

JEAMMY JULIETH SIERRA HERNÁNDEZ

Estadística I

Análisis de datos para la toma
de decisiones informadas en
la administración

Estadística I

Catalogación en la publicación - Biblioteca Luis Oswaldo Beltrán Jara - ESAP

Sierra Hernández, Jeammy Julieth

Estadística I : análisis de datos para la toma de decisiones informadas en la administración /
Jeammy Julieth Sierra Hernández : Bogotá : Escuela Superior de Administración Pública-ESAP,
2024.

192 páginas: figuras y tablas. - (Colección Didáctica)

ISBN 978-958-609-151-0 (papel). - ISBN 978-958-609-152-7 (electrónico)

1. Estadística. 2. Estadísticas matemáticas 3. Estadísticas para administradores-análisis de
datos 4. Administración-Métodos estadísticos

I. Título

CDD-21: 519.5

*Estadística I: Análisis de datos para la toma de decisiones informadas en la
administración*

Jeammy Julieth Sierra Hernández

Facultad de Pregrados

Colección Didáctica

ISBN 978-958-609-151-0 (papel)

ISBN 978-958-609-152-7 (electrónico)

Primera edición, 2024

© Escuela Superior de Administración Pública

Director Nacional: Jorge Iván Bula

Subdirección Nacional de Servicios Académicos

Grupo de Publicaciones

Editorial ESAP

grupo.publicaciones@esap.edu.co

<https://www.esap.edu.co/>

<https://libros.esap.edu.co/>

<https://revistas.esap.edu.co/>

Editores académicos de la colección Jesús María Molina Giraldo y Marcela Castañeda
González

Coordinación editorial Óscar A. Chacón Gómez

Corrección de estilo Lorena Castro e Íkaro Valderrama

Diagramación Yully Cortés

Escuela Superior de Administración Pública (ESAP)

Grupo Publicaciones. Calle 44 #53-37, Bogotá, D. C.

(+57) 601 795 6110



Creative Commons Atribución-NoComercial-CompartirIgual 4.0
Internacional (CC BY-NC-SA 4.0)

Hecho en Bogotá, D. C., Colombia, 2024

Estadística I

Análisis de datos para la toma
de decisiones informadas en la
administración

JEAMMY JULIETH SIERRA HERNÁNDEZ



Este libro es fruto del empeño de la Facultad de Pregrados de la Escuela Superior de Administración Pública, del programa de Administración Pública Territorial, y de los diálogos y aportes de los docentes universitarios que respaldaron la propuesta de renovación curricular y escritura de libros de texto universitario entre 2017 y 2021.

Contenido

Presentación

17

Situación problema	20
Lectura de motivación	22

**Conceptos
básicos**

19

Historia e importancia de la estadística	22
Conceptualizando	25
Importancia de los datos en el sector público	38
Autoevaluación: ¡Practica tus conocimientos!	44
Recursos de aprendizaje	46
Referencias	47
Bibliografía comentada	50

**Distribución
de frecuencias**

53

Situación problema	54
Lectura de motivación	56
Actividad de presaberes	57
Organización y análisis de la información	57
Recursos de aprendizaje	86
Referencias	88
Bibliografía comentada	91

**Medidas de
tendencia central**

93

Situación problema	94
Lectura de motivación	96
Actividad de presaberes	98
Medidas de tendencia central o medidas de posición	98
Autoevaluación: ¡Practica tus conocimientos!	119
Recursos de aprendizaje	120
Referencias	123
Bibliografía comentada	127

Variación y forma

131

Situación problema	132
Actividad de presaberes	133
Lectura de motivación	134
Medidas de dispersión	135
Autoevaluación: ¡Practica tus conocimientos!	143
Autoevaluación: gasto en salud per cápita	151
Recursos de aprendizaje	152
Referencias	154
Bibliografía comentada	154

Números índice

157

Situación problema	158
Lectura de motivación	159
Actividad de presaberes	162
Números índices	163
Periodo base o referencia	164
Periodo actual	164
Índices simples	164
Índices compuestos	170
Autoevaluación: ¡Practica tus conocimientos!	178
Recursos de aprendizaje	178
Referencias	179
Bibliografía comentada	181
Glosario	183
Derechos sobre las imágenes	187
Sobre la autora	189

Lista de figuras

Figura 1. Ejes temáticos de la Unidad 1	21
Figura 2. Etimología de la palabra estadística	23
Figura 3. Línea de tiempo de la estadística	24
Figura 4. Fases de la estadística	25
Figura 5. Cuadro comparativo estadística descriptiva e inferencial	26
Figura 6. Población de hospitales	26
Figura 7. Población de graduados de la Escuela Superior de Administración Pública (ESAP)	27
Figura 8. Muestra de hospitales	27
Figura 9. Esquema de diseño de muestreo	28
Figura 10. Escalas de medición de las variables	31
Figura 11. Variables numéricas: discretas versus continuas	32
Figura 12. Ejemplos de variables dicotómicas	34
Figura 13. Ejemplo variable politómica (grupo étnico)	34
Figura 14. Ejemplo variable ordinal de rango de oficiales de Armada Nacional de Colombia	35
Figura 15. Variables de recuento	37

Figura 16. Fuentes de datos	39
Figura 17. Estadísticas delictivas	40
Figura 18. Estructura del Sistema Estadístico Nacional de Colombia (SEN)	42
Figura 19. Ejes temáticos de la Unidad 2.	55
Figura 20. Planificación de la encuesta	59
Figura 21. Diseño de la investigación	60
Figura 22. Tipos de preguntas que se pueden incluir en un cuestionario	65
Figura 23. Ejemplo de cuestionario	66
Figura 24. Elementos de una tabla de frecuencia para datos categóricos	68
Figura 25. Salida análisis de datos NBI (%) municipios del Chocó según DANE (2018)	77
Figura 26. Tabla de frecuencias NBI (%) en Chocó. Fuente: DANE (2018)	77
Figura 27. Esquema para construir tabla de frecuencia de variable numérica	78
Figura 28. Marca de clase en tabla de frecuencia de variable numérica	78
Figura 29. Configuración tabla de frecuencia de variable numérica con rango de clases.	79
Figura 30. Construcción de la frecuencia absoluta “Fa” de las NBI (%) con límites ajustados.	79
Figura 31. Construcción de la frecuencia relativa Fr de las NBI (%) con límites ajustados	80
Figura 32. Construcción de la frecuencia porcentual “F%” de las NBI (%) con límites ajustados	80
Figura 33. Resultado final tabla de frecuencias NBI (%) con límites ajustados. Chocó	80

Figura 34. Tabla de frecuencias NBI (%) completa con límites ajustados	81
Figura 35. Ejemplo de diagrama circular o tortas	82
Figura 36. Ejemplo de diagrama de barra para datos categóricos	82
Figura 37. Ejemplo de diagrama de barra para datos numéricos	83
Figura 38. Ejemplo de histograma de barra para datos numéricos continuos	84
Figura 39. Ejemplo de diagrama de línea de número de países con primera posición del poder ejecutivo a cargo de mujeres (1960-2022).	84
Figura 40. Elementos de un gráfico	85
Figura 41. Ejes temáticos de la Unidad 3	96
Figura 42. Media geométrica PIB (Excel)	101
Figura 43. Media geométrica en Excel (caso licitación)	102
Figura 44. Distribución de datos en cuartiles	106
Figura 45. Distribución de datos en deciles	107
Figura 46. Medidas de tendencia central (o posición) en Excel	108
Figura 47. Cuestionario Encuesta Calidad de Vida-Fuerza laboral	109
Figura 48. <i>Boxplot</i> o diagrama de caja y bigotes	116
Figura 49. <i>Boxplot</i> de ingresos tributarios en los municipios de Nariño (Colombia), 2021	116
Figura 50. <i>Boxplot</i> de ingresos tributarios Nariño y Santander (Colombia), 2021	116
Figura 51. Ejes temáticos de la Unidad 4	133
Figura 52. Casos de covid-19 en Colombia durante el 2022	134
Figura 53. Comparación presupuesto ejemplo 4.5	139

Figura 54. Salida de Excel medidas de dispersión (TRM enero 2022)	143
Figura 55. Interfaz inicial de Rstudio (a)	144
Figura 56. Interfaz inicial de Rstudio (b)	145
Figura 57. Creación de un vector	146
Figura 58. Asignación de objetos y variables en Rstudio	146
Figura 59. Cargar datos en Rstudio	147
Figura 60. Medidas de resumen en Rstudio	147
Figura 61. Gráficos en Rstudio	148
Figura 62. Histograma en Rstudio	148
Figura 63. Tabla de frecuencias en Rstudio	149
Figura 64. Salvar archivo en Rstudio	149
Figura 65. Importación de archivos en Rstudio	151
Figura 66. Ejes temáticos de la Unidad 5	163
Figura 67. Producción mensual de café en Colombia 2018-2022	166
Figura 68. Evolución mensual de las órdenes de comparendos a nivel nacional en Colombia	169

Lista de tablas

Tabla 1. Diferentes formas de clasificación de las variables	30
Tabla 2. Ejemplos de variables numéricas: discretas y continuas	32
Tabla 3. Ejemplo de variables categóricas	36
Tabla 4. Tipo de datos	38
Tabla 5. Definición de variables	62
Tabla 6. Población desplazada en Colombia según género	68
Tabla 7. Ingresos de clientes del Banco A	69
Tabla 8. Tabla de doble entrada	70
Tabla 9. Tabla de doble entrada (Tipo de sangre y género)	71
Tabla 10. Delitos sexuales	71
Tabla 11. NBI en el departamento del Chocó (población sin ningún grupo étnico)	74
Tabla 12. Gráficos sugeridos por tipo de variable	81
Tabla 13. Métodos de ponderación de la oferta económica en el proceso de contratación pública	97
Tabla 14. Demanda de agua en Colombia	99
Tabla 15. Variación anual del PIB Colombia, 2017-2021	101
Tabla 16. Nacidos vivos de madres de 10-14 años	103
Tabla 17. Déficit cuantitativo habitacional Colombia 2021	106

