de ellas. Por tanto, las mujeres seleccionadas en la muestra son la unidad de

⁵ Para más información sobre los tipos de muestreo, métodos, ventajas y desventajas, véanse Yvón Angulo Reyes, "Muestreo", en Haroldo Elorza Pérez-Tejeda, *Estadística para las ciencias sociales del comportamiento y de la salud*, 3ª ed., México, Cengage Learning, 2007, pp. 179-210; Juan Fernández Durán y Rubén Hernández, "Muestreo", en Víctor Aguirre *et al.*, *Fundamentos de probabilidad y estadística*, 2ª ed., México, Jit Press, 2007, pp. 1-32 (cap. 11).

⁶ Julio César Ángel Gutiérrez, *Estadística general aplicada*, Medellín, Fondo Editorial Universidad Eafit (col. Académica), 2007, p. 49.

observación de la investigación, a fin de que el análisis brinde conclusiones respecto de la población total.

Datos, variables y escalas de medición en estadística

Una vez delimitada la población y la unidad de observación, para responder a preguntas de interés se debe recolectar la información de cada uno de los elementos que la conforman. A estas características les denomina *variables*.

Las variables son cada una de las características que se observan en los elementos de la nuestra población o muestra de estudio, las cuales son susceptibles de ser cuantificadas o registradas. No necesariamente tienen un carácter numérico. En cada elemento se pueden observar y registrar una o más variables; a los posibles valores que toman se les llama *datos*.⁷ A la información recabada de toda la población se le llama *censo* o información censal; o si derivan de una muestra, información muestral.

Las variables se clasifican en cualitativas y cuantitativas (véase el cuadro 4).

Cuadro 4. Tipos de variables usadas en estadística, sus características y ejemplos

Tipos de variables	Características	Ejemplos		
Cualitativos	<ul style="list-style-type: none"> • Corresponden a cualidades o atributos de los elementos de análisis. • Tienen un número fijo de categorías. • Todos los elementos deben clasificarse sólo en una de las categorías. 	Variable Sexo Entidad de residencia Mes de nacimiento Tipo de horario Color de ojos Estado conyugal	Categorías Femenino, masculino. Aguascalientes, Baja California, Baja California Sur... Zacatecas. Enero, febrero, marzo... diciembre. Matutino, vespertino, mixto. Negro, café, verde, azul Soltero, casado, viudo, divorciado, separado.	
Cuantitativos	<ul style="list-style-type: none"> • Se obtienen mediante un proceso de conteo o de medición. • Cuando se obtienen por conteo, se les llama discretos. • Cuando son resultado de una medición son continuos. 	Variables Número de hijos Número de delitos Estatura de adultos Peso de los niños menores de un año Ingreso mensual Número de gestiones recibidas en un día	Valores que puede tomar 0, 1, 2... 0, 1, 2... [1.40, 2.50] [1,8] [0,∞] 0, 1, 2...	Tipo Discreta Discreta Continua Continua Continua Discreta

Fuente: elaboración propia.

⁷ Víctor Aguirre y Begoña Artaloitia, "Análisis exploratorio de datos", en Víctor Aguirre *et al.*, *Fundamentos de probabilidad y estadística*, 2ª ed., México, Jit Press, 2007, pp. 1-68 (cap. 1).

Además de ser cuantitativa o cualitativa, una variable también se puede clasificar en una escala de medición. Dichas escalas se dividen en nominales, ordinales y de intervalo o razón.

Las variables cualitativas sólo pueden tener atributos de escalas nominales y ordinales; una variable cuantitativa puede ser de intervalo o de razón. A continuación mencionamos algunas características y ejemplos de cada tipo de escala.

Escala nominal: son datos que sólo pueden ser clasificados en categorías. No es posible establecer una relación de orden entre ellas, solamente se puede decir si el elemento pertenece a la categoría o no, o si tiene o no el atributo. Además, todos y cada uno de los elementos deben ser clasificados en una y sólo una categoría.

Ejemplos:

- 1) Sexo.
- 2) Estado conyugal.
- 3) Color de ojos.
- 4) Si tiene o no hijos.
- 5) Carácter de la gestión.
- 6) Nacionalidad.
- 7) Tipo de servicio de la gestión.
- 8) Vía de entrada de la gestión.

Escala ordinal: los valores de la variable se agrupan en categorías que permiten establecer una relación de orden entre ellas, según el grado de posesión del atributo que tengan, si la cantidad que posee un elemento es mayor o menor que la cantidad que posee otro. Con las variables de este tipo de escala no es posible hacer ninguna de las operaciones aritméticas básicas: suma, resta, multiplicación y división.

Ejemplos:

- 1) Nivel socioeconómico.
- 2) Orden de llamadas en un día.
- 3) Identificador de gestión.
- 4) Identificador de persona.

Escala de intervalo: cuando un elemento posee mayor o menor cantidad de una variable que otro, también es posible determinar la diferencia de magnitud entre dos elementos medidos. Para hacerlo, es necesario tener una medida origen o *cero*, establecida a conveniencia por el investigador u otra ya definida. Este tipo de variables permiten las operaciones de suma o resta.

Ejemplos:

- 1) Temperatura en °C.
- 2) El resultado de la suma de las caras al lanzar dos dados.
- 3) Edad en años cumplidos.
- 4) El tiempo.

Escala de razón: entre este tipo de variables sí existe el cero e indica la ausencia total de la variable. Además, el cero no es un punto arbitrario de la escala, sino que está fijo. Estas variables permiten las cuatro operaciones aritméticas básicas y, además, realizar comparaciones entre las proporciones o razones.

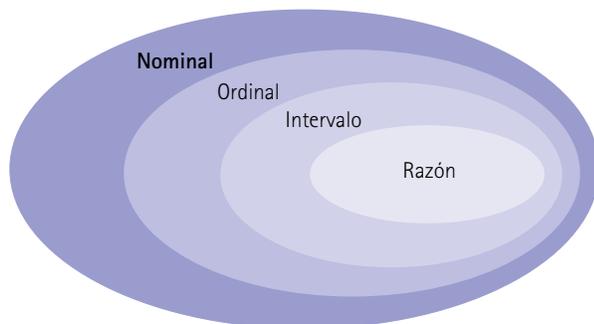
Ejemplos:

- 1) Peso.
- 2) Talla (estatura).
- 3) Ingreso.
- 4) Edad.
- 5) Número de hijos.
- 6) Tiempo de espera en una sala de urgencia.

Debemos destacar que las diferentes escalas de medición son acumulativas, es decir, la escala ordinal posee las propiedades de una nominal, además del orden entre categorías; por lo tanto, la escala ordinal es más fuerte que la nominal, ya que sus datos poseen más información. Asimismo, la escala de intervalo es a su vez nominal y ordinal, mientras que la escala de razón tiene todas las propiedades de la escala de intervalo. Siempre es posible transformar datos que se encuentran en cierta escala a una más débil, simplemente ignorando la información extra que contiene.⁸

En la figura 3 se observa la relación entre las diferentes escalas de medición.

Figura 3. Relación entre las escalas de medición utilizadas en estadística



Fuente: elaboración propia.

⁸ *Ibidem*, pp. 1-5.

Una variable puede ser transformada en otro tipo de escala siempre que sea de rango inferior a la utilizada inicialmente. Como se observa en la figura 3, una variable de razón se puede convertir en una de intervalo, ordinal o nominal; y una de intervalo, en ordinal o nominal. De la misma manera, una variable ordinal se puede transformar en una nominal. Esto se hace mediante la agrupación de categorías.

Ejemplo 5

La variable "número de hijos de una persona" es de razón y toma valores de 0, 1, 2, 3, etc. Formamos dos grupos: las personas que no tienen hijos las ponemos en el grupo A; las que sí tienen hijos, en el grupo B. También les podríamos asignar un valor numérico. Por ejemplo, al grupo A le adjudicamos el valor 1, y a los del B, el 2. De esta manera, la nueva variable, formada por los dos grupos, es una variable nominal, ya que no le podemos asignar valor numérico al hecho de tener o no hijos, simplemente son categorías que clasifican (véase el cuadro 5).

Cuadro 5. Ejemplo de transformación de una variable de razón a una variable nominal

Variable de razón Número de hijos	Variable nueva de tipo nominal ¿Tiene o no tiene hijos?
0	No: A = 1
1	Sí: B = 2
2	
...	
n	

Fuente: elaboración propia.

Este proceso de transformación de las variables es práctico; sin embargo, se pierde información. En el ejemplo 5, al observar el registro de una persona con el valor 2 o la letra B de la nueva variable, es decir, la transformada, sólo sabemos que sí tiene hijos, pero no cuántos. Por lo tanto, se recomienda reunir siempre la información de las variables con la mayor precisión y detalle y, en la medida de lo posible, en escala de razón, ya que después —si fuera necesario— será más fácil transformarlas a la escala que nos convenga para el tipo de análisis estadístico que deseemos realizar.⁹

⁹ Miguel A. Martínez González y Jorkin de Irala Estévez, "Procedimientos descriptivos", en *Bioestadística amigable*, 2ª ed., Madrid, Díaz de Santos, 2006, pp. 17 y 18.

La estadística descriptiva y la estadística inferencial dentro de la investigación estadística

Una vez delimitadas la población, la muestra, la unidad de observación, las variables que se utilizarán, así como la manera de medirlas, el siguiente paso es analizar e interpretar los datos. Clasificar de manera correcta las variables de la investigación servirá para determinar los diferentes modelos matemáticos y el tratamiento que se les dará.¹⁰

Analizar e interpretar los datos requiere de técnicas de la estadística descriptiva y de la estadística inferencial:

- Las técnicas de la estadística descriptiva, que se verán con detalle en los módulos IV y V, se utilizan para el análisis descriptivo de la información con el objetivo de recabar, clasificar, resumir y analizar las características de la población o muestra para luego deducir conclusiones sobre su estructura y composición.¹¹
- Si se usa información de una muestra aleatoria, habrá que utilizar herramientas estadísticas para sacar conclusiones acerca de la población a partir de dicha muestra; a esto se le conoce como *inferencia estadística*. La inferencia estadística¹² atribuye a la población las características más significativas obtenidas en la muestra.¹³ Existen dos métodos de inferencia estadística:
 - *Estimación puntual y por intervalos*: se proponen estimaciones de los valores de la característica de la población que deseamos investigar¹⁴ usando la información de la muestra. Siempre están sujetas a error (la diferencia entre el valor del parámetro de la población y el valor estadístico de la muestra). La probabilidad de cometer este error es calculable.
 - *Pruebas de hipótesis*: en este método hay que establecer una hipótesis estadística respecto al valor que esperamos de la característica de la población que investigamos, la cual se evaluará con información generada a partir de la muestra.¹⁵

Por lo tanto, una hipótesis estadística es una sentencia sobre la naturaleza de una población y, por lo general, se formula a partir de determinada caracte-

¹⁰ *Ibidem*, p. 13.

¹¹ María Teresa Icart Isern et al., *Elaboración y presentación de un proyecto de investigación y una tesis*, Barcelona, Universitat de Barcelona (Salud pública), p. 72.

¹² Esta guía es una introducción a las técnicas de la estadística descriptiva. Respecto a las técnicas de estadística inferencial véanse los capítulos del 9 al 13 de Kenneth D. Hopkins et al., *Estadística básica para las ciencias sociales y del comportamiento*, 3ª ed., México, Prentice Hall, 1997, pp. 143-270; y Haroldo Elorza Pérez-Tejeda, *op. cit.*, pp. 179-210.

¹³ Howard B. Christensen, *Estadística paso a paso*, 3ª ed., México, Trillas, 2008, pp. 20-25.

¹⁴ En estadística, a esta característica se le llama comúnmente *parámetro*.

¹⁵ Haroldo Elorza Pérez-Tejeda, *Estadística para las ciencias sociales...*, *op. cit.*, p. 308.

rística de la población. Para contrastar una hipótesis estadística se decide si parece congruente con los datos de la muestra. A esta hipótesis se le conoce como *hipótesis nula* y se indica con H_0 . Para contrastar la hipótesis nula se recurre a la hipótesis alternativa y se denota como H_1 . Así pues, la H_0 se rechaza si parece incongruente con los datos de la muestra; en caso contrario, se acepta como válida.¹⁶

La diferencia entre el contraste de hipótesis y la estimación puntual y por intervalos consiste en que la primera establece una hipótesis acerca del parámetro antes de realizar el estudio; con fundamento en el resultado del estadístico muestral, se rechaza o no dicha hipótesis. En cambio, en la estimación por intervalo se consideran todos los posibles valores del parámetro.¹⁷

Ejemplo 6

Un político mexicano comenta a los medios de comunicación que la edad promedio de las mujeres asesinadas en Ciudad de Juárez entre 2009 y 2010 es de 30 años. Las personas de la Comisión Estatal de Derechos Humanos consideran que la afirmación no es correcta y, con base en una muestra aleatoria de la información disponible en sus registros de los años correspondientes, desean refutar la aseveración. Por lo tanto, las hipótesis estadísticas que plantean las personas de la Comisión son:

H_0 : el promedio de edad de las mujeres asesinadas entre 2009 y 2010 en Ciudad Juárez es de 30 años.

H_1 : el promedio de edad de las mujeres asesinadas entre 2009 y 2010 en Ciudad Juárez es distinto de 30 años.

El análisis de la población o muestra seleccionada mediante la estadística descriptiva es un paso indispensable y necesario en cualquier investigación cuantitativa, independientemente de que se utilicen o no técnicas de estadística inferencial.

La selección y aplicación de técnicas estadísticas debe ser parte de la planeación de la investigación. Por lo tanto, antes de emplearlas, se deben llevar a cabo los pasos mencionados en la fase de inducción, curso III, módulo III, donde se exponen, entre otros temas, la formulación de objetivos y la elaboración de hipótesis de investigación.¹⁸

¹⁶ Sheldon Ross, *Introducción a la estadística*, Barcelona, Reverté, 2007, pp. 385-387.

¹⁷ *Ibidem*, p. 308.

¹⁸ Mylai Burgos Matamoros, "Fundamentos básicos para la metodología de la investigación aplicada", en *Programa de capacitación y formación profesional en derechos humanos*, México, Comisión de Derechos Humanos del Distrito Federal, 2011, pp. 187-195.

¿Qué es una base de datos?

Una base de datos consiste en la recopilación y ordenamiento de los atributos (datos) asociados a los individuos de la población objetivo. Las características que se medirán son las variables y, como vimos, éstas pueden ser cualitativas o cuantitativas.

Las variables que se consideran para generar una base de datos dependen completa y exclusivamente de los intereses de quien hará uso de ella, así como del medio por el cual reúna la información de su interés. Por esta razón, no es requisito que una base de datos contenga todos los tipos de variables a la vez, sino sólo las que serán de utilidad, ya sea para estudio, comprensión o mejora de los servicios ofrecidos. Al reunir los datos de una población objeto de estudio a través de variables de distinto tipo, ordenada y sistematizadamente, se elabora (confecciona, construye) una base de datos. Para juntar la información se puede recurrir a registros administrativos, encuestas, ejercicios *ad hoc* de levantamiento de información, etcétera.

Es muy importante que los resultados de una encuesta o investigación se presenten de manera clara y concisa, ya que normalmente recopilan gran cantidad de información y es preferible que se puedan leer fácil y rápidamente para tener una idea precisa de lo que representan en conjunto, más allá del comportamiento individual.

Ejemplo de una base de datos es la información que el Sistema Integral de Gestión de Información (Siigesi) de la CDHDF registra de las gestiones que realizan durante un periodo determinado. Para los fines de esta guía, se utilizará una base de datos de 47 328 gestiones recibidas durante 2011 que se encuentra en un archivo de Excel llamado "Tabla de servicios", y de la cual se pueden obtener tablas dinámicas. Una tabla dinámica es un recurso de Excel que permite resumir, organizar y agrupar gran cantidad de datos en un cuadro. Se le llama *dinámica* porque se puede modificar, reorganizar, mostrar u ocultar información de acuerdo con las necesidades del usuario.¹⁹

La base de datos del ejemplo, como cualquier otra, se compone de columnas y renglones. En los renglones se asentarán cada una de las gestiones presentadas a la Comisión y nuestras observaciones; las columnas son las variables o atributos recabados de cada una de las gestiones (véase la figura 4). Las variables que contiene la base de datos de esta guía son:²⁰

¹⁹ Paco Megía Morales, *Ya sé Excel, pero necesito más*, 2ª ed., Madrid, Visión, 2007, p. 168.

²⁰ Subdirección de Estadística de la CDHDF, Sistema Integral de Gestión de Información (Siigesi), México, marzo de 2011, pp. 1-6.

- Folio de gestión (*id_gestion*) permite identificar de manera única la gestión mediante un folio numérico. Es una variable cuantitativa discreta.
- Fecha de creación (*Fecha_creación*): con esta variable sabemos el día, mes y año en que fue creada la gestión, por lo tanto, es una variable cuantitativa discreta.
- Estatus de la gestión (*Status_gestion*): indica el estado de la gestión; en la base de datos todas las gestiones están concluidas.
- Tipo de investigación (*Investigacion*): variable nominal con tres categorías:
 - *Directa*: cuando el servicio es solicitado por el o la peticionario(a).
 - *De reserva*: cuando el servicio es solicitado por el o la peticionario(a), pero menciona que sus datos de identificación personal deben mantenerse en reserva por temor a represalias.
 - *De oficio*: son las gestiones o servicios que se inician *motu proprio*.
- Unidad donde se recibió la gestión (*Nombre_unidad*): es una variable cualitativa nominal que identifica en cuál de las cinco unidades de la CDHDF se recibió la gestión: oficina central, Unidad Norte, Unidad Oriente, Unidad Poniente o Unidad Sur.
- Forma en la que se presentó la gestión (*Via_entrada*): variable cualitativa nominal que cuenta con nueve categorías: correo, correo electrónico, escrito, fax, medios masivos de comunicación, oficina de información pública, personal, por actuación de la CDHDF, telefónica.
- Tipo de servicio ofrecido (*Servicio*): variable nominal que cuenta con las siguientes categorías: aportación a expediente de queja, aportación a remisión, asesoría para formular escrito, canalización con oficio, curso de queja, curso de remisión, documento improcedente, expediente de colaboración, información sobre la CDHDF, medidas precautorias, orientación verbal, orientación verbal con solución inmediata, remisión, respuesta a peticiones por escrito, respuesta a peticiones vía electrónica, revisión de escrito, servicio de colaboración y suplencia de queja.
- Categoría de servicio ofrecido (*Super_tipo*): agrupación de la variable servicio ofrecido y es una variable cualitativa nominal con dos categorías: Servicios de Defensa y Prevención y Servicios de Protección.
- Identificador de persona (*Id_persona*): permite identificar de manera única a la persona que levanta la gestión mediante un folio numérico. Es una variable cuantitativa discreta.
- Carácter de la persona que presenta la gestión (*Character*): variable cualitativa nominal con tres categorías:
 - Peticionario(a): el servicio solicitado es en beneficio de un tercero.
 - Agravado: el servicio lo solicita la presunta víctima (puede ser una sola persona o un colectivo o grupo).
 - Ambos: el servicio solicitado es en beneficio de la persona que lo solicita.

- Edad de la persona que presenta la gestión (*Edad*): variable cuantitativa discreta.
- Sexo de la persona que presenta la gestión (*Sexo*): variable cualitativa nominal y cuenta con cuatro categorías: masculino, femenino, de oficio y colectiva.
- Nacionalidad de la persona que presenta la gestión (*Nacionalidad*): variable cualitativa nominal.
- Rango de edad de la persona que presenta la gestión (*Rango*): variable cualitativa ordinal que resulta de la agrupación de la variable edad.

Figura 4. Estructura de una base de datos

Las columnas corresponden a las variables.