Tabla 18. Función percentil en Excel	108
Tabla 19. Nacimientos según edad de la madre	112
Tabla 20. Nacimientos según edad de la madre	114
Tabla 21. Recursos Fonpet de tres departamentos	117
Tabla 22. Meses buscando empleo	136
Tabla 23. Desviación media del tiempo buscando empleo (meses)	137
Tabla 24. Observaciones con igual media	138
Tabla 25. Medidas de resumen datos ejemplo 4.6	139
Tabla 26. Medidas de resumen datos agrupados (tiempo desempleado)	140
Tabla 27. TRM enero del 2022	142
Tabla 28. Medidas de dispersión en Excel	143
Tabla 29. Límites del IDH	160
Tabla 30. Valores dimensiones del IDH Colombia frente a mínimos y máximos en el mundo	160
Tabla 31. Índice relativo simple de base fija (producción de café)	167
Tabla 32. TRM en Colombia durante julio de 2022	168
Tabla 33. Índices simples de la TRM en Colombia durante julio de 2022	168
Tabla 34. Índices de eslabón (producción de café)	170
Tabla 35. Costos de programa social adulto mayor	172
Tabla 36. Índice compuesto relativo no ponderado (costos programa social adulto mayor)	172
Tabla 37. Tipos de índices compuestos no ponderados	174
Tabla 38. Índices (café versus tomate)	176
Tabla 39. Índice compuesto de café y tomate	177

Presentación

Estadística I. Análisis de datos para la toma de decisiones informadas en la administración es un libro de la Facultad de Pregrados de la Escuela Superior de Administración Pública (ESAP) y del Programa Administración Pública Territorial, fruto de los diálogos y aportes de los docentes universitarios que respaldaron la propuesta de renovación curricular y escritura de libros de texto universitario, desarrollados entre el 2017 y el 2021.

El presente texto académico busca fomentar en el estudiante del campo de la administración, y en especial en la comunidad alrededor de la administración pública, el uso de las herramientas estadísticas.

Al explorar cada unidad, se podrán abordar temáticas de gran interés, por ejemplo las problemáticas sociales, que requieren un análisis profundo. Este contenido reviste valor para el futuro administrador público, puesto que, más allá de recolectar información y describirla, permite, a través del estudio estadístico, obtener conclusiones acerca de las poblaciones (grupos de instituciones públicas, barrios, comunas, municipios, etc.) para actuar en concordancia con los principios de una asignación y manejo óptimo de los recursos.

Se trata de impulsar la comprensión de metodologías y herramientas de la estadística descriptiva para su aplicación investigativa en situaciones económico-administrativas que impliquen obtener y procesar volúmenes de datos para la toma de decisiones, la planificación,

y la solución de problemas administrativos, y así dar al lector la capacidad de resolver situaciones problemáticas.

El documento consta de cinco unidades: conceptos básicos, distribución de frecuencias, medidas de tendencia central, variación y forma, y números índice. En cada una de estas unidades el lector podrá encontrar un texto de motivación para que tenga una perspectiva de la aplicación de los diferentes temas estadísticos en el contexto real. De igual forma, cada uno de los cinco apartados tiene un mapa temático que permite reconocer cómo están articulados los temas en cada unidad. Se resalta que, a lo largo del libro, se encuentran ejemplos y se proponen prácticas que permiten desarrollar habilidades operacionales y manejo de aplicaciones, como Excel, propias de las habilidades del siglo XXI que el futuro administrador público debe dominar.

Conceptos básicos

EN ESTA UNIDAD, SE PRETENDE QUE EL LECTOR RECONOZCA LOS aspectos generales de la estadística, abordando desde los orígenes de la disciplina hasta conceptos propios de su metalenguaje. Esto, en razón a que la estadística cumple un papel importante en la formación del administrador público, pues contribuye a la creación de políticas públicas a partir del análisis de las problemáticas y retos en las comunidades. La estadística permite describir de manera objetiva las características, condiciones y problemáticas del entorno de las comunidades, y también la toma de decisiones acordes a los fenómenos y la realidad social. Ahora bien, debido al enorme aumento de datos, el proceso de decisión está sometido a múltiples escenarios que pueden ser analizados en una forma más simplificada desde la estadística.

Actualmente, los responsables de las decisiones tienen que estar suficientemente familiarizados con el manejo de información, herramientas tecnológicas y técnicas propias de la estadística, pues esta disciplina proporciona la rigurosidad del método científico, y permite interpretar e inferir conclusiones a partir de grandes cantidades de información, al organizarla y presentarla de forma clara. El recorrido por la unidad de conceptos básicos facilita al estudiante elementos lo suficientemente técnicos y con la mayor claridad posible, a través de la ejemplificación en cada caso, para iniciarse en el estudio de la estadística.

Algunos conceptos que se encuentran en esta sección son: estadística, población, muestra, unidad de análisis, unidad de observación, variable, muestreo, muestreo probabilístico y no probabilístico. También

se abordan la importancia de los datos en la administración pública y las principales fuentes de información, así como algunos sitios donde puede encontrar estadísticas oficiales o descargar bases de datos en la sección de recursos. En Colombia, el Departamento Administrativo Nacional de Estadística (Dane) es la entidad responsable de la planeación, levantamiento, procesamiento, análisis y difusión de las estadísticas oficiales. El país cuenta con el Sistema Estadístico Nacional (SEN), creado por la Ley 1753 del 2015 y reglamentado por el Decreto 1743 del 2016.

Situación problema

La conceptualización es un proceso que lleva a poner en contexto lo aprendido, teniendo en cuenta que no es lo mismo aprender, entender y comprender. En la primera etapa (aprender), el estudiante puede resolver problemas similares a los explicados en clase sin tener que consultar sus notas, libros o compañeros. En otras palabras, es capaz de “repetir” (Soto, 2019). En la segunda etapa (entender), puede argumentar su procedimiento. En la última (comprender), tiene un mayor nivel de abstracción que le permite aplicar lo aprendido en diferentes contextos, aun cuando nadie le haya indicado cómo hacerlo en esos otros escenarios. Por lo tanto, la conceptualización es crucial en la construcción del conocimiento, ya que a través de una interpretación adecuada y una interiorización de lo aprendido, se puede pasar de la teoría a la práctica. Sin embargo, es común en la estadística, al igual que en las matemáticas en general, que los procesos se realicen de manera mecánica sin comprender realmente lo que está sucediendo; es decir, se repite, pero no se comprende, lo que dificulta la aplicación práctica de ese aprendizaje.

La conceptualización estadística, como en otras ciencias y disciplinas que tienen un metalenguaje, suscita que el aprendiz construya interpretaciones a partir de conocimientos previos de su propio contexto; sin embargo, es frecuente que en la actualidad se interpreten conceptos erradamente. Por ejemplo, usar de forma indistinta los vocablos *estadista* y *estadístico*, pues según la RAE *estadista* es la persona con gran saber y experiencia en los asuntos del Estado (Real Academia Española, 2022), mientras que *estadístico* se refiere a una persona que, a partir de los datos y usando teorías y herramientas propias de la estadística, concluye respecto a diferentes fenómenos que pueden ser económicos, sociales, culturales, entre otros.



Nota: para ver mejor el diagrama, escanear el código QR.

Ahora bien, ciencias como la matemática, y en este caso la estadística, han sido históricamente estereotipadas como áreas en las que el aprendizaje se dificulta, y en los estudiantes se evidencia predisposición hacia su estudio, incluso en estadios como el de fundamentación. Al parecer la ejemplificación resulta bastante útil, pero es común que los conceptos base se limiten a definiciones y no vayan acompañados de ejemplos, entonces una carencia en el proceso es la ejemplificación.