Filas (renglones) representan cada uno de los casos (gestiones)

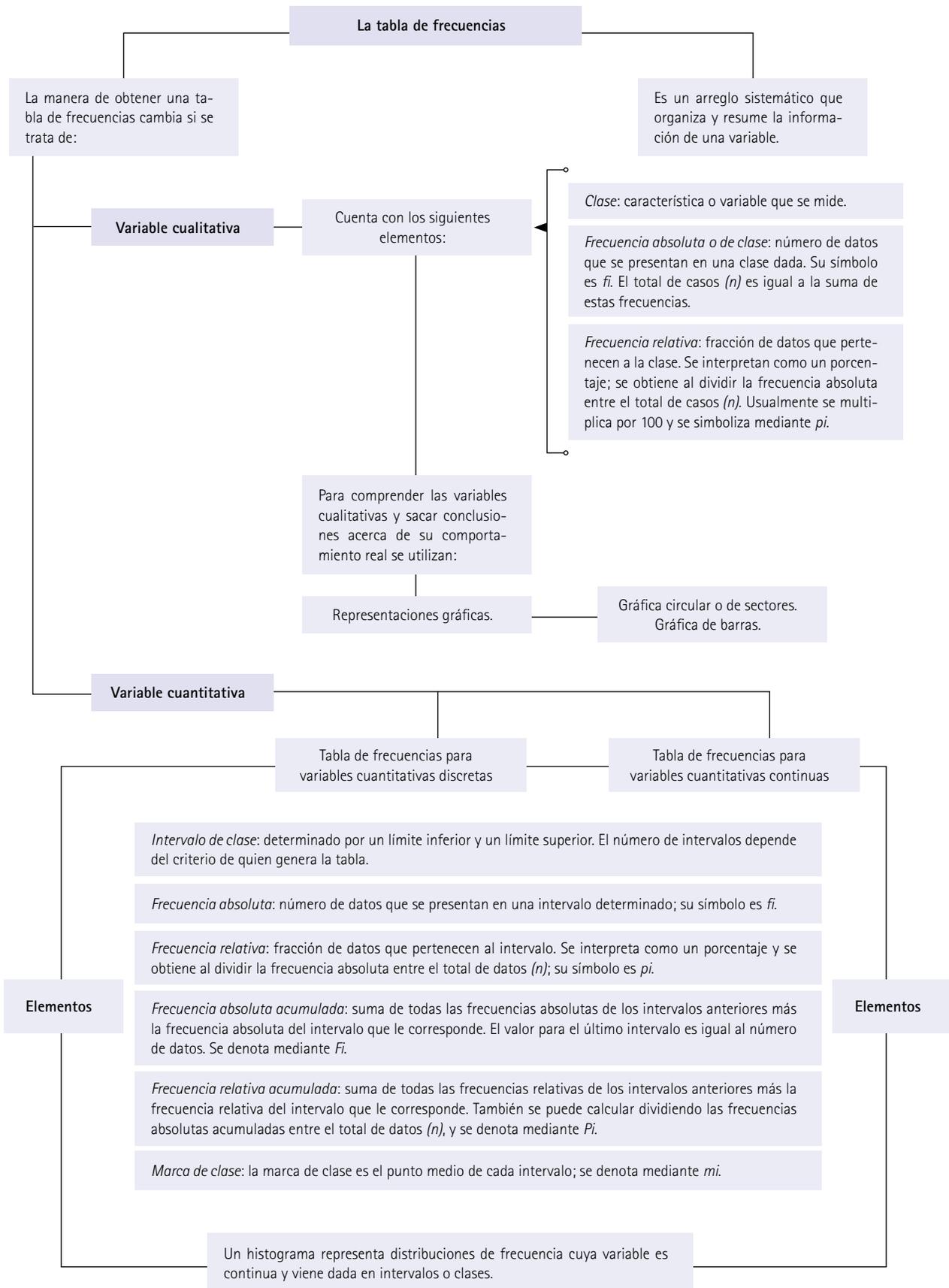
	A	B	C	D	E	F
1	Id_gesti	Fecha_creaci	Status_gesti	Investigaci	Nombre_unide	Via_entrada
2	415936	01/01/2011	Concluida	Directa	Oficina Sede	Telefónica
3	415936	01/01/2011	Concluida	Directa	Oficina Sede	Telefónica
4	415937	01/01/2011	Concluida	Directa	Oficina Sede	Telefónica
5	415937	01/01/2011	Concluida	Directa	Oficina Sede	Telefónica
6	415938	01/01/2011	Concluida	Directa	Oficina Sede	E-mail
7	415939	01/01/2011	Concluida	Directa	Oficina Sede	Telefónica
8	415940	01/01/2011	Concluida	De Oficio	Oficina Sede	Medios Masivos de C
9	415940	01/01/2011	Concluida	De Oficio	Oficina Sede	Medios Masivos de C
10	415941	01/01/2011	Concluida	Directa	Oficina Sede	Telefónica
11	415941	01/01/2011	Concluida	Directa	Oficina Sede	Telefónica

Nombre de las variables

Fuente: elaboración propia.

MÓDULO III.

Algunos ejercicios de estadística descriptiva (tabla de frecuencias)



En los módulos anteriores se revisaron los elementos básicos de una investigación estadística en derechos humanos, los conceptos de población y muestra, el tipo de escalas de medición de las variables, así como la forma como se construye una base de datos a partir de las quejas recibidas en la CDHDF.

El objetivo de este módulo es explicar y ejemplificar algunas técnicas utilizadas en estadística descriptiva, por lo que se centrará en las tablas de frecuencias y en las representaciones gráficas. En el próximo apartado se abordarán algunas medidas de resumen, como la media, mediana, moda, desviación estándar y varianza. Nuestra fuente de información será la base de datos que contiene 47 238 gestiones recabadas en las cinco unidades de la CDHDF, proporcionada por la Subdirección de Estadística de la Comisión.

Las tablas de frecuencias y la representación gráfica de los datos

La tabla de frecuencias, llamada también *distribución de frecuencias*, es un arreglo sistemático que organiza y resume la información de una variable. Además ayuda a identificar las veces que aparece cada categoría o modalidad en la muestra o población, también es el primer paso en la descripción de los datos recogidos.

La manera de obtener una tabla de frecuencias no es la misma para una variable cualitativa que para una variable cuantitativa.

Tablas de frecuencias para variables cualitativas

Las tablas de frecuencias para las variables cualitativas tienen los siguientes elementos:

- *Clase*: característica o variable que se mide. Hay tantas clases como número de categorías en las que se pueden clasificar los datos.
- *Frecuencia absoluta o de clase*: cantidad de datos que se presentan en una clase determinada. Su símbolo es f_i . El total de casos (n) es igual a la suma de estas frecuencias.
- *Frecuencia relativa*: datos que corresponden en la clase. Puede interpretarse como un porcentaje y se obtiene al dividir la frecuencia absoluta (f_i) entre el número total de casos (n). Usualmente, se multiplica por 100 y su símbolo es p_i . Además, la suma de frecuencias relativas es igual a 1 (o a 100%).

Ejemplo 7

Con el archivo de Excel llamado "Tabla de servicios.xls" y con la variable *Investigación* indagaremos si la mayoría de las investigaciones, del total presentado en 2011 ante la CDHDF, son directas. La manera más sencilla de averiguarlo es mediante una tabla de frecuencias. En este caso, la variable de interés *Investigación* es cualitativa nominal, y tiene tres categorías o clases: *de oficio*, *de reserva* y *directas*, además de que el total de casos es $n = 47\ 238$. Por lo tanto, debemos obtener una tabla de frecuencias para variables cualitativas mediante los siguientes pasos:

- 1) Abrir el archivo "Tabla de servicios.xls" en Excel.
- 2) En la hoja de gestiones y con ayuda del *mouse*, seleccionamos el conjunto de datos, incluido el nombre de las variables.
- 3) Después, en el menú *Insertar*, oprimir el ícono de *Tabla dinámica* (véase la figura 5a). Aparecerá la ventana *Crear tabla dinámica* (véase la figura 5b). En el recuadro *Tabla o rango* aparece el rango de datos que se seleccionó previamente.
- 4) Dejar activada la opción de *Nueva hoja de cálculo* y oprimir *Aceptar* (véase la figura 5).
- 5) Automáticamente aparecerá una nueva hoja que contiene la tabla dinámica con la cual se trabajará en los módulos III, IV y V. Esta tabla tiene la misma estructura que la mostrada en la primera parte de la figura 6.
- 6) Selecciona la variable *Investigación* del menú, luego *campos* para agregar al informe y, automáticamente, aparecerá como campo de fila.

- 7) En el recuadro Σ *Valores* arrastra con el *mouse* la variable *Investigación*. Esto da como resultado la frecuencia absoluta o de clase (*fi*) para las clases *De oficio*, *De reserva* y *Directa*, tal como se muestra en el segundo panel de la figura 6.

Figura 5. Menú Insertar, submenú Tabla dinámica

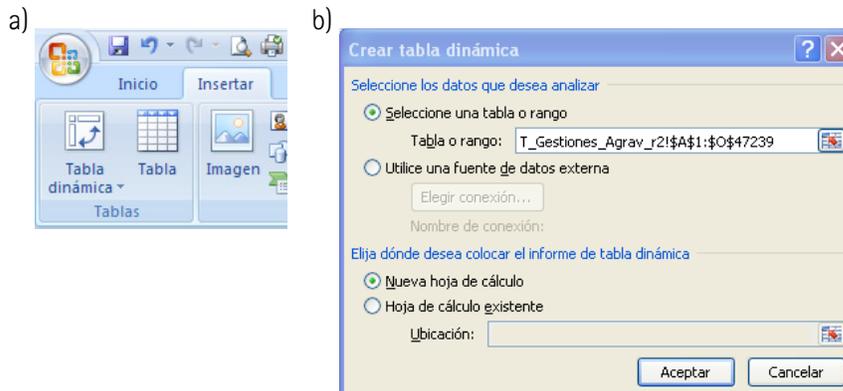
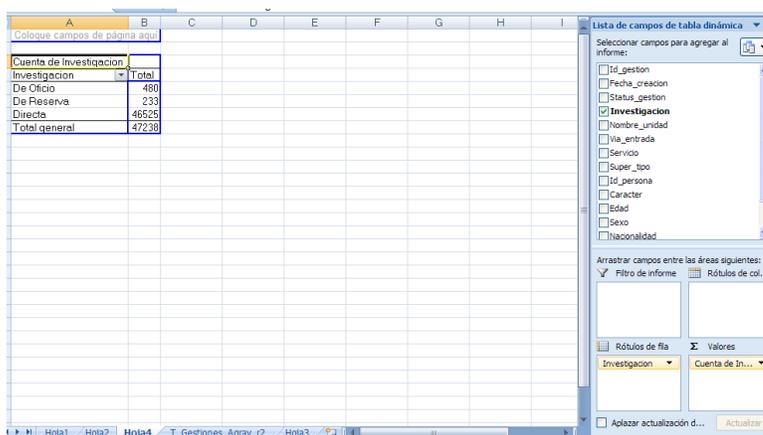
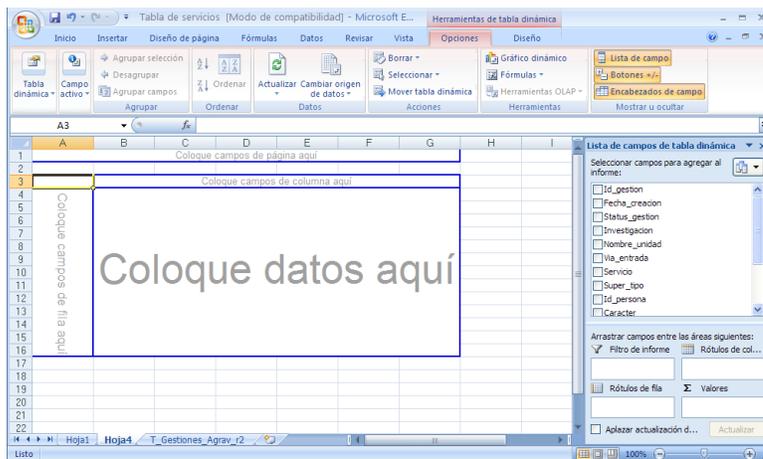


Figura 6. Tabla dinámica generada



Para manipular los valores de la tabla dinámica es necesario realizar la conversión de los valores obtenidos. Primero se seleccionan las celdas de la tabla dinámica. Posteriormente se realiza un *pegado especial* en el que se seleccionan *sólo valores* en una celda diferente, por ejemplo D4. Esto pega todos los valores sin referencia a la tabla dinámica.

Para obtener la frecuencia relativa (pi) se realiza el siguiente procedimiento:

- 1) Seleccionar la celda de la derecha del primer total y teclear, por ejemplo, `=E5/E8` (véase la figura 7). Esto divide el valor de la celda E5 entre el valor de la celda E8, que es la suma de todas las categorías.²¹
- 2) Oprimir *Enter* para obtener el valor de la pi de la clase *De oficio*.
- 3) Copiar la fórmula en las celdas F6 a F8 arrastrando el cursor de la esquina inferior derecha, para obtener la pi de las dos categorías restantes.
- 4) Realizar la suma total de las tres frecuencias absolutas. Esto se hace colocando el cursor en la celda F8 y escribiendo la siguiente fórmula: `= suma(F5,F7)`. Al oprimir *Enter*, aparecerá como resultado un número 1. Es posible hacer lo anterior posicionándose en la celda F8 y dando clic en el ícono Σ Autosuma.
- 5) Para obtener el porcentaje del total de investigaciones que corresponde a cada categoría, se debe obtener pi en porcentaje: dar clic en el ícono de porcentaje después de seleccionar las celdas de F5 a F8 para desplegar el resultado en términos de 100% en lugar de 1. El resultado se muestra en el cuadro 6. De ser necesario, dar clic en el botón *Aumentar decimales*, para una mejor visualización de los resultados.
- 6) Finalmente, se recomienda asignar un título a la tabla, así como a cada columna, tal y como se observa en el cuadro 6.

Figura 7. Generación de las frecuencias absoluta y relativa en Excel a partir de una tabla dinámica

	A	B	C	D	E	F
1						
2						
3	Cuenta de Investigación					
4	Investigación	Total		Investigación	Total	
5	De Oficio	480		De Oficio	480	=E5/\$E\$8
6	De Reserva	233		De Reserva	233	
7	Directa	46525		Directa	46525	
8	Total general	47238		Total general	47238	
9						

²¹ El símbolo \$ en la fórmula hace que la celda E8 se quede fija, por lo que, al ser copiada la fórmula, el valor correspondiente a E8 será el mismo para el resto de los cálculos.

Cuadro 6. Ejemplo de una tabla de frecuencia de una variable cualitativa

Tabla de frecuencias absoluta y relativa para la variable Investigación		
Investigación	Frecuencia absoluta (<i>f_i</i>)	Frecuencia relativa (<i>p_i</i>)
De oficio	480	1.02%
De reserva	233	0.49%
Directa	46 525	98.49%
Total general	47 238	100%

Fuente: elaboración propia con base en la información proporcionada por la Subdirección de Estadística de la CDHDF.

Ahora debemos interpretar la tabla de frecuencias y responder a la pregunta planteada anteriormente. Una forma de hacerlo es la siguiente: del total de investigaciones presentadas durante 2011 ante la CDHDF, 98.49% son de tipo directa; poco más de 1%, de oficio, y el resto (0.49%), de reserva. Por lo tanto, la mayoría de las investigaciones presentadas en las cinco unidades son de tipo directa.

Otra manera de responder a la pregunta del ejemplo 7 es utilizando una representación gráfica.

La representación gráfica de un conjunto de datos resulta muy eficaz para facilitar la comprensión y extraer conclusiones acerca del comportamiento real de las variables. Es necesario que el impacto visual de la representación gráfica resuma la información en forma clara, concisa y atractiva.²² Si bien es similar a la tabla de frecuencias, es distinta para las variables cuantitativas y cualitativas; en el último tipo, las representaciones gráficas más usadas son la gráfica circular o de sectores y las gráficas de barras.

Gráfica circular o de sectores

Este método gráfico es uno de los más simples y usuales, además de un valioso instrumento auxiliar de análisis y presentación de la información. Este diagrama en forma de círculo es particularmente útil para visualizar las diferencias de frecuencia entre algunas categorías de nivel nominal,²³ además de que el ángulo del sector será proporcional a la frecuencia absoluta o relativa de cada una de las clases de la variable.

²² Haroldo Elorza Perez-Tejeda, *op. cit.*, p. 20.

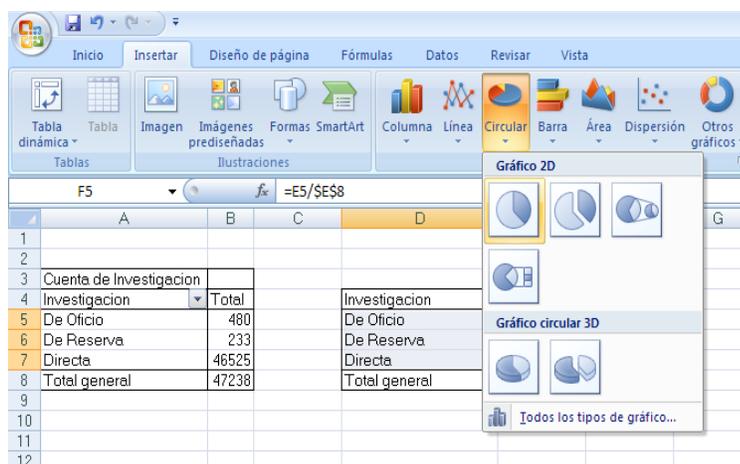
²³ *Ibidem*, p. 21.

Ejemplo 8

Con los datos del ejemplo 7 se obtendrá la gráfica de sectores. Para realizarla en Excel, se parte de la tabla de frecuencias. Los pasos a seguir son:

- 1) Seleccionar las celdas que contienen el nombre de las categorías y la de los valores de las frecuencias relativas (p_i).²⁴ En la opción *Insertar*, seleccionar la opción de gráfica *Circular*. Aparecerá un menú; hay que escoger la primera opción: *Gráfico 2D*. Automáticamente, aparecerá la gráfica que se muestra en la parte derecha de la figura 8.
- 2) Sólo resta añadir un título y la fuente de donde se obtuvieron los datos. Esto se hace con la opción *Herramientas de gráficos*, submenú *Título de gráficos*. Ahí seleccionamos *Título superpuesto centrado* y, en la parte el cuadro donde aparece la leyenda *Texto del gráfico*, se escribe el título (véase la figura 9).
- 3) Para poner la fuente de donde se obtuvieron los datos, en el menú *Insertar* oprimir el icono *Cuadro de texto* y, al final del cuerpo de la gráfica, arrastrar el *mouse* para hacer un cuadro, que aparecerá marcado en azul con líneas discontinuas (véase la figura 9) y se escribe el siguiente texto: Fuente: elaboración propia con base en la información proporcionada por la Subdirección de Estadística de la CDHDF.
- 4) También puede añadirse el porcentaje que corresponde a cada área al seleccionar el menú *Herramientas de gráficos* y oprimir el submenú *Etiquetas de gráficos*, y luego la opción *Ajuste perfecto*. La gráfica resultante se muestra en la figura 10.

Figura 8. Pasos para la elaboración de gráficas circulares o de sectores en Excel



²⁴ También es posible hacer la gráfica de sectores utilizando la frecuencia absoluta. Para eso, en lugar de seleccionar los valores de las frecuencias relativas p_i , se deberán escoger los de las frecuencias absolutas f_i .

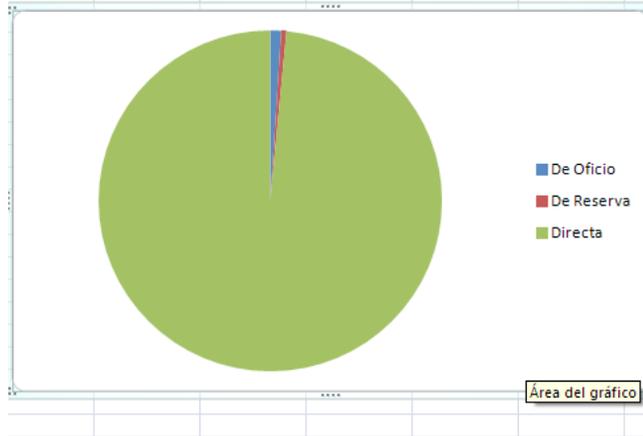


Figura 9. Pasos para dar formato a las gráficas en Excel

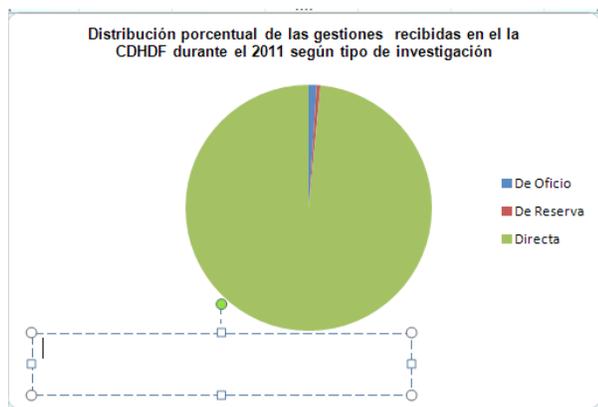
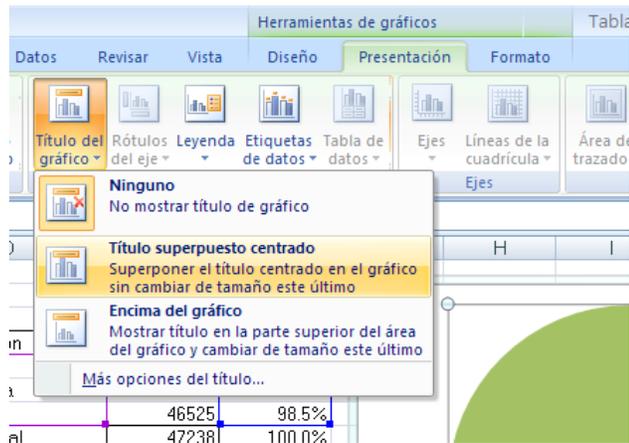


Figura 10. Ejemplo de una gráfica circular o de sectores realizada en Excel



Fuente: elaboración propia con base en la información proporcionada por la Subdirección de Estadística de la CDHDF.