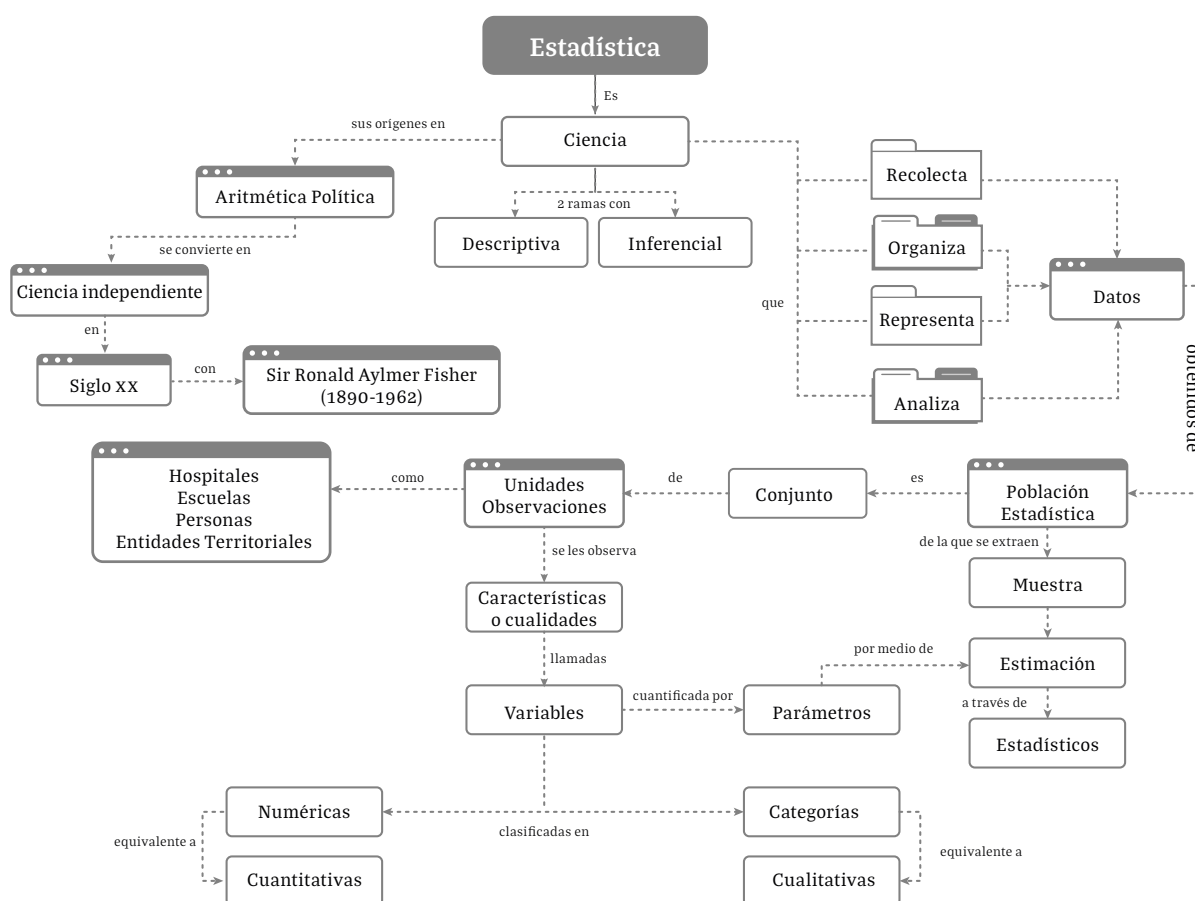


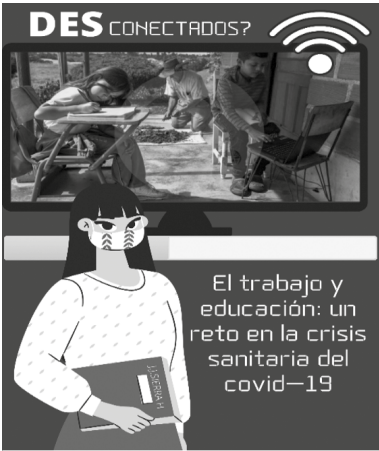
Figura 1. Ejes temáticos de la Unidad 1

Fuente: elaboración propia.

Lectura de motivación

La estadística, una herramienta importante para medir los efectos del SARS-CoV-2

El ámbito educativo fue uno de los tantos lesionados por el sars-cov-2, pues a julio de 2021, un total de 56 319 hogares reportaron que los niños y adolescentes dejaron de asistir a las actividades educativas desde que cerraron las escuelas o colegios por la pandemia, de acuerdo con cifras de la encuesta Pulso Social (Dane – EPS) realizada en 23 ciudades capitales de departamento y áreas metropolitanas¹. Diversas razones se presentan para desconectarse de las actividades educativas; un 13,1 % identificó no contar con dispositivos de computador, tableta o celular, y un 21 % aproximadamente carece de conexión a internet. La pandemia no solo puso en evidencia la desconexión digital de algunos hogares, sino que agudizó la pobreza monetaria, pues con referencia al año 2020, se transitó de 12 333 a 92 974 hogares que tan solo tenían una comida al día; es decir, hubo un aumento del 654 %.



Historia e importancia de la estadística

Etimología

La evolución de la estadística camino a su consolidación como ciencia muestra el desarrollo etimológico histórico de la palabra, como se puede apreciar en la figura 2, sujeta al contexto o cronología, con diversos cambios o modismos. El término *estadística* abordado desde diferentes

1 Para el estudio se tomó una muestra de 3 353 234 hogares de áreas metropolitanas: Barranquilla, Cartagena, Sincelejo, Valledupar, Santa Marta, Riohacha, Bogotá, Tunja, Villavicencio, Neiva, Florencia, Bucaramanga, Cúcuta, Cali, Pasto, Popayán, Manizales, Ibagué, Pereira, Armenia, Medellín, Montería y Quibdó.

	
LATÍN	STATISTICUM COLLEGIUM 'consejo de Estado' STATISTA 'hombre de Estado' o 'político'
ITALIANO	STATISTA, significa 'estadista'
GRIEGO	STATERA, significa 'balanza', debido a la función que cumple la medición de hechos que son objeto de estudio
ALEMÁN	STATISTIK, introducido por Gottfried Achenwall (1749), designaba originalmente el análisis de datos del Estado, es decir, "la ciencia del Estado"

Figura 2. Etimología de la palabra estadística

Fuente: adaptado de Selvamuthu y Das (2018).

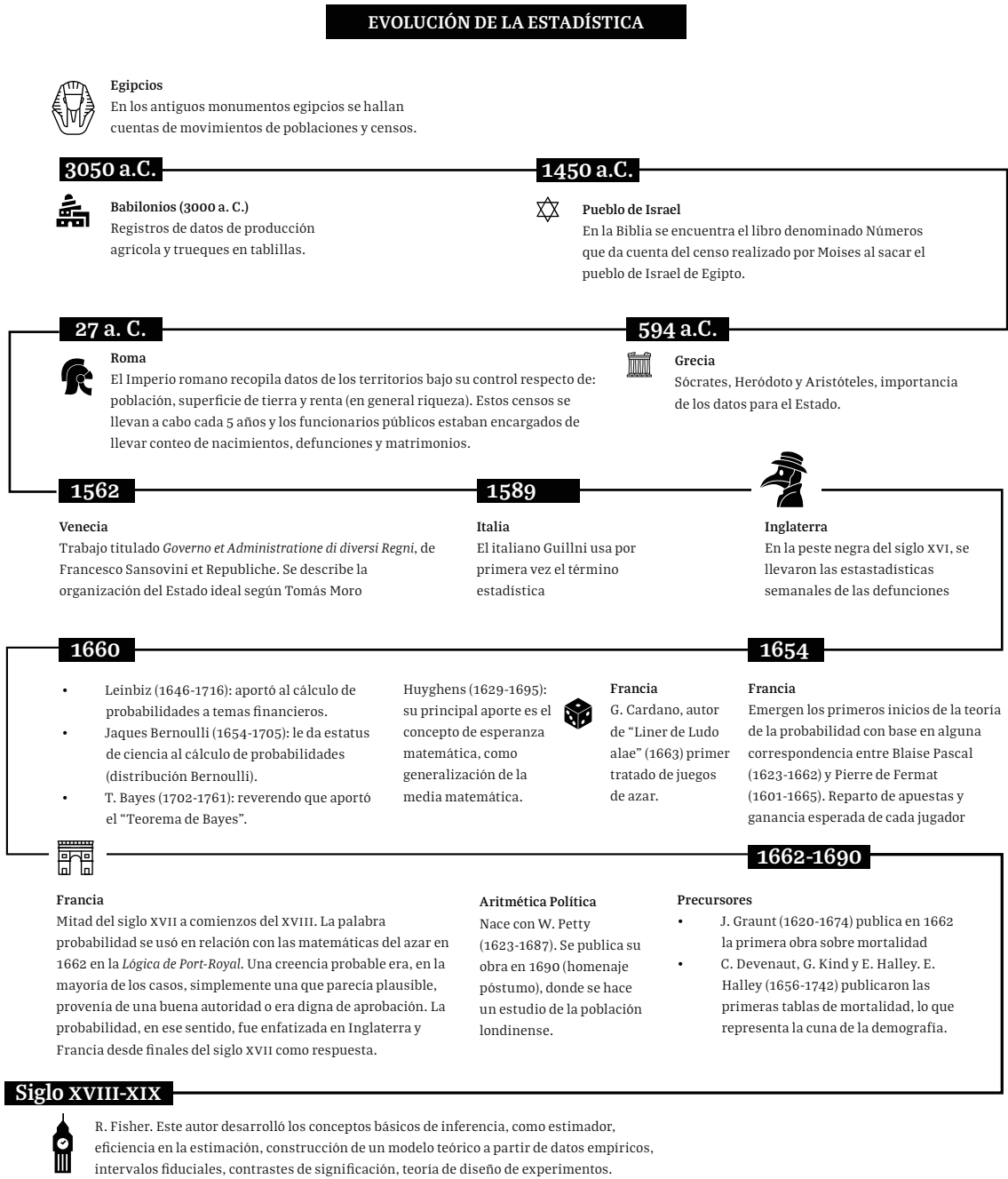
raíces; como el latín *status* que significa 'Estado' o 'Nación', sustantivo al que se le agrega el sufijo *-ist*, *-ista*, *-iste*, creando la palabra *statista* que indica 'hombre de Estado' o 'político', y que cobra el mismo significado en italiano, con el transcurrir del tiempo sufre una metamorfosis, hasta convertirse en el término *estadística*, como actualmente se conoce.

Evolución de la estadística hasta su configuración como ciencia

En el antiguo Egipto, en el año 3050 a. C., se reconocen los primeros vestigios del uso de la estadística, pues se han hallado monumentos que muestran el conteo de personas y sus movimientos, siendo quizás estas manifestaciones el albor de lo que se conoce como censo. Luego, los babilonios mostraron en su arte rupestre una forma de estadística en el ámbito agrícola; a modo de control de producción e ingresos se registraban en tablillas las extensiones de tierra, productos y riqueza lograda en cada caso (Hernández, 2005).

Incluso, en el cuarto libro del Tanaj hebreo, que hace parte de los cinco primeros libros de la Biblia (pentateuco), se presenta en el libro de Números un compendio detallado de registros estadísticos hechos por Moisés en la salida del pueblo de Israel de Egipto. Los 36 capítulos de este libro pormenorizan los clanes, casas paternas, primogénitos, campamentos y ejércitos (capítulos 1 a 3); la cantidad de jefes de las tribus (capítulo 7); número de las poblaciones y libaciones necesarias (capítulo 13); cuenta de la cantidad de hombres sublevados (capítulo 16:2); el tipo y la cantidad de ofrendas (capítulos 28-29); cantidad de botín y su reparto exacto (capítulo 31); agrimensura y dimensiones del territorio (capítulo 35); incluso recuentos minuciosos de las leyes y los relatos contados (Dios habla hoy, La Biblia, 2008, Núm. 1:36).

Más tarde, en Grecia, Sócrates, Heródoto y Aristóteles demuestran la importancia de la recolección y uso de datos para el Estado, con lo cual se convierte en una herramienta fundamental para la toma de decisiones.