En la gráfica circular se observa que el área sombreada mayor corresponde a las investigaciones de tipo directa y que el área de las investigaciones de reserva es mucho menor. Por lo tanto, se concluye que la mayoría de las investigaciones recibidas en la Comisión durante 2011 fueron de tipo directa.

Gráfica de barras

Otra representación gráfica de las variables cualitativas es la gráfica de barras, que consiste en un conjunto de columnas separadas que representan la frecuencia o el porcentaje de cada uno de los valores o categorías de la variable de interés; la altura de cada barra será proporcional a la frecuencia relativa o absoluta.²⁵

Ejemplo 9

Para saber en qué sede se recibieron más gestiones durante 2011, se empleará la gráfica de barras.

Al igual que en la gráfica de sectores, primero se obtiene la tabla de frecuencias. Al realizar todos los pasos del ejemplo 7, pero usando la variable *Nom-bre_Unidad*, se logra las siguientes frecuencias.

²⁵ Haroldo Elorza Perez-Tejeda, *op. cit.*, p. 21.

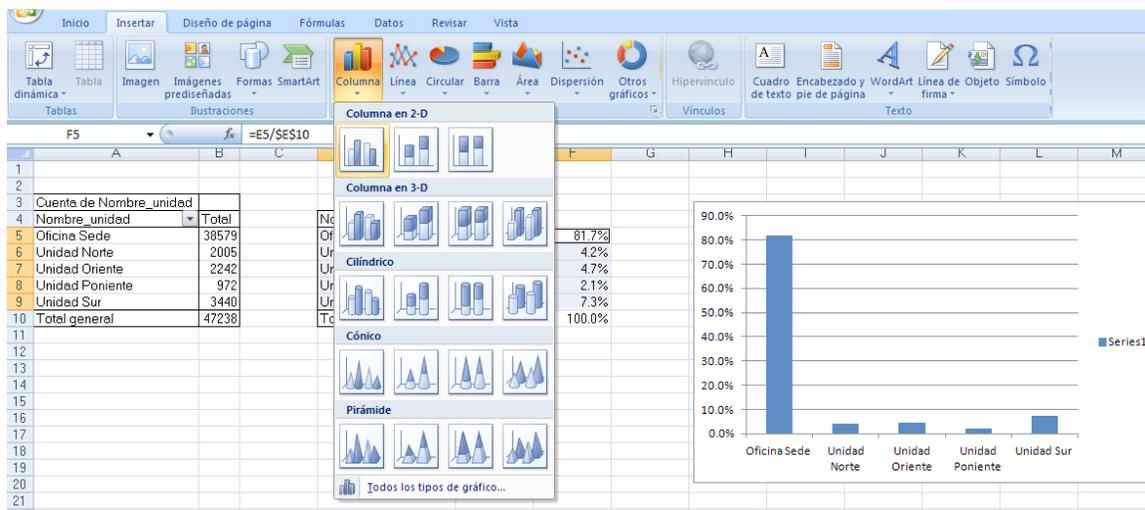
Cuadro 7. Frecuencias absoluta y relativa para la variable Nombre_Unidad

Nombre_Unidad	Frecuencia absoluta (f_i)	Frecuencia relativa (p_i)
Oficina central	38 579	81.67%
Unidad Norte	2 005	4.24%
Unidad Oriente	2 242	4.75%
Unidad Poniente	972	2.06%
Unidad Sur	3 440	7.28%
Total general	47 238	100%

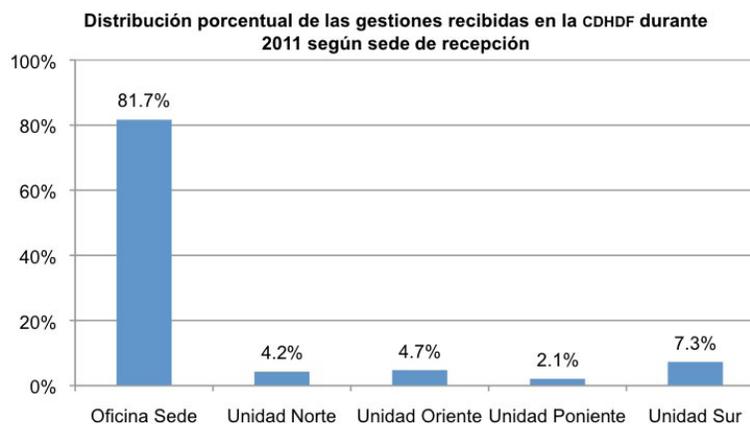
Fuente: elaboración propia con base en la información proporcionada por la Subdirección de Estadística de la CDHDF.

Los pasos para ejecutar una gráfica de barras en Excel son:

- 1) Seleccionar las celdas con el nombre de las categorías y con los valores de las frecuencias relativas (p_i).²⁶ En la opción *Insertar*, seleccionar la opción de gráfica *Columna*. Elegir la primera opción: *Columna en 2D*. Automáticamente aparecerá la gráfica que se muestra en la parte derecha de la figura 11.
- 2) Para añadir el título, la fuente de la cual se obtuvo la información y la etiqueta de los datos, hay que seguir los pasos del 2 al 4 del ejemplo 8.
- 3) Para quitar la guía de la serie de datos, se selecciona y se oprime la tecla *Supr*. Y el resultado es la gráfica que aparece al final de la figura 11.

Figura 11. Pasos para la elaboración de gráficas de barras en Excel

²⁶ También se puede hacer la gráfica de sectores utilizando la frecuencia absoluta; para eso, en lugar de seleccionar los valores de p_i , se optará por los de f_i .



Fuente: elaboración propia con base en la información proporcionada por la Subdirección de Estadística de la CDHDF.

En la gráfica de barras se observa que al eje horizontal (x) le corresponde la variable de interés *Nombre_unidad*; y al eje vertical (y), la frecuencia relativa, es decir, el porcentaje del total de las gestiones de cada sede dentro del total de las gestiones recibidas durante 2011. Además, como la barra perteneciente a la categoría de la oficina central es la más alta, se puede concluir que es la sede que recibió más gestiones. Otra manera sencilla de interpretar la frecuencia relativa (p_i) es la siguiente: de cada 100 gestiones que recibió la CDHDF durante 2011, 82 se recibieron en la oficina central, 7 en la Unidad Sur, 5 en Unidad Oriente, 4 en la Norte y sólo 2 en Unidad Poniente. Como se ve, no hay una única manera de interpretar los resultados. Tanto las tablas de frecuencias como las gráficas de sectores y las de barras son igualmente válidas.

Tablas de frecuencias para variables cuantitativas discretas

Las tablas de frecuencias para las variables de tipo cuantitativo son muy similares a las de los datos cualitativos. Sin embargo, debido a que no hay una división tácita entre valores cuantitativos, es necesario dividirlos en intervalos. Los intervalos serán, entonces, equivalentes a las clases. Los elementos son:

- *Intervalo de clase:* determinado por un límite inferior y un límite superior. El número de intervalos depende del criterio de quien genere la tabla.
- *Frecuencia absoluta:* número de datos que se presentan en un intervalo determinado y se representa mediante f_i .
- *Frecuencia relativa:* fracción de datos que caen en el intervalo. Se puede interpretar como un porcentaje y se obtiene al dividir la fre-

cuencia absoluta entre el número total de datos (n); se representa mediante p_i .

- **Frecuencia absoluta acumulada:** es la suma de todas las frecuencias absolutas de los intervalos anteriores más la frecuencia absoluta del intervalo que le corresponde. El valor para el último intervalo es igual al número de datos. Su símbolo es F_i .
- **Frecuencia relativa acumulada:** es la suma de todas las frecuencias relativas de los intervalos anteriores más la frecuencia relativa del intervalo que le corresponde. También se puede calcular dividiendo las frecuencias absolutas acumuladas entre el número total de datos (n); su símbolo es P_i .

Ejemplo 10

Para conocer el patrón de edad de las personas que levantaron alguna gestión durante 2011 ante la CDHDF, se creará una tabla de frecuencias con la variable *edad*. Esta variable contiene la información de edad en años cumplidos de las personas que presentaron las gestiones durante el periodo analizado, por lo tanto es cuantitativa, discreta y de intervalo. Los pasos para obtener esta tabla son:

- 1) Para obtener la frecuencia absoluta (f_i) y la relativa (p_i) de la tabla de frecuencias de la variable de interés, se seguirán los pasos del 2 al 4 del ejemplo 7.
- 2) Una vez hecho lo anterior, para obtener F_i se posicionará el cursor en la celda G5 y se escribirá la siguiente fórmula =E5. Este valor debe darnos un resultado igual al del primer renglón de f_i , tal y como se muestra en el primer panel de la figura 12.
- 3) Para obtener el siguiente dato, se posiciona el cursor en la celda inferior (E6) y se escribe =E6+G5. Esto suma el valor de la frecuencia absoluta del primer valor más la frecuencia absoluta del segundo valor (véase la figura 12).
- 4) Se copia la fórmula anterior para todas las celdas subsecuentes hacia abajo con excepción de la última (la del *Total general*), pues la penúltima celda acumula los 46 499 casos totales.²⁷ De esta manera, cada celda es la suma de su valor de frecuencia absoluta más todos los anteriores; es acumulativa.

²⁷ Se quitó la opción de categoría *En blanco* debido a que no aportaba información alguna. Sin embargo, es importante mencionar que en esta categoría se agrupan las gestiones colectivas y, por lo tanto, no se puede asignar un valor único a la variable edad.

Figura 12. Generación de la frecuencia relativa, absoluta y absoluta acumulada en una tabla de frecuencias en Excel a partir de una tabla dinámica con una variable cuantitativa discreta

The figure shows two screenshots of an Excel spreadsheet. The top screenshot shows the calculation of relative frequencies (p_i) in column F. The formula bar shows $=E5$. The bottom screenshot shows the calculation of cumulative relative frequencies (P_i) in column G. The formula bar shows $=E6+G5$.

Edad	f_i	p_i	F_i	P_i
0	83	0.2%	83	83
1	47	0.1%	130	179
2	49	0.1%	179	264
3	85	0.2%	264	342
4	78	0.2%	342	420

Para obtener la frecuencia relativa acumulada (P_i), hay que seguir los siguientes pasos:

- 1) Seleccionar la celda junto a la frecuencia relativa (F_5) y escribir la fórmula $=F_5$, por ejemplo, por ser esta la primera celda de la frecuencia relativa (véase la figura 13).
- 2) Seleccionar la celda inferior (G_6) y escribir $=F_6+H_5$ (véase la figura 13).
- 3) Se copia la fórmula anterior para todas las celdas subsecuentes hacia abajo con excepción de la última, pues la penúltima celda acumula 100% de los casos totales (véase la figura 13).

Figura 13. Generación de la frecuencia relativa acumulada en una tabla de frecuencias en Excel a partir de una tabla dinámica con una variable cuantitativa

The screenshot shows the first row of the cumulative relative frequency calculation. The formula bar shows $=F5$. The cell F5 contains the value 0.2%.

Edad	f_i	p_i	F_i	P_i
0	83	0.2%	83	=F5
1	47	0.1%	130	
2	49	0.1%	179	

PEARSON		=F6+H5						
	A	B	C	D	E	F	G	H
1								
2								
3	Cuenta de Edad							
4	Edad	Total		Edad	f_i	p_i	F_i	P_i
5		0	83		0	0.2%	83	0.2%
6		1	47		1	0.1%	130	=F6+H5
7		2	49		2	0.1%	179	
8		3	85		3	0.2%	264	

El resultado de la tabla de frecuencia anterior se aprecia en la cuadro 8.

Cuadro 8. Frecuencia absoluta y relativa, simple y acumulada para la variable *Edad*

Edad	f_i	p_i	F_i	P_i
0	83	0.2%	83	0.2%
1	47	0.1%	130	0.3%
2	49	0.1%	179	0.4%
3	85	0.2%	264	0.6%
4	78	0.2%	342	0.7%
5	71	0.2%	413	0.9%
6	99	0.2%	512	1.1%
7	98	0.2%	610	1.3%
8	78	0.2%	688	1.5%
9	88	0.2%	776	1.7%
10	57	0.1%	833	1.8%
11	74	0.2%	907	2.0%
12	89	0.2%	996	2.1%
13	101	0.2%	1097	2.4%
14	146	0.3%	1243	2.7%
15	182	0.4%	1425	3.1%
16	177	0.4%	1602	3.4%
17	239	0.5%	1841	4.0%
18	238	0.5%	2079	4.5%
19	288	0.6%	2367	5.1%
20	319	0.7%	2686	5.8%
21	342	0.7%	3028	6.5%
22	413	0.9%	3441	7.4%
23	426	0.9%	3867	8.3%
24	468	1.0%	4335	9.3%
25	576	1.2%	4911	10.6%
26	522	1.1%	5433	11.7%
27	560	1.2%	5993	12.9%
28	699	1.5%	6692	14.4%
29	628	1.4%	7320	15.7%
30	711	1.5%	8031	17.3%
31	592	1.3%	8623	18.5%
32	675	1.5%	9298	20.0%
33	687	1.5%	9985	21.5%
34	697	1.5%	10682	23.0%
35	741	1.6%	11423	24.6%
36	793	1.7%	12216	26.3%
37	731	1.6%	12947	27.8%
38	959	2.1%	13906	29.9%
39	737	1.6%	14643	31.5%
40	794	1.7%	15437	33.2%
41	801	1.7%	16238	34.9%
42	841	1.8%	17079	36.7%
43	786	1.7%	17865	38.4%
44	699	1.5%	18564	39.9%
45	791	1.7%	19355	41.6%
46	845	1.8%	20200	43.4%
47	812	1.7%	21012	45.2%
48	709	1.5%	21721	46.7%
49	716	1.5%	22437	48.3%
50	693	1.5%	23130	49.7%
51	598	1.3%	23728	51.0%
52	630	1.4%	24358	52.4%
53	565	1.2%	24923	53.6%
54	577	1.2%	25500	54.8%
55	552	1.2%	26052	56.0%
56	523	1.1%	26575	57.2%
57	487	1.0%	27062	58.2%
58	448	1.0%	27510	59.2%
59	438	0.9%	27948	60.1%
60	493	1.1%	28441	61.2%
61	400	0.9%	28841	62.0%
62	401	0.9%	29242	62.9%
63	320	0.7%	29562	63.6%
64	280	0.6%	29842	64.2%
65	282	0.6%	30124	64.8%
66	234	0.5%	30358	65.3%
67	218	0.5%	30576	65.8%
68	252	0.5%	30828	66.3%
69	228	0.5%	31056	66.8%
70	184	0.4%	31240	67.2%
71	199	0.4%	31439	67.6%
72	147	0.3%	31586	67.9%
73	147	0.3%	31733	68.2%
74	140	0.3%	31873	68.5%
75	127	0.3%	32000	68.8%
76	144	0.3%	32144	69.1%
77	97	0.2%	32241	69.3%
78	72	0.2%	32313	69.5%
79	54	0.1%	32367	69.6%
80	73	0.2%	32440	69.8%
81	74	0.2%	32514	69.9%
82	53	0.1%	32567	70.0%
83	46	0.1%	32613	70.1%
84	35	0.1%	32648	70.2%
85	45	0.1%	32693	70.3%
86	26	0.1%	32719	70.4%
87	24	0.1%	32743	70.4%
88	14	0.0%	32757	70.4%
89	14	0.0%	32771	70.5%
90	22	0.0%	32793	70.5%
91	4	0.0%	32797	70.5%
92	9	0.0%	32806	70.6%
93	3	0.0%	32809	70.6%
94	4	0.0%	32813	70.6%
96	2	0.0%	32815	70.6%
97	1	0.0%	32816	70.6%
9999	13683	29.4%	46499	100%
Total	46499	100%		

Fuente: elaboración propia con base en la información proporcionada por la Subdirección de Estadística de la CDHDF.

En el cuadro 8 se muestra, por ejemplo, que de las 46 499 gestiones presentadas ante la CDHDF, que tienen información sobre edad, 841 fueron presentadas por personas de 42 años de edad, lo que equivale a 1.78% del total de las gestiones. También se observa que poco menos de un tercio (29.4%) de las gestiones tiene como valor 9999 en la variable *Edad*, el cual corresponde a la información de las gestiones realizadas en la categoría *De oficio*, por lo que la variable edad no aplica.

Por otro lado, el cuadro 8 no es muy útil, pues tiene aproximadamente 100 renglones, y la mayoría tiene menos de 2% de los casos debido al amplio rango de edades que se maneja. Para afinar la información, se recomienda abordar la edad como una variable cualitativa ordinal, uniendo las edades en grupos, ya sea de manera quinquenal (de 5 en 5 edades), decenal (de 10 en 10 edades) o cualquier otro tipo de agrupación que satisfaga las necesidades de la investigación. Una vez hecho esto, la tabla de frecuencias se construirá igual que en el ejemplo 7.

En el siguiente ejemplo se describirá el proceso para convertir la variable *Edad* de cuantitativa discreta en cualitativa ordinal.

Ejemplo 11

Para que una variable sea considerada continua, deberá tomar cualquier valor intermedio entre dos valores sucesivos, lo que no sucede con la edad, pues comúnmente se pregunta en años cumplidos. Por lo tanto, como ya se mencionó, la edad en años cumplidos es una variable cuantitativa discreta, ya que no se toman edades intermedias. Sin embargo, debido a la gran cantidad de datos en este caso, es recomendable agrupar las edades para una lectura más simple, tanto de la tabla como de las gráficas. De esta manera, convertiremos la variable *Edad* de cuantitativa discreta en cuantitativa ordinal.

Primero se crearán grupos de edad por grupos decenales, es decir, los grupos irán de 0 a 9 años, de 10 a 19 años, etc. Como la edad mayor en este caso es menor a 99, el último intervalo por considerar será 90-99 años. Los casos sin información en esta variable, o con valor 9999, no se tomarán en cuenta.

Para agrupar las edades en Excel el procedimiento es el siguiente:

- 1) Después de realizar los siete primeros pasos del ejemplo 7, se obtiene una tabla dinámica con la variable *Edad*.
- 2) Para hacer el grupo de 0 a 9 años, se seleccionan las celdas con la etiqueta de estas edades. Después, en el menú *Datos*, se oprime el botón *Agrupar* y, automáticamente, aparece una nueva columna con la leyenda de *Grupo 1*.

- 3) Para formar el grupo de edades de 10 a 19 años, se seleccionan las celdas de las edades 10 a 19 años. Se oprime *Agrupar* del menú *Datos*. Una vez hecho esto, aparecerá el grupo 2.
- 4) Para generar el resto de los grupos, repetimos los pasos anteriores. Recuerde que el grupo 3 corresponde a las edades 20 a 29, el grupo 4 a las edades 30 a 39, y así sucesivamente, en grupos de diez, hasta el grupo 10, de 90 a 99. Sin embargo, como en este ejemplo sólo tenemos hasta los 97 años, se selecciona hasta esa edad para el último grupo. Como los valores 9999 y *En blanco* no interesa conservarlos, estas dos opciones se incorporarán al grupo 11 (véase el primer panel derecho de la figura 14).
- 5) Para colapsar los grupos con el fin de obtener la frecuencia absoluta (f_i), se oprime el botón  que se encuentra del lado izquierdo de la etiqueta de cada número, para cada uno de los grupos (véase el segundo panel izquierdo de la figura 14).
- 6) La tabla final se muestra en el segundo panel derecho de la figura 14.
- 7) Para encontrar f_i , p_i , F_i y P_i , además de dar formato a la tabla, sólo hay que seguir los pasos de los ejemplos anteriores. Recuerde que sólo nos interesa usar la información de los 10 primeros grupos.
- 8) El cuadro 9 es nuestro resultado final.

Figura 14. Ejemplo de cómo transformar una variable cuantitativa discreta en una variable cualitativa ordinal

	A	B	C			
1	Coloque campos de página aquí			79		
2				80		74
3	Cuenta de Edad			81		127
4	Edad2	Edad	Total	82		144
5	Grupo1	0	83	83		97
6		1	47	84		72
7		2	49	85	Grupo9	54
8		3	85	86		73
9		4	78	87		74
10		5	71	88		53
11		6	99	89		46
12		7	98	90		35
13		8	78	91		45
14		9	88	92		26
15	10	10	57	93		24
16	11	11	74	94		14
17	12	12	89	95	Grupo10	14
18	13	13	101	96		22
19	14	14	146	97		4
20	15	15	182	98		9
21	16	16	177	99		3
22	17	17	239	100		4
23	18	18	238	101		2
24	19	19	288	102	Grupo11	1
25	20	20	319	103		9999
				104		(en blanco)
				105	Total general	46499
				106		

	A	B	C
1	Coloque campos de página aquí		
2			
3	Cuenta de Edad		
4	Edad2	Edad	Total
5	⊕ Grupo1		776
6	⊕ Grupo2		1591
7	⊕ Grupo3		4953
8	⊕ Grupo4		7323
9	⊕ Grupo5		7794
10	⊕ Grupo6		5511
11	⊕ Grupo7		3108
12	⊕ Grupo8		1311
13	⊕ Grupo9	80	73
14		81	74
15		82	53
16		83	46
17		84	35
18		85	45
19		86	26
20		87	24
21		88	14
22		89	14
23	⊕ Grupo10	90	22
24		91	4
25		92	9
26		93	3
27		94	4
28		96	2
29		97	1
30	⊕ Grupo11	9999	13683
31		(en blanco)	
32	Total general		46499

	A	B	C
1			
2			
3	Cuenta de Edad		
4	Edad2	Edad	Total
5	⊕ Grupo1		776
6	⊕ Grupo2		1591
7	⊕ Grupo3		4953
8	⊕ Grupo4		7323
9	⊕ Grupo5		7794
10	⊕ Grupo6		5511
11	⊕ Grupo7		3108
12	⊕ Grupo8		1311
13	⊕ Grupo9		404
14	⊕ Grupo10		45
15	⊕ Grupo11		13683
16	Total general		46499
17			

Cuadro 9. Ejemplo de una tabla de frecuencias para una variable de tipo cualitativa ordinal. Distribución en grupos decenales de la edad de las personas que realizaron las gestiones ante la CDHDF durante 2011

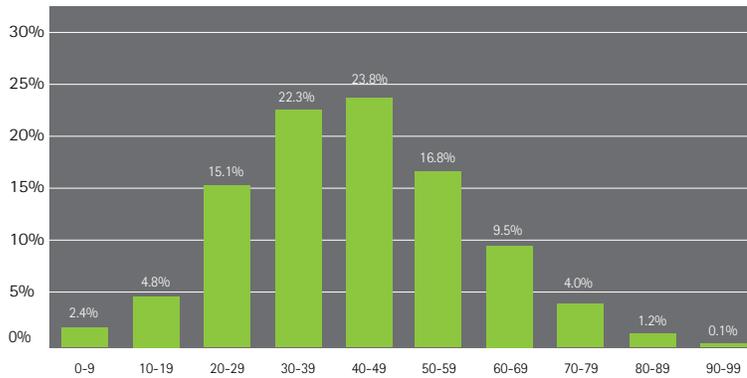
Grupo de edad	Frecuencia absoluta (f_i)	Frecuencia absoluta acumulada (F_i)	Frecuencia relativa (p_i)	Frecuencia relativa acumulada (P_i)
0-9	776	776	2.4%	2.4%
10-19	1 591	2 367	4.8%	7.2%
20-29	4 953	7 320	15.1%	22.3%
30-39	7 323	14 643	22.3%	44.6%
40-49	7 794	22 437	23.8%	68.4%
50-59	5 511	27 948	16.8%	85.2%
60-69	3 108	31 056	9.5%	94.6%
70-79	1 311	32 367	4.0%	98.6%
80-89	404	32 771	1.2%	99.9%
90-99	45	32 816	0.1%	100%
Total	32 816		100%	

Fuente: elaboración propia con base en la información proporcionada por la Subdirección de Estadística de la CDHDF.

Como se observa, el cuadro 9 es más sencillo de leer que el cuadro 8, ya que tiene tan sólo 10 grupos o renglones en vez de 100. Al seguir los pasos del ejemplo 9, se obtiene una gráfica de barras con f_i (véase la figura 15).

Figura 15. Ejemplo de una gráfica de barras para una variable cualitativa ordinal

Distribución de la frecuencia de la edad de las personas en grupo decenales que realizaron las gestiones ante la CDHDF durante 2011



Fuente: elaboración propia con base en la información proporcionada por la Subdirección de Estadística de la CDHDF.

Con la información de la tabla de frecuencias y de la gráfica de barras se observa que la mayoría de las gestiones realizadas ante la Comisión en 2011 fueron interpuestas por mujeres y hombres de entre 20 y 59 años de edad, y que aproximadamente una cuarta parte (23.8%) de las gestiones provinieron de personas de entre 40 y 49 años de edad; los grupos de edad que tienen menos representación son los de los menores de 19 años, 7 de cada 100 gestiones son de mujeres y hombres de este grupo de edad, y sólo 0.1% son mayores de 90 años.

Tabla de frecuencias para variables cuantitativas continuas

Las tablas de frecuencias para las variables cuantitativas continuas tienen los mismos elementos que las cuantitativas discretas (f_i , p_i , F_i y P_i), más una columna titulada *Marca de clase*, que es el punto medio de cada intervalo y se representa mediante m_i .

Ejemplo 12

Como la base de datos utilizada hasta ahora no cuenta con una variable cuantitativa continua, la elaboración de una tabla de frecuencia de este tipo se realiza con datos de una fuente externa.

Supóngase que se cuenta con la información del tiempo que esperaron 27 personas para levantar una denuncia ante la CDHDF de la Unidad Poniente. El tiempo se registró en horas y los datos son:

2.3	2.7	3.3	3.6	3.7	3.7	3.7
3.7	3.8	3.8	3.9	4.0	4.4	4.7
4.3	4.3	4.4	4.4	5.2	5.4	5.9
4.8	4.9	5.1	5.1	8.2	8.9	

Fuente: elaboración propia con base en la información simulada.

Con esta información se analizará el patrón de tiempo de espera para levantar una denuncia en la Unidad Poniente. Para realizar una tabla de frecuencias los pasos son:

- 1) Se registran los datos en Excel en una columna —columna A, por ejemplo—. A este tipo de arreglo se le conoce como *Datos no agrupados*.
- 2) Se ordenan los datos seleccionando toda la columna con la información del tiempo de espera. En el menú *Datos*, se oprime el botón  , el cual ordena de menor a mayor, de manera automática, los datos seleccionados. Así se identifica el valor mínimo (2.3 horas, y se encuentra en la celda B3) y el valor máximo (8.9 horas, localizado en la celda B29).
- 3) Después se obtiene la amplitud, que es el valor máximo menos el valor mínimo. En Excel se sitúa el cursor en la celda B1, por ejemplo, y se escribe la siguiente fórmula: =B29 - B3, la cual dará como resultado 6.6.
- 4) Para decidir los intervalos de clase necesarios, llamaremos k a dicho valor. Sin pérdida de generalidad, se puede optar por una $k = 4$, es decir, una tabla de frecuencias con 4 intervalos de clase.²⁸ Para fijar el ancho de cada intervalo de clase, se le asigna también un valor arbitrario y se le llama c . Se recomienda que $k * c$ sea un poco mayor que la amplitud de los datos. En este caso, se elige $c = 1.7$. El valor c no necesariamente es un número entero. Además, $k * c = 6.8$ es mayor que la amplitud (6.6). Para realizar los cálculos de manera sencilla, se anotan estos valores en las celdas de la hoja de cálculo de Excel (véase la figura 16).
- 5) Es importante mencionar que los extremos de los intervalos de clase se denominan límite inferior (representado por *L inf*) y límite superior (indicado por *L sup*).
- 6) Se elige el valor inicial que limitará el primer intervalo de clase y a partir del cual se obtendrán todos los puntos que limitarán a cada uno de los siguientes intervalos. El primer *L inf* deberá ser un número menor al valor mínimo observado.

En el ejemplo, el mínimo observado es de 2.3 horas, por lo que se puede elegir como *L inf* del primer intervalo el valor 2. Para obtener el

²⁸ Es importante recordar que un intervalo de clase se utiliza para resumir, simplificar y manipular gran cantidad de datos de tipo cuantitativo.

siguiente L_{inf} , se le suma el valor de c . Como ya se tienen los datos en Excel, basta escribir la siguiente fórmula $=E3 + \$C\33 (véase la figura 17). Este procedimiento da como resultado un L_{inf} del segundo intervalo de 3.7. Como del valor de c se encuentra en la celda C33, que se utilizará en el resto de los intervalos, se agregan los signos de \$ para fijar el valor de la celda. Para obtener el L_{inf} del resto de los intervalos, sólo se copia la fórmula en las celdas siguientes, posicionando el cursor en la celda E5 y, poniendo el *mouse* en la parte inferior izquierda del rectángulo (aparecerá una pequeña cruz), arrastrar el *mouse* hasta la celda que corresponde al cuarto intervalo de clase (véase la figura 17).

El L_{sup} del primer intervalo de clase será igual al L_{inf} del segundo, y así sucesivamente. Para hacerlo, sólo se pone en la celda F4 la siguiente fórmula $=E5$. Después se copia en el segundo y tercer intervalo (véase la figura 17).

Note que los valores observados de los L_{inf} no estarán contenidos en dicho intervalo. Por ejemplo, en el segundo intervalo de clase se obtiene el $L_{inf}= 3.7$ y el $L_{sup} =5.4$, por lo que las observaciones con valor igual a 3.7 no serían parte de este intervalo, sino del primer intervalo de clase. Las observaciones con valor igual a 5.4 sí son parte de este segundo intervalo.

- 7) El valor del L_{sup} del último intervalo de clase se obtiene sumando 1.7 (que es el valor c) al L_{inf} . Para hacerlo en Excel, se pone la siguiente fórmula: $= E7 + C33$. El resultado debe ser igual al mostrado en la figura 17.
- 8) Para obtener f_i , se realiza el conteo de las observaciones cuyos valores pertenecen a cada intervalo de clase. Por ejemplo, hay 8 observaciones en los valores 2.0 y 3.7. Una vez obtenidas las f_i , si se suman estos conteos, el valor resultante debe ser el total de las observaciones, es decir, 27.
- 9) Para obtener p_i , F_i y P_i , se siguen del paso 2 al 7 del ejemplo 10. Observe cómo el último valor de F_i es igual al total de observaciones, es decir, 27. De la misma manera, la suma de la columna p_i debe ser 1 o 100, en caso de que se maneje como porcentaje (véase segundo panel de la figura 17).
- 10) Al agrupar las observaciones en intervalos de clase se pierde el valor o magnitud que tenían antes, por ello es conveniente elegir un punto central del intervalo de clase como el valor que represente la magnitud de todas las observaciones que se clasificaron en dicho intervalo. A este valor se le llama, como se mencionó, marca de clase (m_i).²⁹

Para obtenerlo, basta con hacer la siguiente operación para cada intervalo:

²⁹ Víctor Aguirre y Begoña Artaloitia, "Análisis exploratorio de datos", *op. cit.*, pp. 1-16.

$$m_i = \frac{L_{sup} - L_{inf}}{2}$$

En Excel se escribe en la celda K3 la siguiente fórmula: = (E3 + F3)/2. Se añade el título de la tabla de frecuencias, de las columnas y la fuente (véase el segundo panel de la figura 17). El resultado final se muestra en el cuadro 10.

Al arreglo de datos de una tabla de frecuencias como la que se acaba de obtener se le conoce como *datos agrupados*.

Figura 16. Ejemplo para generar una tabla de frecuencias en Excel para una variable cuantitativa continua

	B	C
1		
2	Tiempo de espera (hr)	
3	2.3	
4	2.7	
5	3.3	
6	3.6	
7	3.7	
8	3.7	
9	3.7	
10	3.7	
11	3.8	
12	3.8	
13	3.9	
14	4.0	
15	4.3	
16	4.3	
17	4.4	
18	4.4	
19	4.4	
20	4.7	
21	4.8	
22	4.9	
23	5.1	
24	5.1	
25	5.2	
26	5.4	
27	5.9	
28	8.2	
29	8.9	
30		
31	Amplitud	=B29-B3

Figura 17. Generación de una tabla de frecuencias en Excel para una variable cuantitativa continua

	A	B	C	D	E	F
1						
2		Tiempo de espera (hr)		Intervalos de clase	Límite inferior	Límite superior
3		2.3		1	2	=E5
4		2.7		2	=E4+\$C\$33	=E6
5		3.3		3	=E5+\$C\$33	=E7
6		3.6		4	=E6+\$C\$33	=E7+C33
7		3.7		Total		
8		3.7				
9		3.7				
10		3.7				
11		3.8				
12		3.8				
13		3.9				
14		4				
15		4.3				
16		4.3				
17		4.4				
18		4.4				
19		4.4				
20		4.7				
21		4.8				
22		4.9				
23		5.1				
24		5.1				
25		5.2				
26		5.4				
27		5.9				
28		8.2				
29		8.9				
30						
31		Amplitud	=B29-B3			
32		k	4			
33		c	1.7			
34						

	A	B	D	E	F	G	H	I	J	K
1										
2		Tiempo de espera (hr)	Intervalos de clase	Límite inferior	Límite superior	Frecuencia absoluta	Frecuencia absoluta acumulada	Frecuencia relativa	Frecuencia relativa acumulada	Marcas de clase
3	1	2.3				(fi)	(Fi)	(fi)	(Fi)	(mi)
4	2	2.7	1	2	=E5	8	=G4	=G4/\$G\$8	=I4	=(E4+F4)/2
5	3	3.3	2	=E4+\$C\$33	=E6	16	=G5+H4	=G5/\$G\$8	=I5+J4	=(E5+F5)/2
6	4	3.6	3	=E5+\$C\$33	=E7	1	=G6+H5	=G6/\$G\$8	=I6+J5	=(E6+F6)/2
7	5	3.7	4	=E6+\$C\$33	=E7+C33	2	=G7+H6	=G7/\$G\$8	=I7+J6	=(E7+F7)/2
8	6	3.7	Total			=SUMA(G4:G7)		=SUMA(I4:I7)		
9	7	3.7								
10	8	3.7								
11	1	3.8								
12	2	3.8								
13	3	3.9								
14	4	4								
15	5	4.3								
16	6	4.3								
17	7	4.4								
18	8	4.4								
19	9	4.4								
20	10	4.7								
21	11	4.8								
22	12	4.9								
23	13	5.1								
24	14	5.1								
25	15	5.2								
26	16	5.4								
27	1	5.9								
28	1	8.2								
29	2	8.9								

Cuadro 10. Tabla de frecuencias del tiempo que esperaron las personas para levantar una denuncia ante la CDHDF en la Unidad Poniente

Intervalos de clase	Límite inferior (L_{inf})	Límite superior (L_{sup})	Frecuencia absoluta (f_i)	Frecuencia absoluta acumulada (F_i)	Frecuencia relativa (p_i)	Frecuencia relativa acumulada (P_i)	Marcas de clase (m_i)
1	2.0	3.7	8	8	29.6%	29.6%	2.85
2	3.7	5.4	16	24	59.3%	88.9%	4.55
3	5.4	7.1	1	25	3.7%	92.6%	6.25
4	7.1	8.8	2	27	7.4%	100%	7.95
Total			27		100%		

Fuente: elaboración propia con base en la información simulada.

Con base en los resultados de la tabla anterior, se sabe que 30% de las personas que levantaron una denuncia en la Unidad Poniente esperaron entre 2.0 y 3.7 horas. El 59% esperó más de 3.7 horas y menos de 5.4 horas. Utilizando los valores de F_i , concluimos que 89% de las personas esperaron menos de 5.4 horas para levantar una denuncia.

Cabe mencionar que si se cambia el ancho de las clases, la distribución de frecuencia también cambia. Sin embargo, la interpretación es similar.³⁰

Histograma³¹

Otra forma de conocer el patrón de tiempo de espera es mediante el uso de un histograma de frecuencias.

Un histograma es un diagrama útil para representar distribuciones de frecuencia cuya variable es continua y viene dada en intervalos o clases.³²

Este tipo de gráfica es una forma especial de una gráfica de barras, en la cual los intervalos de clase están representados por el ancho de las barras y las frecuencias de las observaciones son proporcionales a las áreas de los rectángulos. La altura de las barras puede ser la frecuencia absoluta o la relativa, el perfil del histograma será el mismo, sin embargo, es más informativo con frecuencias relativas puesto que así tenemos una idea del porcentaje de observaciones que ocurrieron por intervalo.³³

³⁰ *Ibidem*, pp. 1-17.

³¹ Otras representaciones gráficas se utilizan para las variables de tipo cuantitativas, como los diagramas de caja y brazos, diagramas de puntos, diagrama de tallos y hojas, polígonos de frecuencias y la ojiva. Para consultar la manera de realizarlos, véase Víctor Aguirre y Begoña Artaloitia, *op. cit.*, pp. 1-23.

³² Haroldo Elorza Pérez-Tejeda, *op. cit.*, p. 24.

³³ Víctor Aguirre y Begoña Artaloitia, *op. cit.*, pp. 1-18.

De esta manera, el histograma se construye como la gráfica de barras con la diferencia de que las columnas no están separadas, sino unidas, lo que le da continuidad.

Ejemplo 14

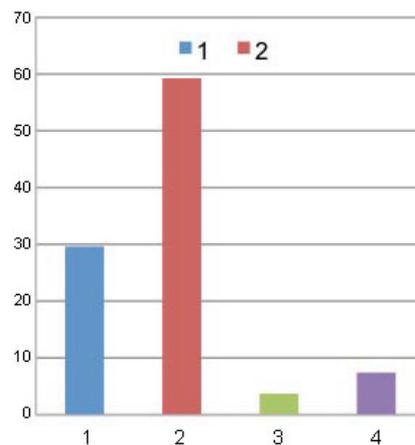
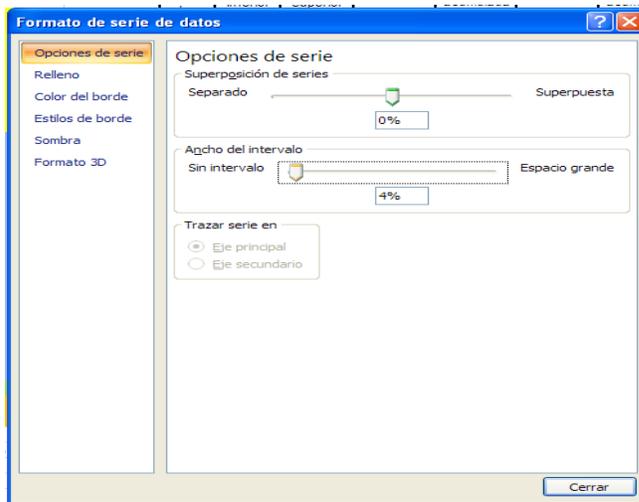
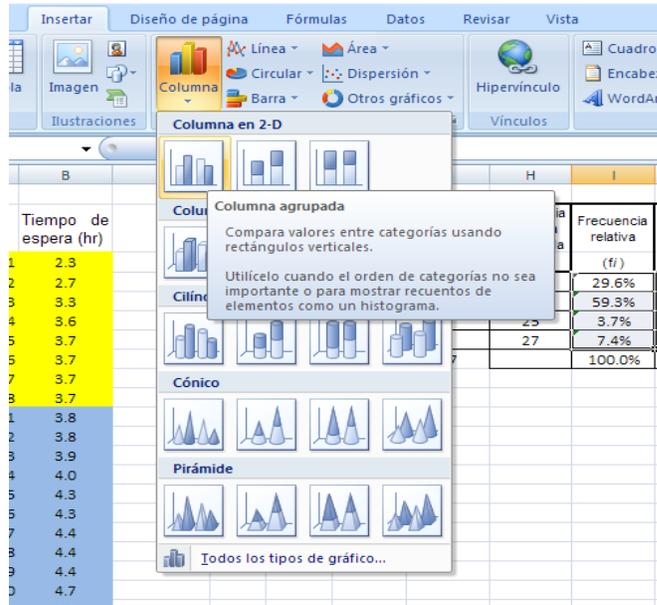
Para realizar el histograma del tiempo de espera de las personas que acudieron a la CDHDF en la Unidad Poniente, se utilizará la tabla de frecuencias del ejemplo anterior.