Conceptualizando

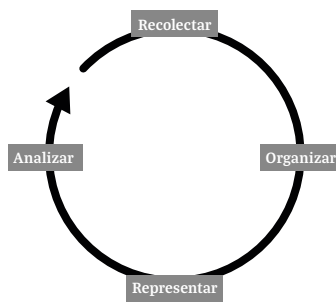


Figura 4. Fases de la estadística

Fuente: elaboración propia.

Definición de estadística

La estadística es el método que permite recolectar, organizar, sintetizar, representar, analizar, cuantificar e interpretar gran cantidad de datos (véase la figura 4), de tal forma que se puedan tomar decisiones, realizar generalizaciones y obtener conclusiones válidas (dar información) acerca de los fenómenos de investigación en estudio.

Algunos de los ejemplos estadísticos más comunes son las ventas periódicas de una empresa, los indicadores económicos de un país, las pruebas médicas de laboratorio, los juegos ganados de un equipo, etc.

La estadística la dividimos en dos tipos: descriptiva e inferencial.

Estadística descriptiva

La estadística descriptiva es un conjunto de procedimientos numéricos y gráficos que tienen por objeto presentar grupos de datos –obtenidos en un estudio o análisis en particular– simplificados en tablas, gráficos o medidas de resumen. La estadística descriptiva, como la palabra misma indica, describe las características de los objetos de análisis a partir de la información obtenida de ellos, la cual no es de fácil comprensión en su presentación original. Así, la estadística descriptiva facilita el manejo de la información y ofrece formas más sencillas de comprender lo que dicen los datos (Ramos *et al.*, 2020).

Estadística inferencial

También llamada estadística inductiva, utiliza técnicas a partir de las cuales se obtienen generalizaciones con base en información de una muestra. Esas generalizaciones tienen un margen de incertidumbre al no trabajar con todos los datos de la población; también es posible que en sí mismo el proceso de recolección de datos contemple errores. En la figura 5 se presenta una comparación entre la estadística descriptiva y la estadística inferencial.

ESTADÍSTICA DESCRIPTIVA	ESTADÍSTICA INFERENCIAL
También conocida como estadística deductiva .	También conocida como estadística inductiva .
Describe los datos de toda una población o muestra.	Permite obtener conclusiones de la población a partir de los hallazgos en una muestra.
Las herramientas y métodos más conocidos son las tablas de frecuencia, gráficos y medidas de resumen.	Aplica diferentes técnicas y métodos como las pruebas de hipótesis, análisis de varianza, regresión y otros.

Figura 5. Cuadro comparativo estadística descriptiva e inferencial
Fuente: adaptación de Mayorga *et al.* (2020).

Población estadística

Un conjunto de elementos o individuos que comparten características en común y que son sujetas a estudio se denomina *población estadística*. Algunos ejemplos se presentan en la figura 6 y la figura 7.

Ejemplo 1.1

Una población estadística puede conformarse por un grupo de instituciones. La figura 6 muestra un conjunto de hospitales.



Figura 6. Población de hospitales
Fuente: elaboración propia.

Ejemplo 1.2

Los grupos de personas pueden ser una población estadística, tal como se ilustra en la figura 7.



Figura 7. Población de graduados de la Escuela Superior de Administración Pública (ESAP)

Fuente: Comunicaciones ESAP, Territorial Tolima.

Otros ejemplos de poblaciones estadísticas son:

- Escuelas del barrio
- Personas a las afueras del centro comercial La Estación
- Peces en un estanque
- Plantas en un cultivo, etc.

Muestra

Un subconjunto extraído de la población a través de algún tipo de muestreo que puede ser o no probabilístico (véase la figura 8).



Figura 8. Muestra de hospitales

Fuente: elaboración propia.

Diseño de muestreo

Se conoce como diseño de muestreo a la metodología que consiste en determinar la cantidad y la forma de selección de los elementos que conformarán la muestra. Para mayor comprensión, se presenta la figura 9.

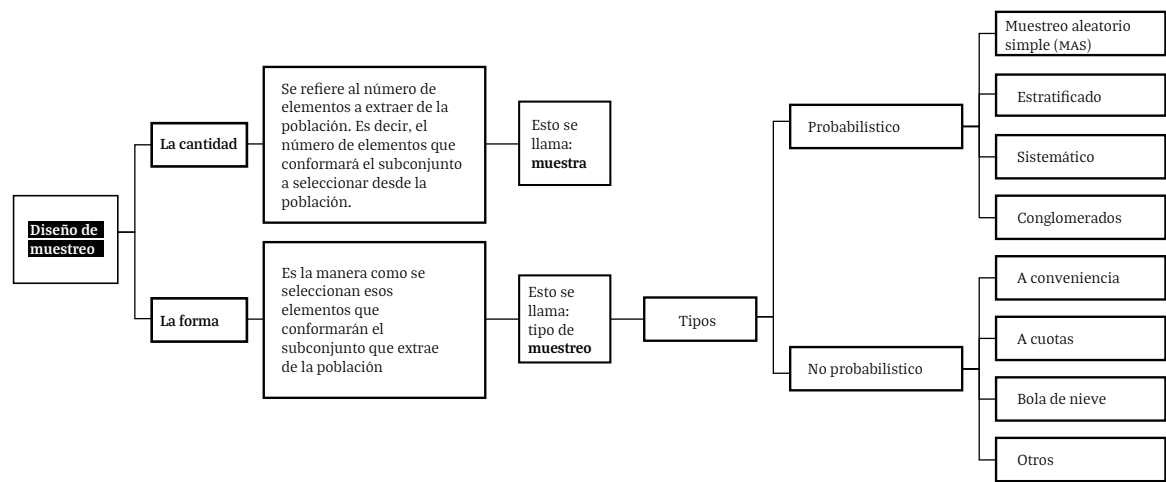


Figura 9. Esquema de diseño de muestreo
Fuente: elaboración propia.

Muestreo probabilístico

Es un método para extraer una muestra de la población, que parte de la premisa de que cada unidad de observación y los elementos que la conforman (personas, entidades territoriales, hospitales, escuelas, etc.) tienen la misma posibilidad de ser seleccionados y ser incluidos en dicha muestra. Son muestreos probabilísticos: muestreo aleatorio simple, muestreo sistemático, muestreo por conglomerados y muestreo estratificado.

Muestreo no probabilístico

Es un método para extraer una muestra de la población, donde cada elemento no tiene la misma posibilidad de ser seleccionado y ser incluido en dicha muestra. Son muestreos no probabilísticos: muestreo a conveniencia, por cuotas, bola de nieve y juicio de expertos.

Censo

Es el conteo de todos y cada uno de los elementos de una población o de un universo, en un lugar determinado y en un tiempo dado. (Departamento Administrativo Nacional de Estadística [Dane], 2021).

Unidad de análisis (UA)

UA es el tipo de objeto delimitado por el investigador para ser investigado. Es una representación abstracta que encarna un conjunto y no a un individuo. Además, solo puede definirse cuando se ha establecido el tema a abordar (pudiendo ser un problema público). La UA permite explicar el problema en más de una forma, por tanto, este puede tener una, dos o más unidades de análisis (Azcona *et al.*, 2013).

Ejemplo 1.2.1

- Desempleo en el municipio de Ibagué
- Unidad de análisis 1: condiciones ocupacionales de hombres y mujeres mayores de edad en el municipio de Ibagué.

Ejemplo 1.2.2

- Residuos sólidos en las calles de Barranquilla dispuestos de forma inadecuada².
- Unidad de análisis 1: cultura en la disposición de basuras en las calles.

Ejemplo 1.2.3

- Calidad de vida de los colombianos
- Unidad de análisis 1: condiciones socioeconómicas de los colombianos.
- Unidad de análisis 2: estado emocional de los colombianos.

Ejemplo 1.2.4

- Movilidad urbana en Bogotá.
 - Unidad de análisis 1: imaginario social de los ciudadanos residentes en la ciudad sobre el transporte de Bogotá.
 - Unidad de análisis 2: percepción de la movilidad en Bogotá de los transportadores públicos.
-

Unidad de observación²

Es el elemento o conjunto de elementos sobre los que se hace la medición de las diferentes variables en una operación estadística (Dane, 2019).

² Según el informe de disposición final de residuos sólidos, el departamento del Atlántico es el cuarto a nivel nacional con mayor número de toneladas diarias en promedio (Superintendencia de Servicios Públicos Domiciliarios, 2019).

Ejemplo 1.2.5 • Desempleo en el municipio de Ibagué • Unidad de observación: personas en edad de trabajar en el municipio de Ibagué.	Ejemplo 1.2.6 • Residuos sólidos en las calles de Barranquilla dispuestos de forma inadecuada. • Unidad de observación: habitantes de barranquilla que transitan por las calles.
Ejemplo 1.2.7 • Calidad de vida de los colombianos • Unidad de observación 1: viviendas • Unidad de observación 2: los hogares • Unidad de observación 3: personas	Ejemplo 1.2.8 • Movilidad urbana en Bogotá. • Unidad de observación 1: transeúntes. • Unidad de observación 2: conductores de servicio público • Unidad de observación 3: automotores.

Variables

Una variable es una característica que se puede observar o medir en cada unidad o sujeto en la población. Las variables, como presenta la tabla 1, se pueden clasificar a partir de diferentes criterios:

Tabla 1. Diferentes formas de clasificación de las variables

Según su papel en la pregunta de investigación	Dependiente (respuesta)
	Independiente (explicativa o exposición)
Según el número de valores posibles	Continuas
	Discretas
Según su escala de medición	Nominal (dicotómica, politómicas)
	Ordinal
	De intervalos (continuas o de escala)
	De razones (continuas o de escala)

Fuente: elaboración propia.

Identificadores

A diferencia de las variables, los identificadores son códigos únicos que permiten diferenciar una unidad de observación de otra. Por ejemplo, un número de cédula, un código del municipio, una ficha del Sisbén o un código estudiantil.