- 1) Con los datos de la tabla del cuadro 10 se seleccionan los valores de la frecuencia relativa (p_i)³⁴ de la tabla, columna 6, referentes a los intervalos de clase. No es necesario seleccionar la última fila, pues no es un intervalo, sino la suma de todos los valores anteriores.
- 2) Una vez hecho lo anterior, en el menú *Insertar* se selecciona la opción *Columna* y luego la primera opción: *Columna en 2D*. Automáticamente aparecerá la gráfica que se muestra en la parte derecha del primer panel de la figura 18.
- 3) La gráfica que se obtuvo tiene las barras separadas como en una variable de tipo cualitativa, y lo que se requiere es un histograma. Para que las barras ocupen todo el ancho del intervalo, se da clic sobre alguna de las barras; con el botón derecho se despliega el menú y se elige *Dar formato a la serie de datos*. Del submenú se selecciona *Ancho del intervalo* y se arrastrar el cursor completamente hacia el lado izquierdo *Sin intervalo*, y se oprime el botón *Cerrar*. De esta manera no hay intervalo ni espacio entre las barras y, por lo tanto, se convierte en un histograma, es decir, las barras son continuas y no hay valores de la tabla de datos que no se encuentren considerados en la gráfica (véase el segundo panel de la figura 18).
- 4) Para incluir los títulos, la fuente de la cual se obtuvo la información y la etiqueta de los datos,³⁵ se siguen los pasos del 2 al 4 del ejemplo 10. La gráfica resultante debe ser igual a la del panel 3 de la figura 18.

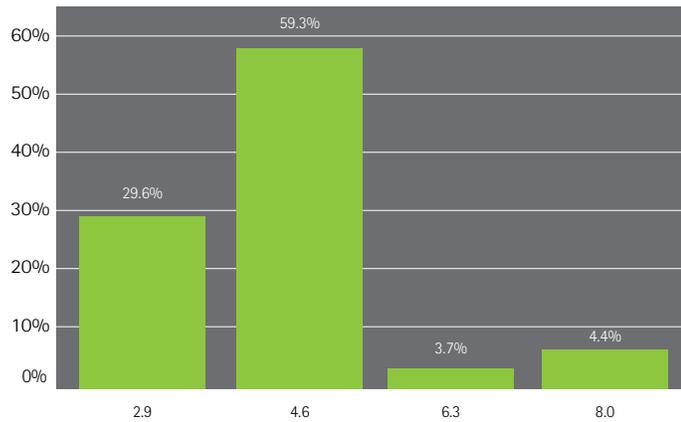
³⁴ Recuerde que también se puede hacer con la frecuencia absoluta, sólo que en lugar de seleccionar la columna de f_i se debe optar por la de p_i .

³⁵ Se recomienda usar las m_i como etiquetas del eje X.

Figura 18. Ejemplo de cómo generar un histograma de frecuencias relativas para una variable continua



Distribución de la frecuencia del tiempo que esperaron las personas para poder levantar una denuncia ante la CDHDF en la unidad Poniente

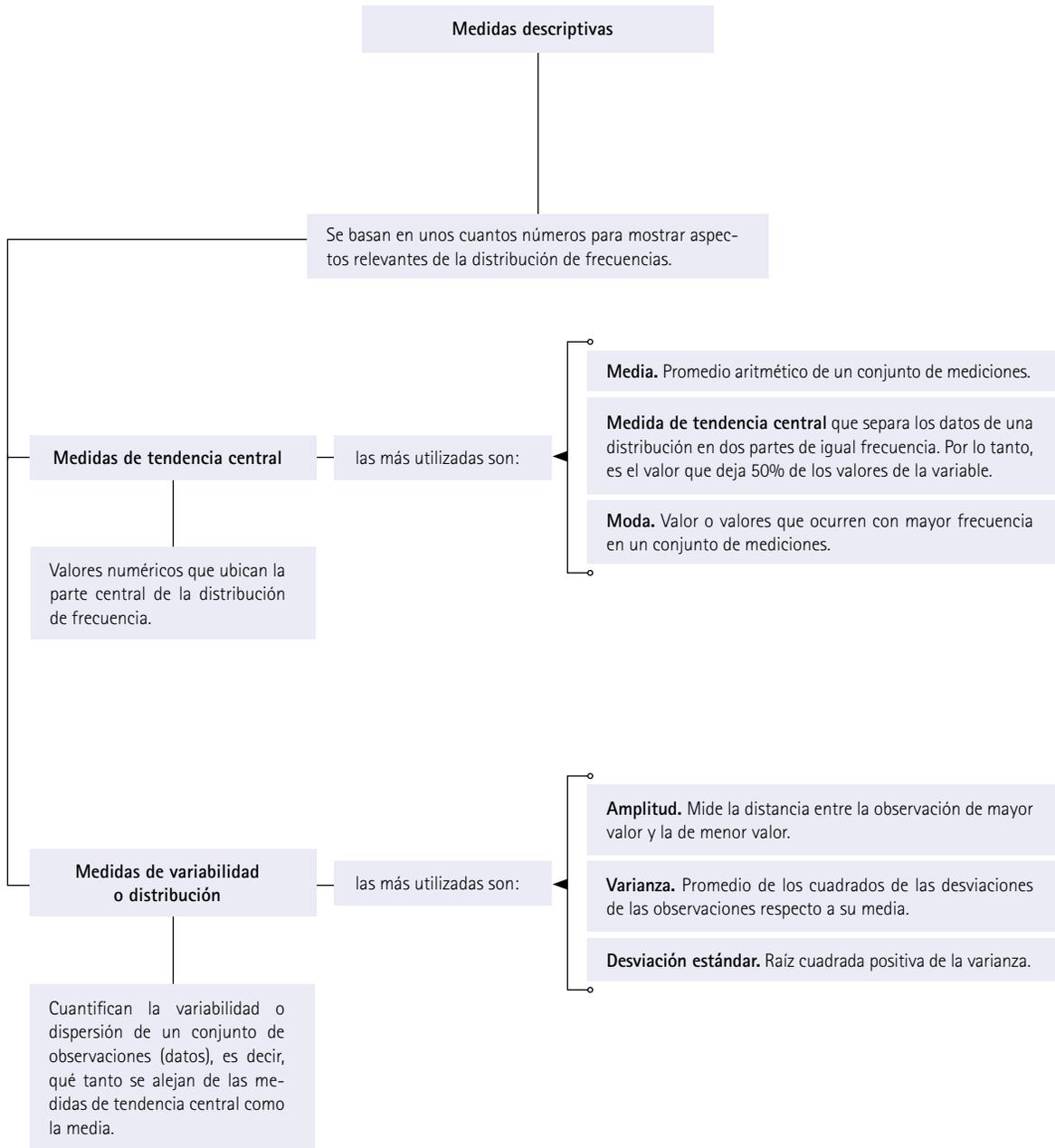


Fuente: elaboración propia con base en la información proporcionada por la Subdirección de Estadística de la CDHDF.

La interpretación es la misma que la de tabla de frecuencias, pero es más sencillo e informativo utilizar un histograma. De la gráfica anterior se deduce que el tiempo de espera de casi 60% de las personas corresponde a 4.6 horas, y que tan sólo 3.7% de las personas tuvieron un tiempo de espera de alrededor de 6.3 horas.

MÓDULO IV.

Algunos ejercicios de estadística descriptiva (medidas descriptivas)



Medidas descriptivas

Hasta el momento se ha visto cómo elaborar tablas de frecuencias y algunas gráficas, tanto para variables cualitativas como cuantitativas, las cuales resumen la información y ayudan a visualizar los valores que aparecen con mayor o menor frecuencia. A pesar de que esta descripción es bastante útil, hay descripciones de otro tipo, basadas en unos cuantos números, que muestran aspectos relevantes de la distribución de frecuencias. Por ejemplo, si lo que interesa es exponer la parte central de la distribución de frecuencias, deberán usarse medidas de tendencia central. Si el objetivo es conocer las variaciones de los datos observados respecto a un valor central de distribución, se utilizarán entonces medidas de variabilidad. A las medidas de tendencia central y de variabilidad se les conoce como medidas descriptivas.³⁶ Este tipo de medidas, con excepción de la moda, sólo deben aplicarse en variables cuantitativas.³⁷ A continuación se definirán y ejemplificarán las medidas descriptivas más utilizadas en la investigación cuantitativa.

³⁶ Víctor Aguirre y Begoña Artaloitia, *op. cit.*, pp. 1-24.

³⁷ Este tipo de medidas se obtienen tanto con datos agrupados como no agrupados; sin embargo, sólo se mostrará su aplicación con datos agrupados, ya que con el uso de las herramientas la manipulación de datos no agrupados es relativamente sencilla.

Medidas de tendencia central

Las medidas de tendencia central son valores numéricos que ayudan a localizar la parte central de la distribución de frecuencia; las más utilizadas son la media, la mediana y la moda.

Media

Es el promedio aritmético de un conjunto de mediciones y es la más común. Con los datos de una población, el promedio de todas sus mediciones es la media poblacional para una población de tamaño N . La media poblacional comúnmente se representa con μ y se obtiene mediante la siguiente fórmula:

$$\mu = \frac{\sum_{i=1}^N x_i}{N}$$

Con los datos de una muestra se utiliza la media muestral para una muestra de tamaño n , cuyo símbolo es \bar{X} y se obtiene de la siguiente manera:

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

La media es un indicador de la tendencia central en el caso de una distribución no muy sesgada y sin observaciones atípicas.³⁸

Mediana

Es la medida de tendencia central que separa los datos de una distribución en dos partes de igual frecuencia. Por lo tanto, la mediana es el valor que deja 50% de los valores de la variable, tanto por encima como por debajo de ella, siempre y cuando estén ordenados de manera ascendente o descendente. Por lo general, se indica con *Me*. La ventaja de esta medida de tendencia central es su cualidad de única, además de que los valores extremos (muy pequeños o muy grandes) no tienen efectos importantes en ella, lo que sí ocurre con la media.

Tanto la media como la mediana se encuentran en la misma unidad de medición de la variable. Por ejemplo, si se dispone de los datos del ingreso mensual en pesos de las personas que viven en la Delegación Iztapalapa, se sabrá que la media del ingreso mensual es de 2 359 pesos. Este resultado deberá leerse de la siguiente manera: el ingreso promedio mensual de las personas que residen en la Delegación Iztapalapa es de 2 359 pesos.

³⁸ Víctor Aguirre y Begoña Artaloitia, "Análisis exploratorio de datos", *op. cit.*, pp. 1-27.

Moda

Para los datos de tipo cuantitativo no agrupados es el valor o valores que ocurren con mayor frecuencia en un conjunto de mediciones. Y en el caso de los datos agrupados en intervalos, se llama clase modal a toda clase cuya frecuencia sea superior o igual a la de sus dos clases contiguas. Al igual que en la media y la mediana, y para las variables cuantitativas, la moda se debe interpretar en unidades de medición de la variable.

Asimismo, ésta es la única medida de tendencia central que se puede obtener para las variables cualitativas. Para encontrar este valor, se utilizará la tabla de frecuencias o algunas gráficas y, mediante la observación, se determinará el valor que más se repite:

- a) Para las variables cualitativas basta observar la clase con mayor frecuencia absoluta (o frecuencia relativa).
- b) También se obtiene observando la gráfica de barras: la moda será(n) aquella(s) clase(s) que tenga(n) la(s) barra(s) más alta(s). Con la gráfica de sectores, será la clase que ocupe mayor área en la gráfica.

Si este valor es único, se dice que la distribución de frecuencias es unimodal. Sin embargo, en la práctica se presentan casos en los que hay dos valores que se repiten más que sus vecinos, y entonces la distribución es *bimodal*. Si hay más, decimos que es *multimodal*. O, en los casos en que no existe la moda, no hay ningún valor que ocurra con mayor frecuencia.

Medidas de variabilidad o dispersión

Las medidas de variabilidad, también conocidas como de dispersión, cuantifican la variabilidad o dispersión de un conjunto de observaciones (datos), es decir, qué tanto se alejan los datos de las medidas de tendencia central como la media. Estas medidas de variabilidad serán pequeñas si no hay diferencias muy grandes entre los valores observados, pero resultarán grandes en caso contrario.³⁹

Entre las medidas de variabilidad más utilizadas se encuentran la amplitud, la varianza y la desviación estándar.

³⁹ *Ibidem*, pp. 1-30 a 1-31.

Amplitud

Es la más simple de todas y también se conoce como rango. Esta medida descriptiva mide la distancia entre la observación de mayor valor y la de menor valor, por lo que está en la misma unidad de medición que la variable utilizada para su obtención. Se saca mediante:

$$R = \text{Amplitud} = \text{valor máximo} - \text{valor mínimo.}$$

Una de sus ventajas es que se obtiene de manera fácil; sin embargo, al no considerar todos los datos, es muy inestable e insensible al tamaño de la muestra. Esta medida ayuda a comparar la dispersión entre variables que tengan la misma unidad de medición.

Varianza

Es el promedio de los cuadrados de las desviaciones de las observaciones respecto a su media. Se calcula para la población usando la siguiente fórmula:

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

Pero si los datos son de una muestra en vez de ser de una población, se aplica

$$S^2 = \frac{\sum (x_i - \bar{X})^2}{n-1} = \frac{\sum x_i^2 - n\bar{X}^2}{n-1}$$

Las unidades en que se expresa la varianza son el cuadrado de las unidades originales de la medición, de aquí su dificultad para interpretarla. Además, su valor se puede distorsionar por la presencia de valores extremos, muy pequeños o muy grandes.⁴⁰

Desviación estándar

Como ya se mencionó, la varianza tiene por unidades el cuadrado de las unidades de los datos, por ello es común usar su raíz cuadrada positiva, a la que se denomina desviación estándar. La finalidad es obtener una medida de dispersión con las mismas unidades que los datos. Su fórmula es:

⁴⁰ *Ibidem*, pp. 1-31 a 1-32.

$s = \sqrt{s^2}$ si tenemos datos de la población.

$S = \sqrt{S^2}$ si tenemos datos de la muestra.

Esta medida de dispersión mide la variabilidad en términos de la distancia entre cada dato y la media.

La varianza y la desviación estándar son estadísticas que se pueden distorsionar si hay datos atípicos, es decir, si las observaciones presentan datos muy pequeños o muy grandes.⁴¹

Ejemplo 14

Se analizarán algunas características de edad de 45 personas (mujeres y hombres) que presentaron una denuncia en la oficina central de la CDHDF (los datos se muestran en el cuadro 11).⁴² Las preguntas que se deberán responder son: la edad de las mujeres, en promedio, ¿es mayor que la de los hombres?, ¿cuál es edad modal del grupo de las mujeres y cuál es la de los hombres?, ¿la edad mediana de las mujeres es igual a la de los hombres?, ¿cuál de los dos grupos presenta mayor variabilidad?

Cuadro 11. Datos de edad de 45 mujeres y hombres que presentaron una denuncia en la oficina central de la CDHDF

Número	Edad	
	Mujeres	Hombres
1	26	59
2	58	24
3	28	59
4	48	23
5	16	58
6	22	47
7	27	37
8	47	93
9	13	61
10	46	26
11	41	23
12	41	36
13	41	77
14	48	52
15	55	38
16	48	22
17	55	44
18	19	28

⁴¹ *Idem.*

⁴² Las 45 personas serán la población con la que se harán las comparaciones de manera puntual. En caso de que se tratara de una muestra, se utilizarían las técnicas de la estadística inferencial.

Número	Edad	
	Mujeres	Hombres
19	51	33
20	42	29
21	72	25
22	48	34
23	37	

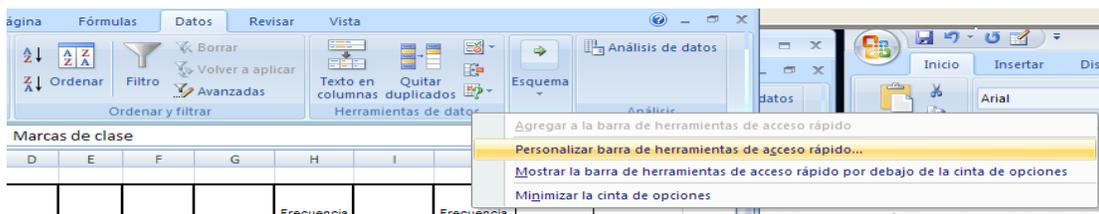
Fuente: elaboración propia con base en la información proporcionada por la Subdirección de Estadística de la CDHDF.

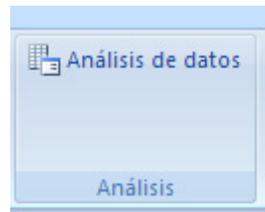
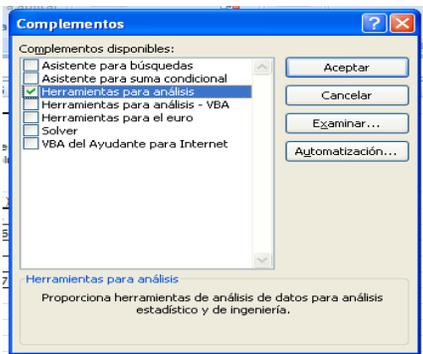
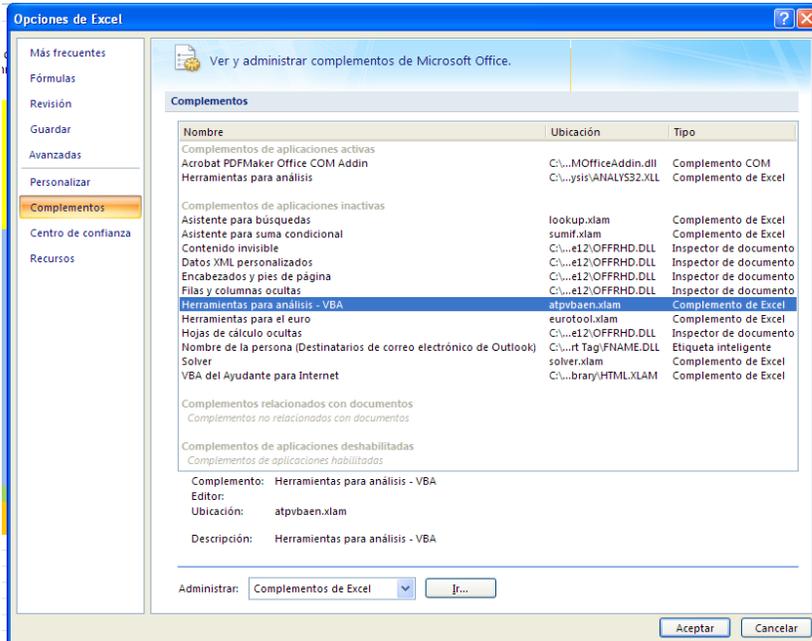
Para contestar los cuestionamientos anteriores, se deben obtener las medidas de tendencias central y la variabilidad de la variable *Edad* para cada grupo, es decir, para mujeres y para hombres. Una manera de realizarlo es mediante las herramientas de *Análisis de datos*, que permiten procedimientos de estadística descriptiva e inferencial de manera rápida y sencilla.

En ocasiones, Excel no trae cargado este menú. La incorporación de este conjunto de herramientas se hace siguiendo estos pasos:

- 1) Desde el menú *Datos*, hacer clic en el botón derecho del *mouse*. Aparecerá una ventana con cuatro opciones, elegir *Personalizar herramientas de acceso rápido*.
- 2) En la ventana *Opciones de Excel*, seleccionar la opción *Complementos*.
- 3) En el cuadro del centro, optar por *Herramientas para análisis*. En la parte inferior de esta ventana, oprimir el botón *Ir*. Aparecerá una ventana igual a la del tercer panel de la figura 19. Una vez allí, seleccionar *Herramientas para análisis* y oprimir *Aceptar*.
- 4) Con lo anterior se iniciará el proceso de instalación.
- 5) Al finalizar, deberá aparecer la opción de *Análisis de datos* en el menú *Datos*. Tal como aparece en la parte derecha del panel final de la figura 19.

Figura 19. Procedimiento para instalar las herramientas para análisis de Excel





Cuando tenemos incorporada a Excel la opción *Análisis de datos*, el procedimiento para obtener las medidas descriptivas y responder a las preguntas planteadas al principio de este ejemplo es el siguiente:

- 1) Los datos del cuadro 11 deberán estar en una hoja de Excel, tal como se ejemplifica en el primer panel de la figura 20. Observe que la primera columna sólo indica el número de casos y, con este número, se sabe que hay 23 observaciones de mujeres y 22 de hombres. Además, la información de la columna B es la edad de cada una de las mujeres que acudieron a la oficina central de la CDHDF a presentar una queja. Los datos de la columna C corresponden a la edad de los hombres.
- 2) Después, del menú *Datos*, se elige la opción *Análisis de datos*. Automáticamente aparece una ventana, seleccionar *Estadística descriptiva* y *Aceptar*.
- 3) Cuando aparezca una ventana como la del tercer panel de la figura 20, hay que ingresar los datos de entrada y salida. En la opción *Rango de entrada*, se oprime el ícono  y se seleccionan los datos

de las edades del grupo de las mujeres (celdas de B3 a B25). Se deja seleccionada la opción *Columnas*, ya que los datos se encuentran de esta forma. Como no se eligió el título de la columna, debemos dejar deseleccionada la opción *Rótulos en la primera fila* (véase el tercer panel de la figura 20).