Escalas de medición de variables

Reconocer el tipo de dato y las escalas de medida es importante por la precisión implícita; además, esto permite establecer la herramienta y el método estadístico idóneos para su análisis. En la figura 10, se muestra la clasificación de las variables según su escala de medida.

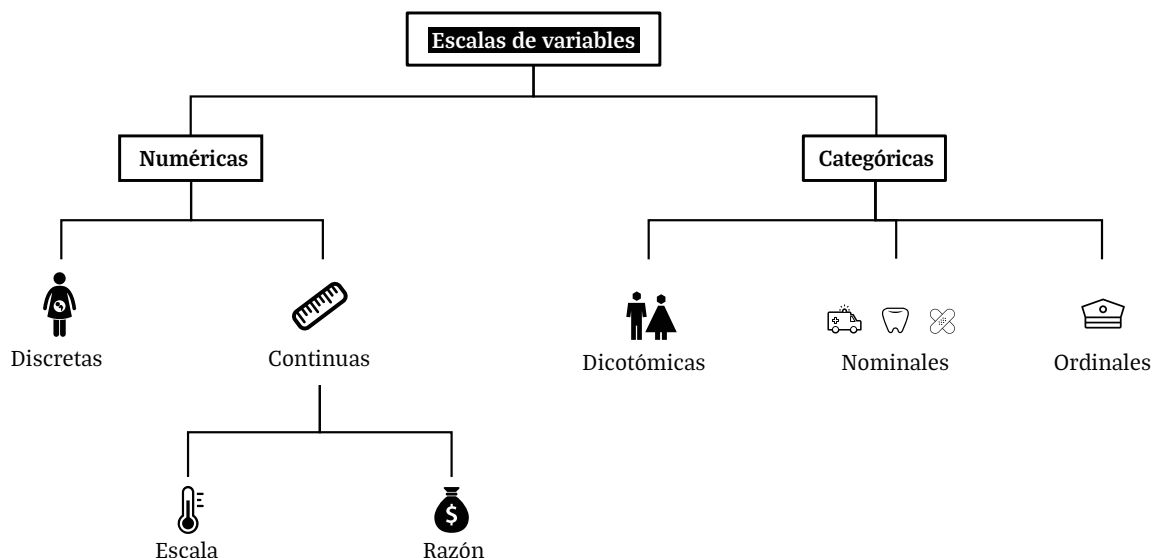


Figura 10. Escalas de medición de las variables

Fuente: elaboración propia.

Numéricas o cuantitativas

Las hay de dos tipos: discretas o continuas (véase la figura 11). El primer tipo solo admite valores enteros ($\mathbb{Z}$) y el segundo toma valores en un conjunto infinito o un intervalo de números reales ($\mathbb{R}$). Un ejemplo de variable en escala discreta es el número de hijos en una familia; algunos valores que puede tomar esta variable son 0, 1, 2 o 3 hijos. ¡No es posible pensar que una familia tenga 1,5 hijos! Entre tanto, una variable en escala continua es la estatura, puesto que admite valores intermedios (entre dos valores es posible un tercero, entre 1 metro y 2 metros es posible tener 1,6 metros).



Figura 11. Variables numéricas: discretas versus continuas
Fuente: elaboración propia.

Tabla 2. Ejemplos de variables numéricas: discretas y continuas

Ejemplo

Variables numéricas discretas	Variables numéricas continuas
1. Número de hijos	1. Edad
2. Número de aparatos electrónicos en el hogar	2. Kilómetros pavimentados en el municipio
3. Cantidad de mujeres en la institución educativa	3. Peso del niño
4. Cantidad de salones	4. Presupuesto del municipio
5. Cantidad de camillas	5. Presupuesto asignado para atender emergencias
6. Cantidad de automotores	6. Tiempo en días transcurrido hasta que recibió una ayuda humanitaria
7. Subsidios entregados de adulto mayor	7. Tiempo que lleva vinculado a su actividad económica
8. Kits de alimentación para niños asignados	8. Distancia que hay de la vivienda a la escuela
9. Profesionales de la salud en la institución prestadora de servicios.	9. Tiempo que se tarda en llegar
10. Número de Instituciones educativas en el municipio	10. Extensión de la institución educativa en metros cuadrados

Fuente: elaboración propia.

De intervalo

Una variable de intervalo carece de cero absoluto, es decir que valores cercanos a cero no indican la ausencia de la variable. Un ejemplo de este tipo de variables es la “temperatura”. Una temperatura de 1°C no indica la inexistencia de temperatura, caso contrario a una variable que no es de intervalo, como la “edad”, en la cual una edad de 1 año o menos sugiere que cuanto más se acerque al cero es porque la persona

acaba de nacer. Además, operaciones como la multiplicación y la división no son realizables con datos en este tipo de escala. Regresando al ejemplo de la temperatura, podemos decir que la distancia entre 10 y 12 grados es la misma que la existente entre 15 y 17 grados, pero lo que no podemos establecer es que una temperatura de 10 grados equivale a la mitad de una temperatura de 20 grados.

De razón

Estas variables, a diferencia de las variables de intervalo, cuentan con cero absoluto; es decir, el valor cero (0) representa la ausencia total de medida, por lo que se puede realizar cualquier operación aritmética (suma, resta, multiplicación y división) y lógica (comparación y ordenamiento). Las variables altura, peso, distancia o el salario son algunos ejemplos de este tipo de escala de medida.

Categóricas

La escala de medida de una variable de respuesta categórica es fundamental para elegir el análisis estadístico apropiado. Las variables de respuesta categórica pueden ser:

- Dicotómicas
- Ordinales
- Nominales
- De recuento

Dicotómicas

Son aquellas que tienen dos posibles resultados. A menudo solo toman un valor de sí o no, por ejemplo, ¿se desarrollaría la enfermedad?, ¿el votante votaría por el candidato A o por el B?, ¿aprobaría el examen? Ahora bien, supongamos que se quiere hacer un estudio sobre el nivel de desempeño municipal, tomando como población de referencia los municipios de Colombia. Una posible variable dicotómica sería: “municipio PDET” (Programas de Desarrollo con Enfoque Territorial), pues acá los valores que puede tomar la variable son: sí/no; es decir, el municipio es o no reconocido como PDET. Otros ejemplos de este tipo de variable están en la figura 12.

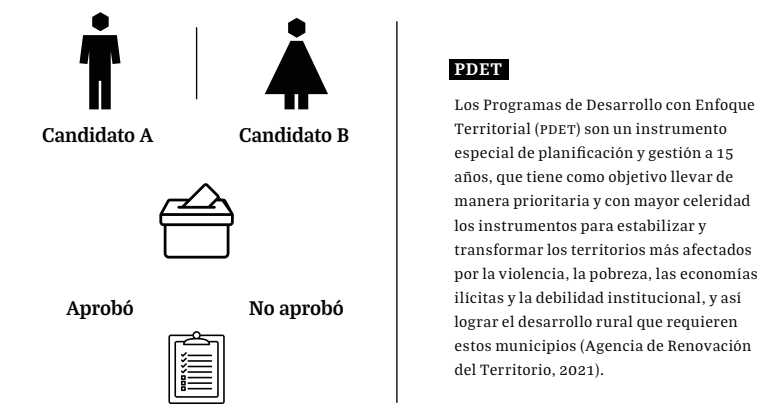


Figura 12. Ejemplos de variables dicotómicas
Fuente: elaboración propia.

Politómicas

Con frecuencia, las respuestas categóricas representan más de dos resultados posibles y por tanto no se pueden clasificar como dicotómicas; en este caso se considera que la variable está en una escala politómica o sencillamente nominal. Un ejemplo de variable politómica es “Grupo étnico”, como se aprecia en la figura 13. En Colombia existen tres grupos étnicos:

- Los negros, palenqueros y raizales
- Los rom o población gitana
- Los indígenas³

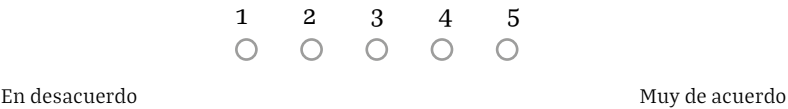


Figura 13. Ejemplo variable politómica (grupo étnico)
Fuente: elaboración propia.

3 A modo de ejemplo, el censo de 2018 dice que la población indígena en Colombia creció un 36,8 %, representando el 4,4 % de la población total del país, siendo 1 905 617 indígenas de todos los pueblos. El crecimiento poblacional de los pueblos indígenas fue seis veces más que el total nacional, aumento explicado en buena parte por una tasa de fecundidad que duplica la media nacional (Departamento Administrativo Nacional de Estadística [Dane], 2019).

Ordinales

Frecuentemente, en estos resultados de características categóricas es posible considerar algún orden inherente. Estas variables tienen una escala de respuesta ordinal de medición.



Veamos los siguientes ejemplos. ¿Qué tan favorable es su percepción frente a la política de educación actual? (Nada de acuerdo, algo de acuerdo, muy de acuerdo) o ¿cuál es el nivel de educación más alto alcanzado por usted? En esos resultados hay una escala subyacente y una forma aparente de ordenarlos. Pues bien, se puede decir, según la respuesta obtenida, que existe una jerarquía; es posible comparar dos datos y concluir cuál de ellos es superior al otro –lo que no se puede hacer en variables como el sexo– ¡No se puede decir si mujer está por encima de hombre! Otro ejemplo: rango de los oficiales y suboficiales navales de la Armada Nacional como se muestra en la figura 14.

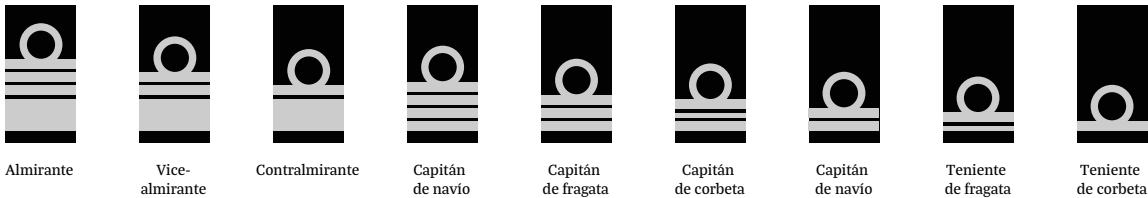


Figura 14. Ejemplo variable ordinal de rango de oficiales de Armada Nacional de Colombia
Fuente: tomada de Grados Armada (s. f.).