- 4) Para este ejemplo es conveniente que los resultados aparezcan en esta hoja de trabajo, por ello, en *Opciones de salida* se selecciona la opción *Rango de salida* y se oprime el ícono . Con ayuda del *mouse* seleccione un conjunto de celdas. Finalmente, elija la opción *Resumen de estadísticas y Aceptar* (véase el tercer panel de la figura 20).
- 5) Automáticamente aparece una tabla con el título *Columna 1*, tal como se muestra en la parte derecha del tercer panel de la figura 20.
- 6) Para obtener las estadísticas descriptivas de los hombres, deben repetirse todos los pasos anteriores, pero esta vez seleccionando las observaciones de este grupo.
- 7) Finalmente, se incluyen los títulos, la fuente de donde se obtuvieron los datos y sólo se usarán las medidas descriptivas antes revisadas. La tabla final se muestra en el cuadro 12.⁴³

Figura 20. Ejemplo de cómo obtener las medidas de descriptivas utilizando la opción Análisis de datos de las herramientas para análisis de Excel

	A	B	C
1		Edad	
2	Número	Mujeres	Hombres
3	1	26	59
4	2	58	24
5	3	28	59
6	4	48	23
7	5	16	58
8	6	22	47
9	7	27	37
10	8	47	93
11	9	13	61
12	10	46	26
13	11	41	23
14	12	41	36
15	13	41	77
16	14	48	52
17	15	55	38
18	16	48	22
19	17	55	44
20	18	19	28
21	19	51	33
22	20	42	29
23	21	72	25
24	22	48	34
25	23	37	

⁴³ Para el resto de las medidas descriptivas que muestra Excel, véase Víctor Aguirre y Begoña Artaloitia, *op. cit.*, pp. 1-24 a 1-35, y Haroldo Elorza Pérez-Tejeda, *op. cit.*, pp. 39-71.

Cuadro 12. Estadísticas descriptivas de edad de 45 mujeres y hombres que presentaron una denuncia en la oficina central de la CDHDF

Medida descriptiva	Mujeres	Hombres
Media	40.39	42.18
Mediana	42.00	36.50
Moda	48.00	59.00
Desviación estándar	14.85	19.18
Varianza de la muestra	220.52	367.97
Rango	59.00	71.00
Mínimo	13.00	22.00
Máximo	72.00	93.00

Fuente: elaboración propia con base en la información proporcionada por la Subdirección de Estadística de la CDHDF.

En respuesta a las preguntas planteadas, se observa que la edad promedio de las mujeres que presentaron una denuncia en la oficina central de la CDHDF es de 40.4 años, mientras que, en promedio, los hombres tienen 42.2 años; de lo que se concluye que las mujeres son más jóvenes que los hombres. Lo anterior se corrobora con la edad modal, pues la de los hombres es 11 años mayor que la de las mujeres, con 59 y 48 años, respectivamente. Sin embargo, si se observa el valor de la edad mediana de ambos grupos, se ve que es menor la de los hombres, es decir, mientras que la mitad de las mujeres tiene 42 años, 50% de los hombres tiene 36.5 años, es decir, casi seis años menos. Esto se debe a que en ambos grupos de edad hay casos atípicos, personas muy jóvenes y personas de la tercera edad. Entre las mujeres, la edad mínima es de 13 años y la máxima de 72, lo que significa una amplitud de 59 años. En el grupo de los hombres, la edad mínima es de 22 años y la máxima de 93, con una amplitud de 71 años. Por lo tanto, no se puede afirmar que, en general, las mujeres que presentaron una denuncia sean más jóvenes que los hombres, pero sí se puede decir que, en promedio, las denunciadas son más jóvenes que los denunciados.

En cuanto a la variabilidad, en promedio, los hombres se alejan más de su media, con un valor de la desviación estándar de 19 años; en cambio, las mujeres se alejan un poco menos (casi 15 años). Por lo tanto, se concluye que entre los hombres que presentaron una denuncia en la oficina central de la CDHDF hay más variabilidad en la edad que entre las mujeres.

Si no se utiliza la opción *Análisis de Datos*, Excel tiene un conjunto de funciones estadísticas mediante las cuales se obtienen las medidas descriptivas de un conjunto de observaciones revisadas en este módulo. El siguiente cuadro muestra la manera como se debe escribir la descripción de cada una de ellas.

Cuadro 13. Descripción de algunas funciones estadísticas que proporciona Excel

Medida descriptiva	Descripción de lo que realiza Excel	Función en Excel
Media o promedio	Devuelve el promedio de los argumentos.	=promedio()
Mediana	Devuelve la mediana de los números dados.	=mediana()
Mínimo	Devuelve el valor mínimo de una lista de argumentos.	=min()
Máximo	Devuelve el valor máximo de una lista de argumentos.	=max()
Moda	Devuelve el valor más frecuente en un conjunto de datos.	=moda()
Desviación estándar	Calcula la desviación estándar basada en una muestra.	=desvest()
Varianza	Calcula la varianza de una muestra.	=var()

Ejemplo 15

De todas las gestiones recibidas en la CDHDF durante 2011, se analizará la moda de la variable *Sexo*, es decir, se verá si hay más gestiones levantadas por hombres, por mujeres, colectivas o de oficio.

Lo anterior se responde usando una tabla de frecuencias para variables cualitativas o mediante una representación gráfica de barras o de sectores. Al realizar todos los pasos de los ejemplos 7, 8 y 9 usando la variable *Sexo*, se obtienen el cuadro y las gráficas que se muestran a continuación.

Cuadro 14. Frecuencia absoluta y relativa del sexo de las personas que realizaron una gestión en la CDHDF durante 2011

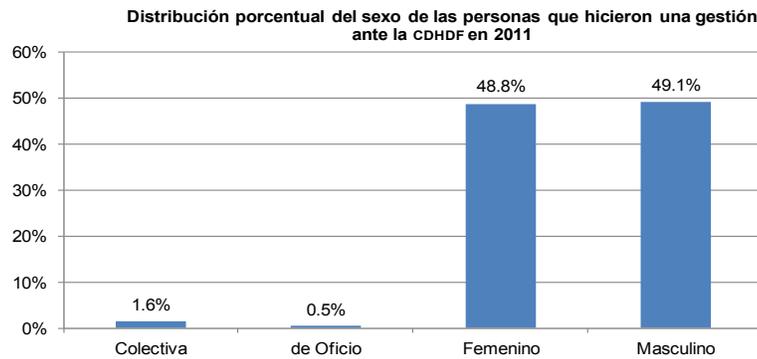
Sexo	Frecuencia absoluta (f_i)	Frecuencia relativa (p_i)
Colectiva	739	1.6%
De oficio	226	0.5%
Femenino	23 064	48.8%
Masculino	23 209	49.1%
Total	47 238	100%

Fuente: elaboración propia con base en la información proporcionada por la Subdirección de Estadística de la CDHDF.

Nótese que la clase o categoría con mayor f_i y p_i es *Masculino*. Por lo tanto, la moda de la variable *Sexo* de las gestiones recibidas por la Comisión en 2011 es *Masculino*.

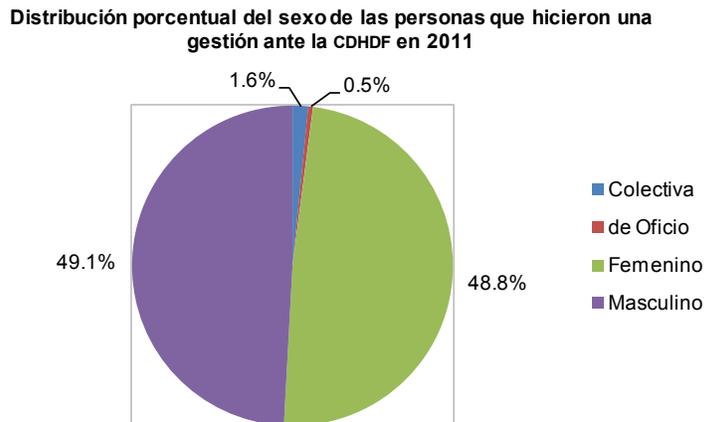
Lo anterior se corrobora con la gráfica de barras de la figura 21. A pesar de que la diferencia en la altura de las barras de las categorías *Femenino* y *Masculino* es muy pequeña, la más alta corresponde a los hombres. Además, se observa que la categoría *De oficio* se encuentra con menos frecuencia entre las gestiones realizadas en 2011 ante la Comisión. Así pues, la categoría modal de dichas gestiones es *Masculino*.

Figura 21. Ejemplo de cómo obtener la categoría modal de una variable cualitativa utilizando una gráfica de barras



De la misma manera, en la gráfica circular la mayor área corresponde a la categoría *Masculino*.

Figura 22. Ejemplo de cómo obtener la categoría modal de una variable cualitativa utilizando una gráfica circular



Fuente: elaboración propia con base en la información proporcionada por la Subdirección de Estadística de la CDHDF.

Ejemplo 16

¿Cuál es la moda del tiempo de espera de 27 personas para levantar una denuncia ante la CDHDF de la Unidad Poniente?

Para resolver este ejemplo se usarán los datos del ejemplo 12 y la función de Excel para obtener la moda. Los pasos son:

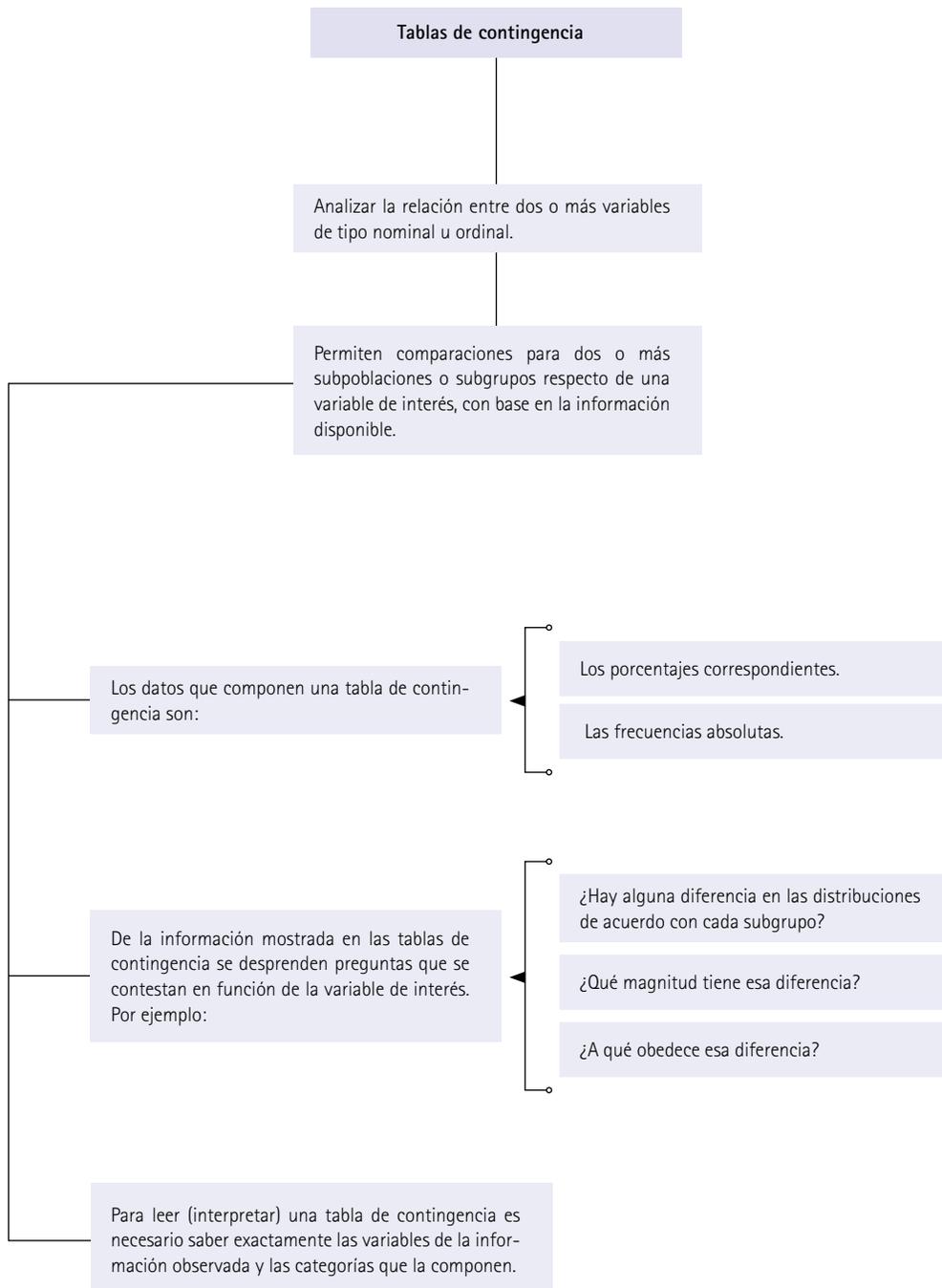
- Capturar las observaciones en una hoja de Excel.
- Para obtener la moda, posicionar el cursor en una celda, por ejemplo la D3, escribir =moda(B3, B29) y oprimir *Enter* (véase la figura 23).
- El resultado obtenido es 3.7, por lo que el tiempo de espera con mayor frecuencia entre las 27 personas que levantaron una denuncia ante la CDHDF en 2011 es de 3.7 horas.

Figura 23. Ejemplo de cómo obtener la moda de una variable cuantitativa

	A	B	C	D
1				
2		Tiempo de espera (hr)		
3	1	2.3		=moda(B3:B29)
4	2	2.7		
5	3	3.3		
6	4	3.6		
7	5	3.7		
8	6	3.7		
9	7	3.7		
10	8	3.7		
11	1	3.8		
12	2	3.8		
13	3	3.9		
14	4	4.0		
15	5	4.3		
16	6	4.3		
17	7	4.4		
18	8	4.4		
19	9	4.4		
20	10	4.7		
21	11	4.8		
22	12	4.9		
23	13	5.1		
24	14	5.1		
25	15	5.2		
26	16	5.4		
27	1	5.9		
28	1	8.2		
29	2	8.9		

MÓDULO V.

Tablas de contingencia



Las tablas de contingencia analizan la relación entre dos o más variables de tipo nominal u ordinal.⁴⁴ Por ejemplo, se examina la relación entre la actitud de un grupo de personas hacia la igualdad entre los sexos, y para hacerlo se divide la población en hombres y mujeres. De esa forma, se verá cómo se distribuyen las frecuencias entre quienes están a favor de la igualdad o en contra, y cómo se comportan por diferencia de sexo. Las tablas de contingencia son útiles para analizar la información de variables cualitativas. Con ello se hacen comparaciones entre dos o más subpoblaciones o subgrupos respecto de una variable de interés, con base en la información disponible. En el ejemplo del siguiente esquema se muestran dos variables de tipo nominal; la variable de interés es la *Actitud* de las personas a favor o en contra de la igualdad entre los sexos. Los subgrupos se determinan por las diferencias entre las dos categorías de la variable *Sexo*, es decir, mujeres u hombres.

⁴⁴ Como se vio en la sección *Datos, variables y escalas de medición en la estadística*, una variable nominal se refiere a datos que sólo pueden ser clasificados en categorías, mientras que una variable ordinal se agrupa en categorías de acuerdo con el grado de la magnitud del atributo que tienen. Respecto al uso de tablas de contingencia y del análisis a partir de éstas en la inferencia estadística, véase Juan Javier Sánchez Carrión, *Análisis de tablas de contingencia*, Centro de Investigaciones Sociológicas, Siglo XXI, Madrid, 1992. Las variables de tipo nominal y ordinal se explicaron en el módulo II.

Cuadro 15. Ejemplo de una tabla de contingencia

Sexo	Actitud (Igualdad entre los sexos)		Total por fila
	A favor	En contra	
Mujer	w	x	w + x
Hombre	y	z	y + z
Total por columna	w + y	x + z	Gran total = w + x + y + z

Fuente: elaboración propia.

Los datos de una tabla de contingencia pueden ser las frecuencias absolutas de la variable, o bien los porcentajes correspondientes. Si se tienen las frecuencias absolutas, se pueden obtener:

- Los porcentajes de cada casilla; a estos valores se les llama *frecuencias relativas conjuntas* y se calculan dividiendo el valor de cada casilla entre el valor ubicado en el *Gran total*, para luego multiplicarlo por 100. Estas frecuencias relativas dan una idea de la frecuencia con que se presentan simultáneamente ambos atributos en la población o muestra (véase el cuadro 16).
- Los porcentajes de los totales por fila o por columna; a estos valores se les llama *frecuencias relativas marginales*. Se denominan así porque se presentan en los márgenes de la tabla de contingencia para indicar la frecuencia con la que se presenta el atributo en la población por sí mismo. Para calcularlos, debemos dividir el valor total de categoría entre el valor ubicado en el *Gran total* y, después, se debe multiplicar por 100 (véase el cuadro 16).⁴⁵

Cuadro 16. Ejemplo de cómo obtener las frecuencias relativas conjuntas y marginales en una tabla de contingencia

Sexo	Actitud (Igualdad entre los sexos)		Total por fila
	A favor	En contra	
Mujer	$\frac{w}{(w + x + y + z)} \times 100$	$\frac{x}{(w + x + y + z)} \times 100$	$\frac{(w + x)}{(w + x + y + z)} \times 100$
Hombre	$\frac{y}{(w + x + y + z)} \times 100$	$\frac{z}{(w + x + y + z)} \times 100$	$\frac{(y + z)}{(w + x + y + z)} \times 100$
Total por columna	$\frac{(w + y)}{(w + x + y + z)} \times 100$	$\frac{(x + z)}{(w + x + y + z)} \times 100$	Gran total = $\frac{(w + x + y + z)}{(w + x + y + z)} \times 100$

Fuente: elaboración propia.

⁴⁵ Víctor Aguirre y Begoña Artaloitia, *op. cit.*, pp. 1-37 a 1-38.

Las frecuencias relativas condicionales para cada variable se obtienen al dividir las frecuencias conjuntas entre la correspondiente frecuencia marginal (véanse los cuadros 16 y 17).⁴⁶

Cuadro 17. Ejemplo de cómo obtener las frecuencias relativas condicionales para la variable Sexo

Sexo	Actitud (Igualdad entre los sexos)		Totales por fila
	A favor	En contra	
Mujer	$\frac{w}{(w+x)} \times 100$	$\frac{x}{(w+x)} \times 100$	$\frac{(w+x)}{(w+x)} \times 100 = 100$
Hombre	$\frac{y}{(y+z)} \times 100$	$\frac{z}{(y+z)} \times 100$	$\frac{(y+z)}{(y+z)} \times 100 = 100$
Total por columna	$\frac{(w+y)}{(w+x+y+z)} \times 100$	$\frac{(x+z)}{(w+x+y+z)} \times 100$	Gran Total= $\frac{(w+x+y+z)}{(w+x+y+z)} \times 100 = 100$

Fuente: elaboración propia.

Cuadro 18. Ejemplo de cómo obtener las frecuencias relativas condicionales para la variable Actitud

Sexo	Actitud (igualdad entre los sexos)		Totales por fila
	A favor	En contra	
Mujer	$\frac{w}{(w+y)} \times 100$	$\frac{x}{(x+z)} \times 100$	$\frac{(w+x)}{(w+x+y+z)} \times 100$
Hombre	$\frac{y}{(w+y)} \times 100$	$\frac{z}{(x+z)} \times 100$	$\frac{(y+z)}{(w+x+y+z)} \times 100$
Totales por columna	$\frac{(w+y)}{(w+y)} \times 100 = 100$	$\frac{(x+z)}{(x+z)} \times 100 = 100$	Gran Total= $\frac{(w+x+y+z)}{(w+x+y+z)} \times 100$

Fuente: elaboración propia.

De la información mostrada en las tablas de contingencia, regularmente se formulan preguntas que pueden responderse en función de la variable de interés, como, ¿hay alguna diferencia en las distribuciones de acuerdo con cada subgrupo?, ¿qué tan grande es la diferencia?, ¿a qué obedece esa diferencia?