Las variables categóricas a veces contienen recuentos. En lugar de considerar las categorías que presenta cada observación, (sí, no) (bajo, medio, alto), los resultados que se estudian son los números mismos. El tamaño de la camada, ¿fue de 1, 2, 3, 4 o 5 animales? La casa tiene ¿1, 2, 3 o 4 equipos de aire acondicionado?

Las variables de recuento se definen como: “número de sucesos o eventos que ocurren en una misma unidad de observación en un intervalo temporal o espacial definido” (Lindsey, 1995). De esta definición

se desprende, tal como se ha indicado anteriormente, que los sucesos a recontar corresponden a fenómenos recurrentes. Ejemplos:

- Número de errores en una tarea de reconocimiento de palabras
- Número de asignaturas suspendidas durante una licenciatura
- Número de veces que un fumador ha intentado dejar de fumar
- Número de accidentes laborales durante un periodo de tiempo

Tabla 3. Ejemplo de variables categóricas

Dicotómicas	Politómicas (nominales)	Ordinales
• Sexo⁴ • Carácter de la institución de educación (privada o pública) • Ubicación (urbano o rural) • Municipio categorizado como priorizado para el posconflicto: sí o no • Registro en cámara de comercio (sí o no)	• Estado civil • Vocación del municipio (minera, turística) • Servicios prestados por los hospitales: urgencias, quirófano, odontología, laboratorio, hospitalización • Tipo de electrodomésticos en el hogar: TV, PC, nevera, etc. • Zona de la vivienda en el municipio: zona 1, zona 2, etc.	• Estrato socioeconómico (bajo, medio o alto) • Nivel de escolaridad • Categoría del municipio: especial, I, II, III, y IV. Ver Ley 617 del 2000. • Desempeño municipal: crítico, bajo, medio, satisfactorio y sobresaliente⁵ • Nivel del hospital I, II, III

Fuente: elaboración propia.

4 Acá se habla desde el concepto biológico, pero si quiere hacer una distinción entre los roles socialmente contruidos, use el término *género*; sin embargo, ya no sería una variable dicotómica. Para más información, véanse los tipos de género en BBC (2018).

5 Véase el MDM del DNP (2019); en especial, vale la pena revisar el informe de eficacia de cada departamento.

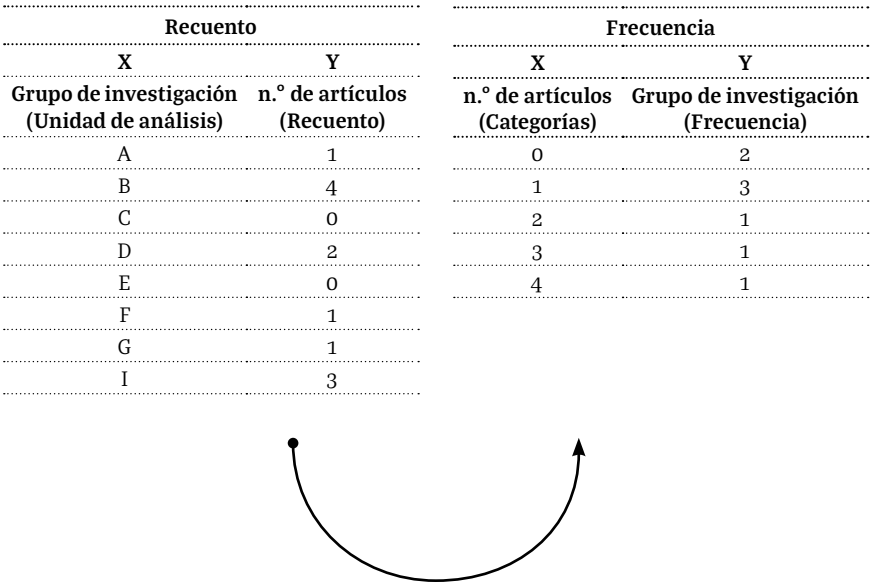


Figura 15. Variables de recuento
Fuente: tomado de Vives Brosa (2002).

Dato

Es el valor o la categoría que puede tomar una variable analizada en una unidad de observación. Identifica y define la unidad básica de información a partir de la cual se realiza el intercambio de información, de acuerdo con los requerimientos funcionales definidos dentro del proceso o servicio de intercambio de información (Ministerio de Tecnologías de la Información y las Comunicaciones, 2019).

Ejemplo 1.3

En Bogotá, la capital de Colombia, una de las unidades de planeamiento zonal (UPZ) que hace parte de la localidad de Teusaquillo, como es Galerías, tiene aproximadamente un área de espacio verde⁶ de 2,09 m²/hab.

6 La Organización Mundial de la Salud (OMS) recomienda un umbral mínimo de 9 m² por habitante (ONU-Hábitat, 2015). Puede consultar las estadísticas de espacio verde por ciudadano en los diferentes territorios en Colombia en Datos Abiertos Colombia (s. f.).

- Población: Bogotá
- Unidad de observación: Galerías
- Variable: área de espacio verde por habitante
- Dato: 2,09 m²/hab.

Tipos de datos

Los datos tienen características con respecto a la manera en que se presentan, esto en relación con datos estructurados, no estructurados y semiestructurados.

Tabla 4. Tipo de datos

Tipo de dato	Descripción	Ejemplos
Estructurados	Son datos que tienen una estructura interna identificable. Tienen definida la longitud, el formato y su tamaño	Hojas de cálculo, bases de datos relacionales y texto delimitado por comas, cuya extensión es .csv. Ejemplo: fichas del Sisbén IV
		FichaNombreApellido
		Tipo de documentoMunicipio
Semiestructurados	No siguen un sistema de bases de datos convencional. Aunque suelen tener la forma de datos estructurados, no están organizados en modelos de bases de datos relacionales	Páginas web, señales de tráfico, formatos XML y JSON. Ejemplo: correos electrónicos Parte estructurada: destinatario, receptor, asunto. Parte no estructurada: cuerpo del mensaje
No estructurados	No tienen una estructura interna identificable	Documentos de PDF o Word, mensajes de texto, grabaciones, videos con audio

Fuente: adaptado de DNP (2020), Gómez (2021) y Calvo (2017).

Importancia de los datos en el sector público

El mundo actual produce cada segundo una cantidad exponencial de datos y, según Mohamed y Weber (2020), un 90 % de los datos disponibles en la actualidad se ha generado en los últimos dos años.

Tales datos se asocian a las vivencias y comportamientos de las comunidades, lo cual permite identificar las necesidades de la ciudadanía y a la vez pensar en diversas soluciones para suplirlas. Por tanto, “los datos se han convertido en activos estratégicos para la generación de valor social y económico tanto en el sector público como en el privado” (Departamento Nacional de Planeación [DNP], 2020).

La obtención de datos y su uso adecuado permiten una gestión pública más eficiente y transparente. Estos son la base para la toma de decisiones de la administración pública, ya que permiten generar proyectos con intenciones específicas, de acuerdo con la realidad presentada. Hay datos por doquier. Los distintos ámbitos de interés de la administración pública producen datos; cada minuto se producen en los sectores: agropecuario, salud, educativo, transporte y muchos más, de ahí que una eficiente captura de información y análisis contribuye a la toma de decisiones en las diferentes áreas.

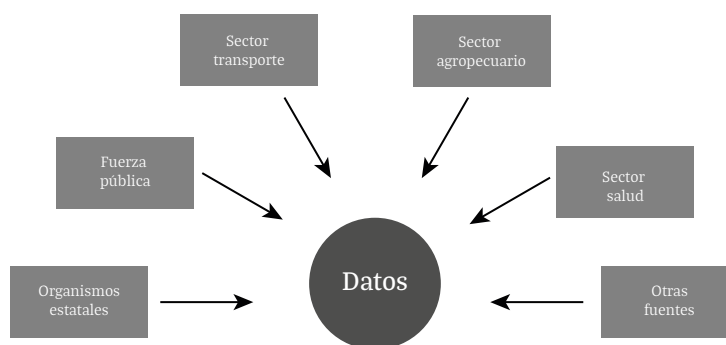


Figura 16. Fuentes de datos

Fuente: elaboración propia.

Datos delictivos. El comportamiento delictivo en las diferentes zonas de una ciudad está disponible en las bases de datos que reposan en la Policía Nacional, como el Sistema de Información Estadístico Delincuencial Contravencional y Operativo de la Policía Nacional⁷ (Siedco).

7 Puede consultar información estadística de delitos como: delitos sexuales, terrorismo, secuestro, extorsión, homicidios, violencia intrafamiliar, entre otros, en la base de datos de la Policía Nacional de Colombia (s. f.).

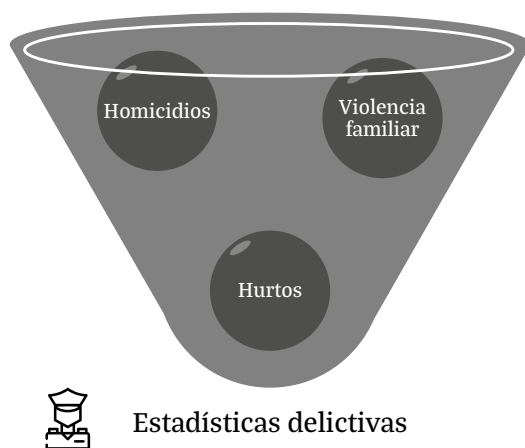


Figura 17. Estadísticas delictivas

Fuente: elaboración propia a partir de Sierra (2022) y la página web del Grupo Información de Criminalidad (Gicri) de la Policía Nacional de Colombia (s. f. b).