⁴⁶ *Idem.*

Para leer (interpretar) de manera correcta una tabla de contingencia es necesario saber exactamente cuáles son las variables de la información que se observa y las categorías que la componen. En nuestro ejemplo, la variable *Sexo* se compone de dos categorías: *Hombre* y *Mujer*, y la variable *Actitud* también: *A favor* y *En contra*. Posteriormente, se observan los totales respecto de cada subgrupo de población y se leen los resultados en función de la variable de interés, por ejemplo, qué porcentaje de los que están a favor son hombres respecto de quienes están a favor y son mujeres. Lo mismo se lleva a cabo para ambos grupos en la opción *En contra*.

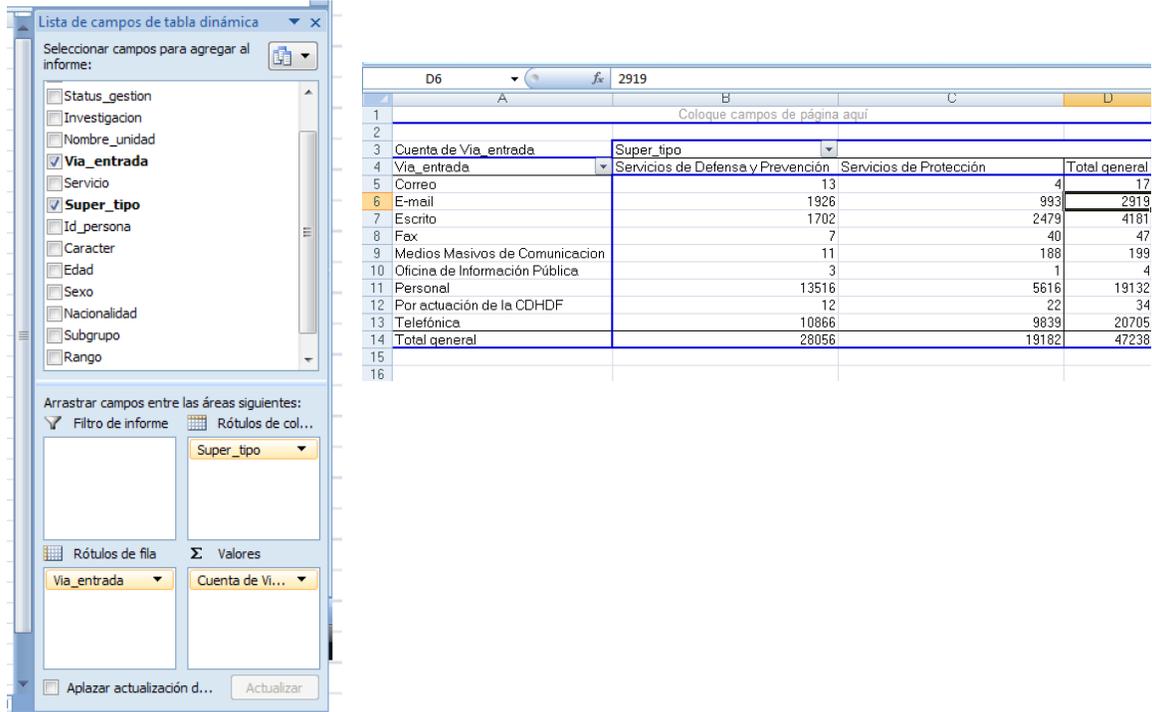
A continuación se muestra un ejemplo de cómo elaborar una tabla de contingencia en Excel utilizando una tabla dinámica.

Ejemplo 17

Para comparar la vía de entrada de las gestiones recibidas en la CDHDF durante 2011 con el tipo de servicio de la gestión: de defensa y prevención o de protección. Para saber si estas dos características de las gestiones se relacionan, se obtiene una tabla de contingencia con las variables de tipo nominal *Vía de entrada* y *Super_tipo*, así como sus respectivas frecuencias relativas conjuntas, marginales y condicionales. Los pasos a seguir, utilizando el archivo de trabajo "Tabla de servicios", son:

- 1) Seguir los siete primeros pasos del ejemplo 8 para obtener una tabla dinámica de la variable *Vía de entrada*. Ésta será la variable de los renglones o filas. Observe que tiene nueve categorías (véase primer panel de la figura 24).
- 2) La variable de las columnas será *Super_tipo*, con dos categorías. Para colocarla en la tabla, se arrastra con el *mouse* a la ventana *Rótulos de columna*. Automáticamente aparece una tabla de contingencia de nueve renglones (filas) y dos columnas (véase el segundo panel de la figura 24).
- 3) Para dar un formato adecuado, se escribe el título de la tabla, de la variable columna, de la variable renglón y de la fuente de información. El resultado final se muestra en el cuadro 19.

Figura 24. Ejemplo de cómo obtener una tabla de contingencia utilizando una tabla dinámica en Excel



Cuadro 19. Distribución del tipo de vía de entrada de la gestión y del tipo de servicio de las gestiones recibidas en la CDHDF durante 2011

Vía de entrada de la gestión	Tipo de servicio		Total
	Servicios de defensa y prevención	Servicios de protección	
Correo	13	4	17
Correo electrónico	1 926	993	2 919
Escrito	1 702	2 479	4 181
Fax	7	40	47
Medios masivos de comunicación	11	188	199
Oficina de información pública	3	1	4
Personal	13 516	5 616	19 132
Por actuación de la CDHDF	12	22	34
Telefónica	10 866	9 839	20 705
Total	28 056	19 182	47 238

Fuente: elaboración propia con base en la información proporcionada por la Subdirección de Estadística de la CDHDF.

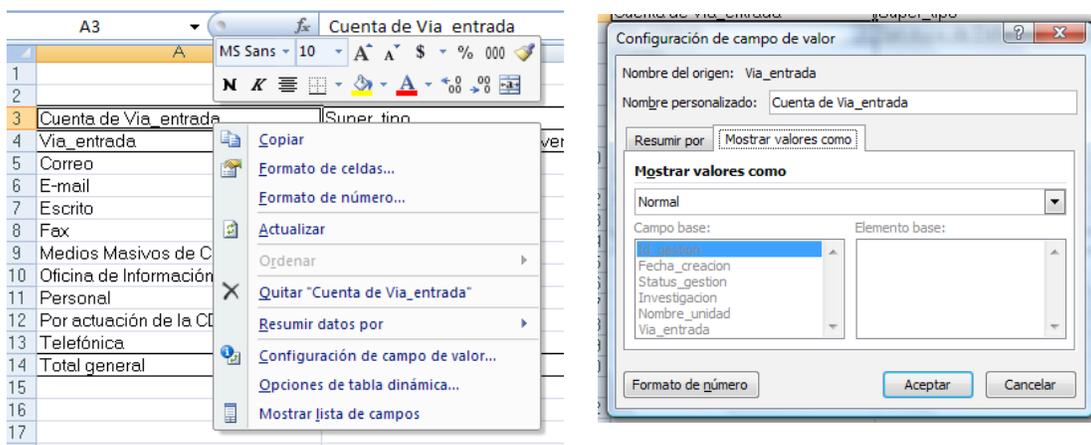
En el cuadro 19 se observa que la combinación más frecuente entre las gestiones recibidas en la Comisión son las de Servicios de defensa y prevención con vía de entrada personal, mientras que las menos frecuentes son las de Servicios de protección con vía de entrada en alguna oficina de información pública. Esto se sabe porque la primera combinación tiene el mayor

número en la tabla de contingencia (13 516 gestiones), y la segunda tiene el menor número (1 gestión). En los totales por columna se observa que la categoría modal del tipo de servicio es Servicios de defensa y prevención, puesto que presenta un número mayor de observaciones (28 056 gestiones). Mientras que la categoría modal de la Vía de entrada es la Telefónica, con 20 705 gestiones.

Para obtener las frecuencias relativas conjuntas y las marginales los pasos son:

- 1) Al dar clic derecho en la celda *Cuenta de Vía_entrada* aparecerá un menú como el que se muestra en el primer panel de la figura 25. Al seleccionar la opción *Configuración de campo de valor* aparecerá un recuadro titulado *Configuración de campo de valor*. En este último, seleccione la pestaña *Mostrar valores como* y observe que en la ventana aparece la leyenda *Normal* (véase la figura 25).
- 2) Al dar clic en la flecha que está junto a *Normal* aparecerá un menú con diferentes opciones; seleccione *% de total* y oprima *Aceptar*. Esto conduce a la tabla de frecuencias relativas conjuntas y marginales que se muestra en el segundo panel de la figura 25.
- 3) Si le ponemos el título a la tabla, a la variable *Columna*, a la variable *Renglón* y a la fuente de información, el resultado es el que se muestra en el cuadro 20.

Figura 25. Ejemplo de cómo obtener las frecuencias relativas conjuntas y las marginales en una tabla de contingencia utilizando una tabla dinámica en Excel



Vía de entrada	Super_tipo	Servicios de Defensa y Preven	Servicios de Protección	Total general
Correo		0.03%	0.01%	0.04%
E-mail		4.08%	2.10%	6.18%
Escrito		3.60%	5.25%	8.85%
Fax		0.01%	0.08%	0.10%
Medios Masivos de Comunicacion		0.02%	0.40%	0.42%
Oficina de Información Pública		0.01%	0.00%	0.01%
Personal		28.61%	11.89%	40.50%
Por actuación de la CDHDF		0.03%	0.05%	0.07%
Telefónica		23.00%	20.83%	43.83%
Total general		59.39%	40.61%	100.00%

Cuadro 20. Distribución porcentual del tipo de vía de entrada de la gestión y del tipo de servicio de las gestiones recibidas en la CDHDF durante 2011

Vía de entrada de la gestión	Tipo de servicio		Total (%)
	Servicios de defensa y prevención (%)	Servicios de protección (%)	
Correo	0.028	0.008	0.036
Correo electrónico	4.077	2.102	6.179
Escrito	3.603	5.248	8.851
Fax	0.015	0.085	0.099
Medios masivos de comunicación	0.023	0.398	0.421
Oficina de información pública	0.006	0.002	0.008
Personal	28.613	11.889	40.501
Por actuación de la CDHDF	0.025	0.047	0.072
Telefónica	23.003	20.829	43.831
Total	59.393	40.607	100

Fuente: elaboración propia con base en la información proporcionada por la Subdirección de Estadística de la CDHDF.

Con las frecuencias relativas conjuntas y marginales se obtienen las mismas conclusiones que con las frecuencias absolutas, sólo que los porcentajes permiten una interpretación en términos de 100% de las gestiones recibidas en 2011 en la CDHDF. Por ejemplo:

- De cada 100 gestiones recibidas en la Comisión, 29 fueron para atender Servicios de defensa y prevención y la vía de entrada fue personal, lo que la convierte en la mayor frecuencia relativa conjunta.

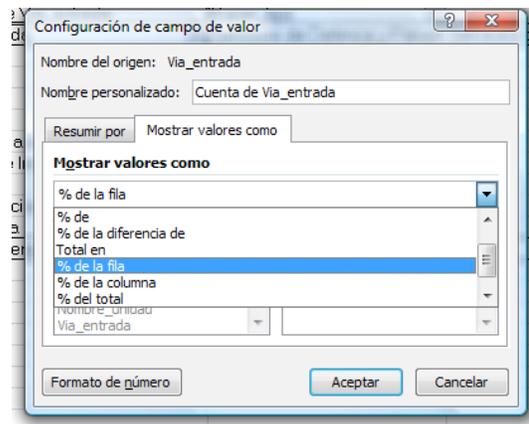
- Las gestiones de Servicios de defensa y prevención atendidas de manera telefónica equivalen a 23% del total de las gestiones recibidas en la Comisión durante 2011.
- Sólo 0.002% de las gestiones fueron de Servicios de protección, las cuales ingresaron al sistema de la CDHDF mediante una oficina de información pública.

Asimismo, con las frecuencias relativas marginales, las gestiones se distribuyen en 100% en cada una de las categorías de cada variable (de la columna y del renglón), por ejemplo:

- Del total de gestiones recibidas, y sin tomar en cuenta la variable *Tipo de servicio*, la categoría *Telefónica* fue la más frecuente (44%); le sigue la *Personal*, ya que 4 de cada 10 gestiones recibidas entraron con esta modalidad. La vía de entrada menos frecuente es la de *Oficina de información pública*.
- Por su parte, sin diferenciar entre la vía de entrada de las gestiones, poco más de 59% del total de las gestiones recibidas en 2011 en la CDHDF son de *Servicios de defensa y prevención* y el resto de *Servicios de protección*.

Como se observa, con los resultados del cuadro 20 se obtienen las mismas conclusiones que con los resultados del cuadro 19. A pesar de que los resultados anteriores son útiles para conocer la distribución de las gestiones, quizá el interés se centre en saber si el tipo de servicio varía relativamente con la vía de entrada de las gestiones. Para averiguarlo se debe condicionar la variable *Vía de entrada* respecto al *Tipo de servicio*. Para obtener esta tabla en Excel se siguen los tres pasos anteriores, pero en lugar de seleccionar *% de total* debemos escoger *% de la fila* (véase la figura 26); el resultado se muestra en el cuadro 21.

Figura 26. Ejemplo de cómo obtener la frecuencia relativa condicionada de la variable de la fila en una tabla de contingencia utilizando una tabla dinámica en Excel



Cuadro 21. Distribución porcentual del tipo de vía de entrada de la gestión según el tipo de servicio de las gestiones recibidas en la CDHDF durante 2011

Vía de entrada de la gestión	Tipo de servicio		Total (%)
	Servicios de defensa y prevención (%)	Servicios de protección (%)	
Correo	76.47	23.53	100
Correo electrónico	65.98	34.02	100
Escrito	40.71	59.29	100
Fax	14.89	85.11	100
Medios masivos de comunicación	5.53	94.47	100
Oficina de información pública	75.00	25.00	100
Personal	70.65	29.35	100
Por actuación de la CDHDF	35.29	64.71	100
Telefónica	52.48	47.52	100
Total	59.39	40.61	100

Fuente: elaboración propia con base en la información proporcionada por la Subdirección de Estadística de la CDHDF.

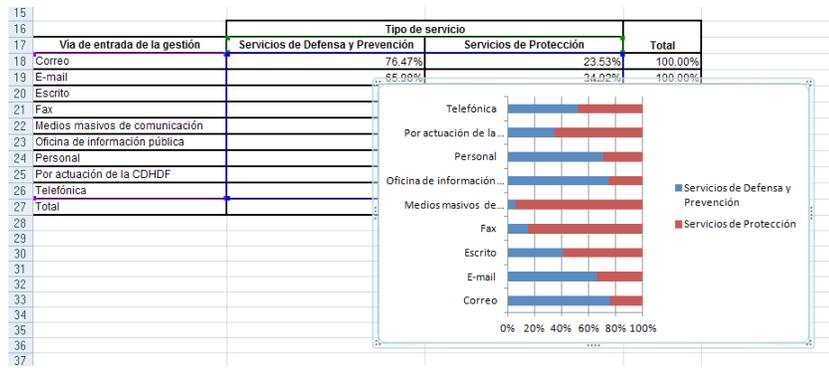
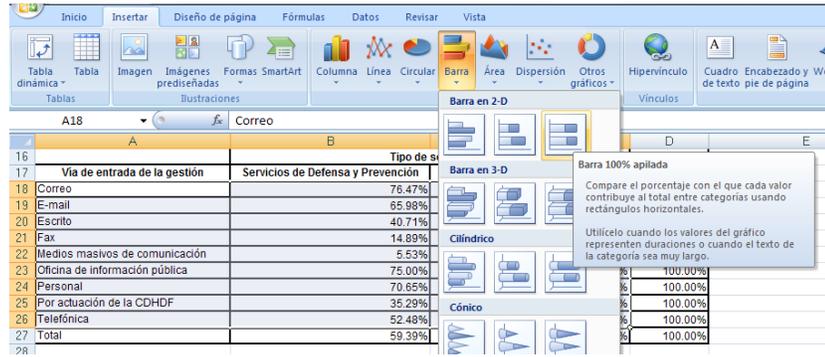
El cuadro 21 debe leerse por renglones (filas), ya que en esa dirección va la suma de 100%. Por ejemplo, del total de gestiones recibidas en la CDHDF durante 2011 por correo, la mayoría (77%) fueron de Servicios de defensa y prevención; en cambio, de las gestiones recibidas vía fax, poco más de 85% fueron de Servicios de protección. Un dato que llama la atención es que de cada 100 gestiones recibidas por medios masivos de comunicación, 95 son de Servicios de protección.

La información del cuadro 21 se puede transferir a una gráfica de barras apiladas a 100% para una interpretación más sencilla. El procedimiento en Excel es el siguiente:

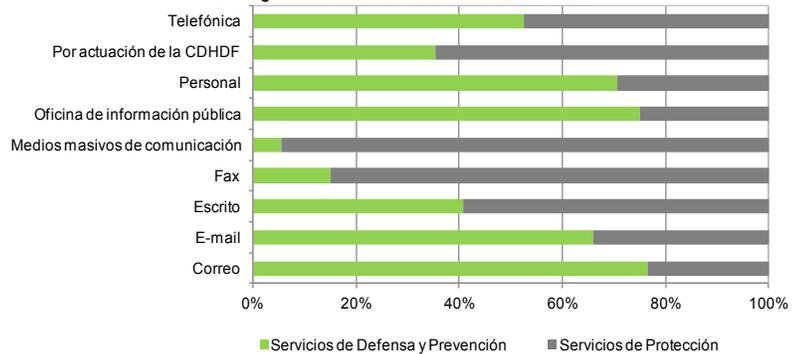
- 1) Seleccionar todas las celdas de la columna, incluidas las etiquetas de cada categoría, excepto la categoría de *Total*.
- 2) En el menú *Insertar* se selecciona la opción *Barra*.⁴⁷ Aparecerá un menú, en el cual se elige la tercera opción: *Barra 100% apilada*. Automáticamente aparecerá la gráfica que se muestra en el segundo panel de la figura 27.
- 3) Para añadir el título, la fuente de la cual se obtuvo la información y la etiqueta de los datos, hay que seguir del paso 2 al 4 del ejemplo 8. Puede suprimirse la guía de la serie de datos seleccionando y oprimiendo la tecla *Suprimir*. La gráfica resultado es la que aparece al final de la figura 27.

⁴⁷ La opción *Columna* en este caso no es recomendable debido a que se tienen muchas categorías y con nombres extensos; es mejor usar una gráfica de barras horizontales.

Figura 27. Ejemplo de cómo obtener una gráfica de barras horizontal apiladas a 100% en Excel



Distribución porcentual del tipo de vía de entrada de la gestión según el tipo de servicio de las gestiones recibidas en la cdhdf durante el 2011



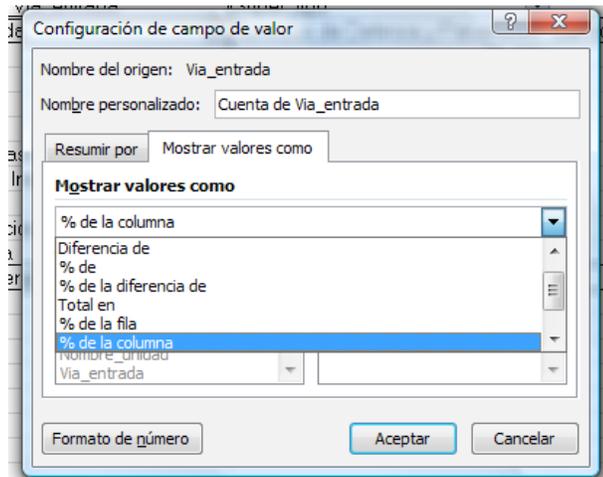
Fuente: elaboración propia con base en la información proporcionada por la Subdirección de Estadística de la CDHDF.

Desde la gráfica anterior se puede concluir que, al parecer, hay cierta relación entre la forma en que se reciben las gestiones en la CDHDF y el tipo de servicio. Si se agrupan los medios a los cuales las personas recurrieron, como correo, correo electrónico y personal, se observa que la mayoría de las gestiones pertenece a la categoría *Servicios de defensa y prevención*. Pero si se toman las vías de entrada, Medios masivos de comunicación, Por actuación de la CDHDF y Fax, que pueden considerarse medios de comunicación colectivos, los tipos de

gestión más frecuentemente recibidos son Servicios de protección. Finalmente, no hay una tendencia clara en las vías de entrada Escrito y Telefónica, ya que las gestiones se reparten de manera muy similar entre los dos tipos de servicios.

Para saber si el tipo de servicio varía respecto a la vía de entrada, debe obtenerse la tabla de frecuencias relativas siguiendo los tres primeros pasos de la primera tabla de contingencia, pero debe optarse por *% de columna* en lugar de *% del total* (véase la figura 28). El cuadro resultante es el número 22.