La estadística en la planeación del desarrollo territorial

Recuerde que la palabra *estadística* procede del vocablo *Estado*, pues era función principal de los gobiernos de los Estados establecer registros de población, nacimientos, defunciones, impuestos, cosechas, etc. Desde que se establecieron sociedades humanas organizadas se ha tenido la necesidad de poseer datos cifrados sobre la población. Así, para el gobierno nacional y, en particular, para los gobiernos subnacionales, reconocer el territorio es fundamental para formular los ejes de acción para el desarrollo en las diferentes dimensiones del municipio, departamento o distrito; de ahí que en el Plan de Desarrollo Territorial (PDT) sea tan importante el diagnóstico del territorio, y la estadística es una herramienta fundamental para este propósito.

El diagnóstico del territorio se puede equiparar al que hace el médico: es la descripción del estado actual del paciente, que para el caso, es el ente territorial. En este sentido, el territorio se debe analizar desde diferentes ámbitos, como sugiere la metodología propuesta por el Departamento Nacional de Planeación (DNP). Una lectura sectorial y transversal del territorio es necesaria, o un análisis desde diferentes dimensiones (poblacional, ambiental, sociocultural y económica).

Un PDT debe fundamentarse en hechos no en supuestos; así, los datos, cifras e indicadores revisten gran valor. Por ejemplo, para señalar que un municipio es de vocación turística, antecede un detallado

análisis del número de restaurantes, hoteles, cafeterías, balnearios y demás establecimientos que permitan ubicarlo en tal categoría. En el caso de hoteles, se puede determinar la capacidad de habitaciones por hotel en el municipio, y con ellos se estima el porcentaje de ocupación hotelera o el número de turistas que llegan por semana, mes o periodo del año. Además, se puede definir el tipo de turismo como: turismo de sol y playa, arqueológico, de aventura, ecológico, cultural, de negocios, deportivo, rural y místico, reconociendo así la oferta y potencial de la región (Velazco, 2011).

Las estadísticas oficiales

Las fuentes de información deben ser confiables. En Colombia, el Dane es la entidad responsable de la planeación, levantamiento, procesamiento, análisis y difusión de las estadísticas oficiales. El país cuenta con el Sistema Estadístico Nacional (SEN)⁸, el cual fue creado por la Ley 1753 del 2015 y reglamentado en el Decreto 1743 del 2016. Este último define el SEN como “el conjunto articulado de componentes que, de manera organizada y sistemática, garantiza la producción y difusión de las estadísticas oficiales que requiere el país”. Por otra parte, la Ley 1753 dispone que el SEN sea coordinado y regulado por el Dane.

8 El Decreto 0574 del 21 de marzo de 2012 establece que, para ser parte de la Ode, Colombia requiere “una revisión de todo el sistema de producción de estadísticas” y el Dane debe ser la entidad coordinadora del directorio estadístico del sector público.

El directorio estadístico del sector público:

- Detalla la información jurídica y económica de las entidades del sector público a nivel nacional y territorial.
- Caracteriza las entidades del nivel central y descentralizado que producen bienes y prestan servicios esenciales para la sociedad.
- Contiene información económica importante para generar estadísticas de finanzas públicas del país.
- Garantiza que el país cuente con un sistema de información único y de calidad para la producción de estadísticas fiscales.

El Dane presenta el Plan Estadístico Nacional (PEN), que es el “documento que establece los lineamientos estratégicos y las acciones para el desarrollo de la producción y difusión estadística teniendo en cuenta las necesidades de información del país, incluyendo enfoques diferenciales. El Plan incluye la oferta de operaciones estadísticas, la demanda de información, los requerimientos de intercambio de información entre los integrantes del SEN y las estrategias sobre la disposición final de los resultados de las operaciones estadísticas” (Decreto 1743 del 2016, Artículo 1). El PEN 2017-2022 es la hoja de ruta del país para la producción estadística durante el referido quinquenio.

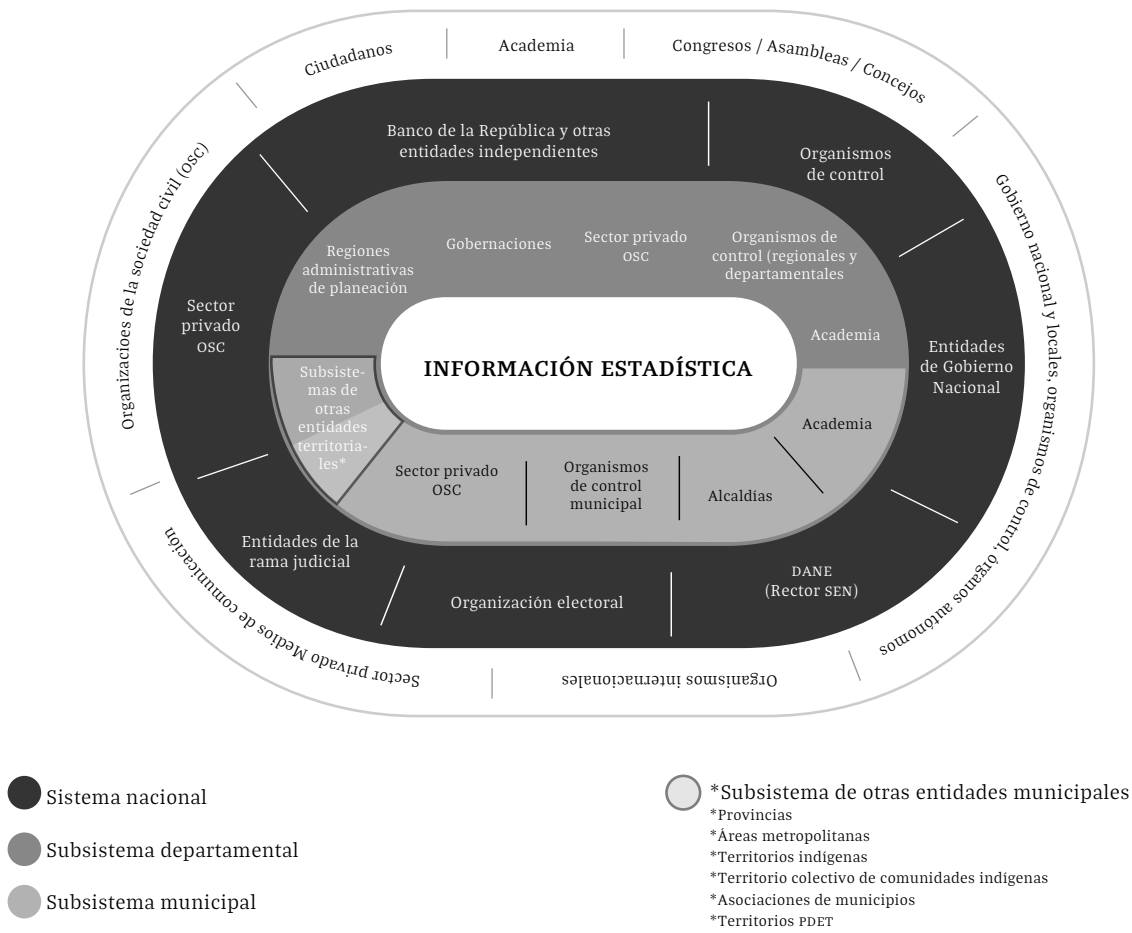


Figura 18. Estructura del Sistema Estadístico Nacional de Colombia (SEN)

Fuente: Dane y Dirpen (2020).

El censo nacional agropecuario

Según el Departamento Administrativo Nacional de Estadísticas (DANE) en 2017, la tercera edición del Censo Nacional Agropecuario (CNA) de 2014 tuvo una cobertura integral de todas las actividades productivas, tanto agropecuarias como no agropecuarias, llevadas a cabo en las áreas rurales dispersas de Colombia. Se abordaron todas las unidades de análisis, sin importar su actividad productiva, tamaño, tenencia de tierras o ubicación. Adicionalmente, se estableció la Unidad Productiva no Agropecuaria (UPNA) con el fin de recopilar datos específicamente relacionados con actividades exclusivamente no agropecuarias.

Como no se había realizado una investigación estadística similar desde el último censo, el país carecía de información actualizada y completa sobre las unidades agropecuarias y no agropecuarias a nivel nacional. El módulo principal del censo agropecuario se diseñó para abarcar el conjunto mínimo de datos recomendado por la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO), los cuales incluyen:

- Identificación y ubicación de la explotación agrícola
- Condición jurídica del productor agrícola
- Sexo del productor agrícola
- Edad del productor agrícola
- Sexo de los miembros del hogar
- Finalidad principal de la producción de la explotación
- Área de la explotación según los tipos de utilización del suelo
- Área total de la explotación
- Regímenes de tenencia del suelo en la explotación
- Presencia de riego en la explotación
- Tipos de cultivos transitorios en la explotación
- Tipos de cultivos permanentes en la explotación y posible ubicación en plantaciones compactas
- Cantidad de animales en la explotación por cada tipo
- Presencia de acuicultura en la explotación
- Presencia de bosques y otras tierras boscosas en la explotación
- Otras actividades económicas de producción en la explotación

La FAO proporciona directrices para la recolección de datos a nivel nacional, pero cada país adapta sus cuestionarios y métodos según sus particularidades. Esta flexibilidad permite ajustar las investigaciones a las necesidades y condiciones específicas de cada nación.

Las recomendaciones y los conceptos de la FAO son un punto de referencia, y cada país desarrolla sus cuestionarios y los procedimientos para la recolección de datos de acuerdo con sus condiciones particulares. La FAO insiste en que debe haber integración entre los censos agropecuarios y los censos de población, mediante la vinculación de los datos de las dos investigaciones; la coordinación en la recolección de campo; el intercambio de preguntas; el uso de conceptos estandarizados; el uso de la misma unidad de observación y el canje de material de campo. Se plantea hacer un seguimiento anual sobre una batería de indicadores (FAO, 2009, p. 9). Sin embargo, esta última condición lleva a que las variables que se miden en su mayoría sean de carácter coyuntural, como el área cosechada; la producción por cultivo; la productividad por cultivo; el número de nacimientos de animales; el número de animales sacrificados; la producción de leche y carne; el precio promedio anual a nivel nacional para productos de origen agropecuario; y el costo de los insumos para la producción. Además, el nivel de complejidad y el grado de profundidad que requiere la medición objetiva de algunas de esas variables conduce a que sean objeto de operaciones estadísticas diferentes a un censo.