Figura 28. Ejemplo de cómo obtener la frecuencia relativa condicionada de la variable de la columna en una tabla de contingencia utilizando una tabla dinámica en Excel



Cuadro 22. Distribución porcentual del tipo de servicio de las gestiones según la vía de entrada de las gestiones recibidas en la CDHDF durante 2011

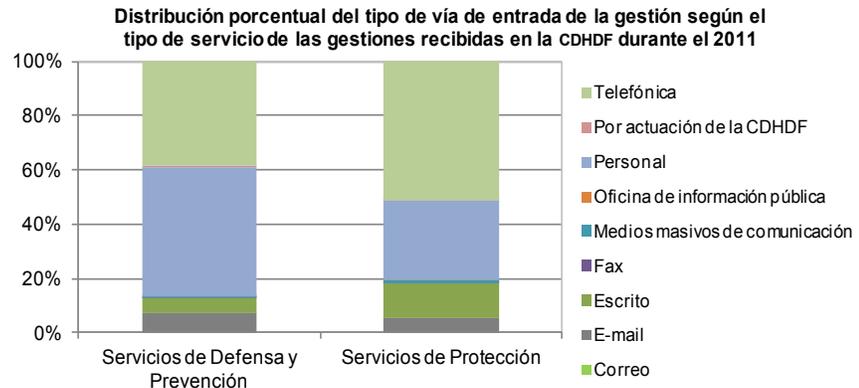
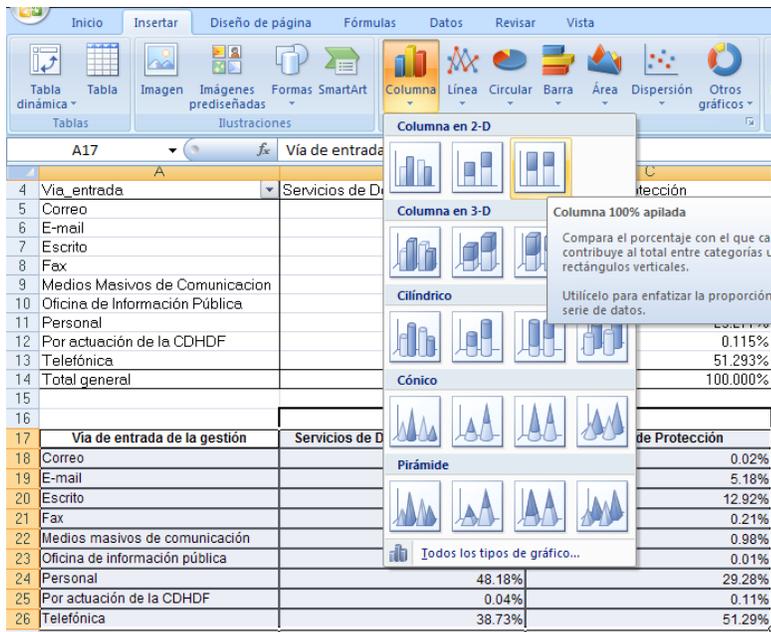
Vía de entrada de la gestión	Tipo de servicio		Total (%)
	Servicios de defensa y prevención (%)	Servicios de protección (%)	
Correo	0.05	0.02	0.04
Correo electrónico	6.86	5.18	6.18
Escrito	6.07	12.92	8.85
Fax	0.02	0.21	0.10
Medios masivos de comunicación	0.04	0.98	0.42
Oficina de información pública	0.01	0.01	0.01
Personal	48.18	29.28	40.50
Por actuación de la CDHDF	0.04	0.11	0.07
Telefónica	38.73	51.29	43.83
Total	100	100	100

Fuente: elaboración propia con base en la información proporcionada por la Subdirección de Estadística de la CDHDF.

Como en el caso anterior, se obtiene la gráfica de barras apilada a 100%. En este caso, la opción son las barras verticales, porque es más fácil interpretar los resultados. Los pasos para hacerlo en Excel son:

- 1) Seleccione todas las celdas de la columna, incluidas las etiquetas de cada categoría, excepto la del *Total*.
- 2) En el menú *Insertar* se selecciona *Columna*. Aparecerá otro menú, en el cual se debe escoger *Columna 100% apilada*.
- 3) Al igual que en la gráfica anterior, se le da formato y el resultado es el que aparece al final de la figura 29.

Figura 29. Ejemplo de cómo obtener una gráfica de barras vertical apiladas a 100% en Excel



Fuente: elaboración propia con base en la información proporcionada por la Subdirección de Estadística de la CDHDF.

La gráfica anterior debe leerse por columnas, ya que en esa dirección se suma 100%. Con esta gráfica se concluye que en ambos tipos de servicios analizados, las vías de entrada más utilizadas son la Personal y la Telefónica.

BIBLIOGRAFÍA

- Aguirre, Víctor, y Begoña Artaloitia, "Análisis exploratorio de datos", en Víctor Aguirre *et al.*, *Fundamentos de probabilidad y estadística*, 2ª ed., México, Jit Press, 2007.
- Ángel Gutiérrez, Julio C., *Estadística general aplicada*, Medellín, Fondo Editorial Universidad Eafit (col. Académica), 2007.
- Angulo Reyes, Yvón, "Muestreo", en Haroldo Elorza Pérez-Tejeda, *Estadística para las ciencias sociales del comportamiento y de la salud*, 3ª ed., México, Cengage Learning, 2007.
- Asher, Jana, *et al.*, *Statistical Methods for Human Rights*, Nueva York, Springer, 2007.
- Burgos Matamoros, Mylai, "Fundamentos básicos para la metodología de la investigación aplicada", en *Programa de capacitación y formación profesional en derechos humanos*, México, Comisión de Derechos Humanos del Distrito Federal (serie Cuadernos de capacitación para el SPDH), 2011.
- Christensen, Howard B., *Estadística paso a paso*, 3ª ed., México, Trillas, 2008.
- Elorza Pérez-Tejeda, Haroldo, *Estadística para las ciencias sociales del comportamiento y de la salud*, 3ª ed., México, Cengage Learning, 2007.
- Fernández Durán, Juan, y Rubén Hernández, "Muestreo", en Víctor Aguirre *et al.*, *Fundamentos de probabilidad y estadística*, 2ª ed., México, Jit Press, 2007.
- Hopkins, Kenneth D., *et al.*, *Estadística básica para las ciencias sociales y del comportamiento*, 3ª ed., México, Prentice Hall, 1997.
- Icart Isern, María Teresa, *et al.*, *Elaboración y presentación de un proyecto de investigación y una tesina*, Barcelona, Universitat de Barcelona, 2006.
- Jabine, Thomas, *Human Rights and Statistics, Getting the Record Straight*, Filadelfia, University of Pennsylvania Press, 1992.
- Martínez González, Miguel A. y Jorkin de Irala Estévez, "Procedimientos descriptivos", en *Bioestadística amigable*, 2ª ed., Madrid, Díaz de Santos, 2006.
- Megía Morales, Paco, *Ya sé Excel, pero necesito más*, 2ª ed., Madrid, Visión, 2007.
- Ross, Sheldon, *Introducción a la estadística*, Barcelona, Reverté, 2007.
- Sánchez Carrión, Juan Javier, *Análisis de tablas de contingencia*, Centro de Investigaciones Sociológicas, Siglo XXI, Madrid, 1992.

Schedler, Andreas, *et al.*, *The Self-restraining State, Power and Accountability in New Democracies*, Boulder, Lynn Rienner, 1999.

Subdirección de Estadística de la CDHDF, Sistema Integral de Gestión de Información (Siigesii), México, marzo de 2011.

EJERCICIOS

Ejercicio 1

Objetivo: repasar los conceptos más importantes del curso.

Instrucciones: relacione los conceptos de la primera columna con las definiciones de la segunda.

Concepto	Definición
1. Población	a) Es aquella en la que los valores de la variable se agrupan en categorías, las cuales permiten establecer una relación de orden entre las categorías, de acuerdo con el grado de posesión del atributo que tienen.
2. Muestra	b) En este tipo de variable el cero indica la ausencia total de la cantidad a medir y no es un punto arbitrario de referencia.
3. Unidad de análisis	c) Es la que hace referencia a datos que sólo pueden ser clasificados en categorías. No es posible establecer una relación de orden entre ellas, solamente se puede decir si el elemento pertenece a la categoría o no, o si tiene o no el atributo.
4. Variable de escala nominal	d) Es la recopilación y ordenamiento de los atributos (datos) referentes a los individuos de la población objetivo.
5. Base de datos	e) En este tipo de variables, además de afirmar que un elemento posee más o menos cantidad de variable que otro, también es posible determinar la magnitud de la diferencia entre dos elementos medidos.
6. Inferencia estadística	f) Conjunto de elementos delimitados en un espacio y tiempo determinados. Dentro de una investigación, se define al especificar qué o cuáles elementos son de interés y qué características deben tener para formar parte del estudio.
7. Variable de tipo ordinal	g) Es la entidad (persona, grupo, institución, país, etc.) sobre la cual se realizará el análisis estadístico y que se desea estudiar para sacar conclusiones al respecto. La unidad de análisis es la entidad que sintetiza e identifica la población y en función de la cual se definirá el tamaño de la muestra.
8. Variable de escala de intervalo	h) Es un proceso por el cual se atribuyen a la población las características más significativas obtenidas en la muestra.
9. Variable de escala de razón	i) Es aquella que se toma como base para recoger la información que sirve de sustento para las mediciones estadísticas y que brindará insumos en forma de datos con los que se realizará el estudio de nuestra unidad de análisis.
10. Unidad de observación	j) Es un subconjunto de la población. Debe contener un número determinado de elementos para ser representativa de la población.

Respuesta 1	Respuesta 2	Respuesta 3	Respuesta 4	Respuesta 5

Respuesta 6	Respuesta 7	Respuesta 8	Respuesta 9	Respuesta 10

Instrucciones: para obtener lo que se pide en los siguientes ejercicios, utilice la tabla dinámica obtenida del archivo de Excel llamado "Tabla de servicios" con la cual se trabajó a lo largo del curso.

Ejercicio 2

Objetivo: obtener estadísticas descriptivas y representaciones gráficas de variables cualitativas.

- 1) Para saber cuál es la forma principal en la que se presentaron las gestiones en 2011 ante la CDHDF, es decir, cuál es la principal vía de entrada de ellas. Responda y obtenga:
 - a) ¿Qué tipo de variable es *Via_entrada*?
 - b) Obtenga la tabla de frecuencia de la variable.
 - c) Interprete los resultados.

- 2) Realice una gráfica de sectores en la que se aprecie la proporción de investigaciones recibidas en cada una de las sedes de la CDHDF. Para esto responda y obtenga:
 - a) ¿Qué tipo de variable es *Nombre_unidad*?
 - b) Obtenga la tabla de frecuencias de la variable.
 - c) Obtenga la gráfica de sectores. No olvide poner el título y fuente a la gráfica.
 - d) Interprete los resultados obtenidos.

- 3) Realice una gráfica de barras para determinar la frecuencia absoluta del carácter de las gestiones realizadas durante 2001 ante la CDHDF. Para esto responda y obtenga:
 - a) ¿Qué tipo de variable es *Caracter*?
 - b) Obtenga la tabla de frecuencias de la variable.
 - c) Obtenga la gráfica de barras. No olvide poner el título y la fuente a la gráfica.
 - d) Interprete los resultados obtenidos.

Ejercicio 3

Objetivo: transformación de una variable de intervalo en una variable ordinal.

- 1) Sin tomar en cuenta los casos con edad no especificada, es decir, casos con valor de 9999 y sin información, transforme la variable *Edad* de cuantitativa de intervalo en cualitativa ordinal agrupando los valores en grupos quinquenales (cada cinco años). Realice también la tabla de frecuencias y la gráfica de barras de esta nueva variable. Interprete los resultados y no olvide poner el título y la fuente de la gráfica.

Ejercicio 4

Objetivo: obtener estadísticas descriptivas y representaciones gráficas de variables cuantitativas.

- 1) La siguiente tabla representa la frecuencia absoluta del tiempo en días, en el que fueron resueltas 157 de las gestiones presentadas ante la oficina central de la CDHDF en 2011. Complete la información de la tabla (p_i , F_i , P_i y m_i) y realice un histograma de frecuencias relativas. Interprete los resultados y no olvide poner el título y la fuente de la gráfica.

Cuadro 1. Distribución del tiempo (en días) en que fueron resueltas las gestiones presentadas en la oficina central de la CDHDF en 2011

Intervalo	Límite inferior (Linf)	Límite superior (Lsup)	Frecuencia absoluta (f_i)	Frecuencia absoluta acumulada (F_i)	Frecuencia relativa (p_i)	Frecuencia relativa acumulada (P_i)	Marcas de clase (m_i)
1	0	10	5				
2	10	20	20				
3	20	30	40				
4	30	40	45				
5	40	50	38				
6	50	60	9				
Total			157				

Fuente: elaboración propia con base en la información simulada.

- 2) Interesa conocer algunas características de la edad de las personas que presentaron su denuncia vía correo electrónico y por escrito ante la Unidad Norte de la CDHDF; los datos se muestran en la tabla 2. Entre las preguntas que se quiere responder están: ¿cuál es la edad promedio, modal y la edad mediana de ambos grupos?, ¿cuál de los dos grupos presenta mayor variabilidad? Interprete los resultados comparando los valores de ambos grupos.

Cuadro 2. Datos de la edad de las personas que presentaron una denuncia en la Unidad Norte de la CDHDF

Número	Edad de las personas que presentaron su queja vía:	
	Escrito	Correo-e
1	52	85
2	27	85
3	39	85
4	34	30
5	7	87
6	61	40
7	57	46
8	40	19
9	33	
10	57	

Cuadro 2. Datos de la edad de las personas que presentaron una denuncia en la Unidad Norte de la CDHDF (*continuación*)

Número	Edad de las personas que presentaron su queja vía:	
	Escrito	Correo-e
11	66	
12	52	
13	46	
14	52	
15	83	
16	76	
17	80	
18	80	
19	76	
20	46	
21	46	
22	44	
23	46	
24	37	
25	53	
26	46	
27	44	
28	33	

Fuente: elaboración propia con base en la información proporcionada por la Subdirección de Estadística de la CDHDF.

Ejercicio 5

Objetivo: obtener e interpretar una tabla de contingencia.

- 1) Compare el sexo de las personas que iniciaron una gestión ante la CDHDF durante 2011 y diga si el tipo de servicio de la gestión es de Defensa y prevención o de Protección. Para saber si hay relación entre estas dos características de las gestiones, obtenga una tabla de contingencia con las variables *Sexo* y *Super_tipo*, con sus respectivas frecuencias relativas conjuntas, marginales y condicionales. Interprete los resultados y no olvide poner el título y la fuente de la tabla.

AUTOEVALUACIÓN

1. Rama de las matemáticas que se encarga de recopilar datos, describirlos y analizarlos para obtener conclusiones acerca de la población objeto de estudio.
 - a) Estadística descriptiva.
 - b) Estadística.
 - c) Estadística inferencial.
 - d) Estadística experimental.

2. ¿Cuál(es) de los siguientes enunciados es (son) verdadero(s)?
 - I. Las variables cualitativas sólo pueden tener atributos de escalas nominales u ordinales.
 - II. Las variables cuantitativas sólo pueden ser continuas.
 - III. Las variables cuantitativas sólo pueden ser de escala o de intervalo.
 - a) Sólo III.
 - b) Sólo I y II.
 - c) Sólo I y III.
 - d) Sólo II y III.

3. Para cada una de las siguientes variables, si su escala de medición es nominal, coloque "N"; en caso de ser ordinal, una "O"; si es de intervalo, identifíquelo con una "I", y si es razón, con una "R":
 - I. Sexo.
 - II. Ingreso mensual.
 - III. Edad en años cumplidos.
 - IV. Fecha en la cual se levanta una queja.
 - V. Tipo de servicio de la gestión.
 - VI. Número de quejas gestionadas en un mes.
 - VII. Tiempo que se tardan en registrar una queja.
 - VIII. Años aprobados de escolaridad.
 - XI. Identificador de la gestión.
 - a) I-N; II-R; III-I; IV-O; V-N; VI-R; VII-R; VIII-I; IX-O.
 - b) I-N; II-R; III-I; IV-R; V-N; VI-O; VII-I; VIII-I; IX-N.
 - c) I-N; II-N; III-N; IV-O; V-O; VI-R; VII-R; VIII-I; IX-I.
 - d) I-N; II-R; III-I; IV-N; V-N; VI-R; VII-O; VIII-I; IX-O.

4. ¿Cuál(es) de los siguientes enunciados es (son) verdadero(s)?
 - I. Población es un conjunto de elementos delimitados en un espacio y tiempo determinados.
 - II. Todas las muestras son de tipo probabilísticas.
 - III. Todos los elementos de la muestra son también elementos de la población.
 - IV. La unidad de análisis es la entidad que sintetiza e identifica la población y en función de la cual se deberá definir el tamaño de la muestra.

- a) Sólo I, II y III.
- b) Sólo II, III y IV.
- c) Sólo III y IV.
- d) Sólo I, III y IV.

5. Es la recopilación y ordenamiento de los atributos (datos) referentes a los individuos de la población objetivo.

- a) Población.
- b) Base de datos.
- c) Muestra.
- d) Tabla dinámica.

6. Es un arreglo sistemático que tiene como objetivo organizar y resumir la información de una variable ya sea cualitativa o cuantitativa.

- a) Tabla de frecuencias.
- b) Base de datos.
- c) Muestra.
- d) Intervalo de clase.

7. ¿Cuál(es) de los siguientes enunciados es (son) verdadero(s)?

- I. La moda siempre es única.
- II. La moda es una medida de tendencia central que puede ser obtenida tanto para variables de tipo cualitativo como cuantitativo.
- III. La moda en una variable puede no existir.

- a) Sólo III.
- b) Sólo I y II.
- c) Sólo I y III.
- d) Sólo II y III.

8. ¿Cuál(es) de los siguientes enunciados es (son) verdadero(s)?

- I. Las gráficas de un conjunto de datos facilitan la comprensión y extracción de conclusiones acerca del comportamiento de las variables.
- II. Un histograma es un diagrama útil para representar distribuciones de frecuencia cuya variable es continua y viene dada en intervalos o clases.
- III. De las representaciones gráficas más usadas para las variables cualitativas son las gráficas circulares y las gráficas de barras.

- a) Sólo I.
- b) Sólo II.
- c) Sólo III.
- d) I, II y III.

9. Las medidas de tendencia central más utilizadas en la estadística descriptiva son:

- a) Amplitud y desviación estándar.
- b) Media, moda y mediana.
- c) Máximo y mínimo.
- d) Todas las anteriores

10. Sirve para analizar la relación entre dos o más variables de tipo nominal u ordinal

- a) Tabla de frecuencias.
- b) Media.
- c) Tabla de contingencia.
- d) Desviación estándar.

CLAVE DE RESPUESTAS

1	b
2	c
3	a
4	d
5	b
6	a
7	d
8	d
9	b
10	c

*Fase de formación especializada.
El empleo de la estadística aplicada
a la investigación en derechos humanos,*
se terminó de editar en octubre de 2012.
Para su composición se utilizaron los tipos
Futura y Rotis Sans Serif.

En el marco del Programa de Derechos Humanos y Medio Ambiente
y comprometida con la ecología y el cuidado del planeta,
la Comisión de Derechos Humanos del Distrito Federal
edita este material en versión electrónica para
reducir el consumo de recursos naturales, la generación
de residuos y los problemas de contaminación.

**COMISIÓN DE DERECHOS HUMANOS
DEL DISTRITO FEDERAL**

Oficina sede

Av. Universidad 1449,
col. Florida, pueblo de Axotla,
del. Álvaro Obregón,
01030 México, D. F.
Tel.: 5229 5600

Unidades desconcentradas

Norte

Payta 632
col. Lindavista,
07300 México, D. F.
Tel.: 5229 5600, ext.: 1756

Sur

Av. Prol. Div. del Norte 5662,
Local B, Barrio San Marcos,
del. Xochimilco,
16090 México, D. F.
Tel.: 1509 0267

Oriente

Cuauhtémoc 6, 3er piso,
esquina con Ermita,
Barrio San Pablo,
del. Iztapalapa,
09000 México, D. F.
Tels.: 5686 1540, 5686 1230
y 5686 2087

Centro de Consulta y Documentación

Av. Universidad 1449,
edificio B, planta baja,
col. Florida, pueblo de Axotla,
del. Álvaro Obregón,
01030 México, D. F.
Tel.: 5229 5600, ext.: 1818

www.cd hdf.org.mx