Autoevaluación: ¡Practica tus conocimientos!

Ubica en cada uno de los círculos los números del 1 al 5 que están en la lista de la izquierda, asociando la palabra subrayada a uno de los conceptos básicos de la estadística vistos; sigue el ejemplo:

① Población

② Variable

③ Muestra

④ Unidad de observación

⑤ Dato

En una muestra tomada de 800 ciudadanos residentes

en las 13 ciudades principales en Colombia de

un estudio respecto al uso de mecanismos de

participación ciudadana que analizó entre otros el

tipo de mecanismo usado en los últimos 5 años, las

siguientes son 10 de esas 800 respuestas obtenidas:

Práctica 2. Clasificación de variables

Clasifica las variables de la lista que se construyó en la práctica 1 para el diagnóstico de tu municipio:

Unidad de observación	Variable	Tipo de variable

Recursos de aprendizaje

- Anderson, D., Sweeney, D. y Williams, T. (2012). *Estadística para administración y economía* (10.ª ed.). Cengage Learning.
- Anderson, D., Sweeney, D. y Williams, T. (2012). *Estadística para los negocios y la economía* (11.ª ed.). Cengage Learning.
- Martínez, C. (2012). *Estadística y muestreo* (13.ª ed.). Ecoe Ediciones.
- Lind, D., Marchal, W. y Wathen, S. (2012). *Estadística aplicada a los negocios y a la economía* (15.ª ed.). McGraw Hill.
- Departamento Administrativo Nacional de Estadística [Dane]. (2017, junio 21). Tercer censo Nacional Agropecuario. *Anda-Microdatos del Dane*. http://microdatos.dane.gov.co/index.php/catalog/513/get_microdata⁹

9 Microdatos de la página del Dane del Tercer Censo Nacional agropecuario realizado en el 2014 en Colombia.
Los datos se encuentran divididos en archivos por cada departamento en una carpeta comprimida, dentro de la cual se encuentra el archivo tipo comas extensión CSV. De la misma manera, se encuentran archivos extensión SPSS. Igualmente hay un archivo consolidado de todos los departamentos.
Para usar estos datos, se puede descargar la información de un departamento

Referencias

- Agencia de Renovación del Territorio. (10 de agosto del 2021). PDET. https://www.renovacionterritorio.gov.co/especiales/especial_PDET/#descargables
- Almanzor, J. R., Murias, P. G. y Rodríguez, M. T. (1997, octubre). Resumen Histórico de la Evolución de la Estadística. *Estudios de Economía Aplicada*, 8(3), 141. <https://d1wq-txts1xzle7.cloudfront.net/50260463/Dialnet-ResumenHistoricoDeLaEvolucionDeLaEstadistica-176008-with-cover-page-v2.pdf?Expires=1631646423&Signature=CT9BOqU-4DocuwVhy166OeTiK9zhUQlMLoyxK8wTI1TUIBHVcEe-pUz4dj4MNZo4PnKypBaAtpLQ5oGj8I4z5ekmzeed2DM~5a>
- American Statistical Association. (2021, julio 4). ASA. History of Statistics. <https://www.amstat.org/ASA/History-of-Statistics.aspx>
- Azcona, M., Manzini, F. y Dorati, J. (2013). Precisiones metodológicas sobre la unidad de análisis y la unidad de observación. IV Congreso Internacional de Investigación de la Facultad de Psicología. [Memorias]. Universidad Nacional de La Plata. <http://sedici.unlp.edu.ar/handle/10915/45291>
- British Broadcasting Corporation [BBC]. (2018). Pansexual, intersexual, 'queer', homosexual, transexual... ¿cuántas formas hay de definir la identidad sexual? <https://www.bbc.com/mundo/noticias-38995644>

.....
de Colombia sobre el cual trabajará y presentará un informe estadístico del diagnóstico de tal departamento.

- *Proyecto*: “Diagnóstico” en el Plan de Desarrollo Territorial.
- *Dimensión*: económica (ver dimensiones Ley 152 y Ley 388)
- *Línea estratégica*: por definir
- *Sector*: agropecuario

Se deja para consulta el objeto virtual de aprendizaje “Kit de Planeación Territorial”, que permite comprender el proceso de Diagnóstico en la elaboración del Plan de Desarrollo Territorial con todas sus etapas, ubicado en el portal territorial del Departamento Nacional de Planeación (DNP) (2019).

- Calvo, D. (2017, 21 de noviembre). Tipos de datos: estructurados, semiestructurados y no estructurados. [Mensaje en un blog]. <https://www.diegocalvo.es/tipos-de-datos-estructurados-semiestructurados-y-no-estructurados/>
- Datos Abiertos Colombia (s. f.). Espacio público verde por habitante por UPZ. <https://www.datos.gov.co/widgets/276j-m5rd>
- Departamento Administrativo Nacional de Estadística [Dane]. (2019, 16 de septiembre). Enfoque diferencial e intersectorial. <https://www.dane.gov.co/index.php/estadisticas-por-tema/enfoque-diferencial-e-interseccional/autorreconocimiento-etnico>
- Departamento Administrativo Nacional de Estadística [Dane]. (2019, julio). Manual de conceptos. *Nacional de Población y Vivienda (CNPV) 2018*. <https://www.dane.gov.co/files/censo2018/informacion-tecnica/CNPV-2018-manual-conceptos.pdf>
- Departamento Administrativo Nacional de Estadística [Dane]. (2021, julio). Glosario Estadístico. *Dane*: <https://www.dane.gov.co/files/Glosario.pdf>
- Departamento Nacional de Planeación [DNP] (2019). Medición del desempeño municipal. <https://www.dnp.gov.co/programas/desarrollo-territorial/Estudios-Territoriales/Indicadores-y-Mediciones/Paginas/desempeno-integral.aspx>
- Departamento Nacional de Planeación [DNP] (2020, septiembre). Aprovechamiento de datos para la toma de decisiones en el sector público. *Dirección de Desarrollo Digital*: <https://www.dnp.gov.co/DNPN/Documents/Borrador-Aprovechamiento-de-datos.pdf>
- Dirección de Regulación, Planeación, Estandarización y Normalización [Dirpen]. (2020). *Plan Nacional Estadístico. Actualización 2020-2022*. Departamento Administrativo Nacional de Estadística. <https://www.sen.gov.co/files/PEN/PEN%202020.pdf>
- Gómez, A. (2021). Big data, un sistema de gestión de datos. Manuscrito inédito, Tecana American University, USA. https://tauniversity.org/sites/default/files/articulo_big_data_de_angel_gomez_degraves.pdf
- Grados Armada (s. f.). <http://1.bp.blogspot.com/-CwH1RSsP9Nc/UC4sKTSx1pI/AAAAAAAAADk/FZRMmglQ3iQ/s1600/GRADOS+ARMADA.bmp>

- Hernández, S. (mayo del 2005). Historia de la estadística. *Revista de Divulgación Científica y Tecnológica de la Universidad Veracruzana*, 18(2). <https://www.uv.mx/cienciahombre/revistae/vol18num2/articulos/historia/>
- Lindsey, J.K. (1995). *Introductory Statistics: A Modelling Approach*. Oxford University Press.
- López Dimas, B. O., Santillán Cuéllar, Y. A. y Trujillo Rivera, J. A. (s. f.). Línea de tiempo “Historia de la estadística y probabilidad”. <https://prezi.com/cljkusgdotjn/linea-del-tiempo-historia-de-la-estadistica-y-la-probabilidad/?frame=abf639408634cc5353a9165b0b50ff1afd8680b4>
- Mayorga, R. B., Sillis, K., Martínez, A., Salazar, D. y Mota, U. I. (2020). Cuadro comparativo “Estadística inferencial y descriptiva”. *Educación y Salud Boletín Científico Instituto de Ciencias de la Salud Universidad Autónoma del Estado de Hidalgo*, 8(16), 93-95. <https://doi.org/10.29057/icsa.v8i16.5806>
- Ministerio de Tecnologías de la Información y las Comunicaciones. (agosto del 2019). Marco de interoperabilidad. *MinTic*. http://lenguaje.mintic.gov.co/sites/default/files/archivos/marco_de_interoperabilidad_para_gobierno_digital.pdf
- Mode, E. (2021). Introducción. En E. Mode, *Elementos de probabilidad y estadística* (p. 2). Reverté, S. A.
- Mohamed, M. y Weber, P. (2020). Trends of digitalization and adoption of big data & analytics among UK SMEs: Analysis and lessons drawn from a case study of 53 SMEs. *2020 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, (pp. 1-6). <https://ieeexplore.ieee.org/document/9198545>
- O’Neill, A. (20 de junio del 2022). Number of countries where the highest position of executive power was held by a woman, in each year from 1960 to 2022. *Statista*. <https://www.statista.com/statistics/1058345/countries-with-women-highest-position-executive-power-since-1960/>
- Policía Nacional de Colombia (s. f. a). Estadística delictiva. <https://www.policia.gov.co/grupo-informacion-criminalidad/estadistica-delictiva>
- Policía Nacional de Colombia (s. f. b). Grupo Información de Criminalidad (Gicri). <https://www.policia.gov.co/grupo-informacion-criminalidad>

- Real Academia Española. (2022). Estadista. En *Diccionario de la Lengua Española* (edición tricentenario). <https://dle.rae.es/estadista>
- Ramos, J., del Águila, V. y Bazalar, A. (2020). *Estadística básica para los negocios*. Fondo Editorial de la Universidad de Lima.
- Selvamuthu, D. y Das, D. (2018). *Introduction to Statistical Methods, Design of Experiments and Statistical Quality Control*. Springer Singapore.
- Sierra, J. J. (2021, agosto 03). *Historia de la estadística*. <https://my.vismeco.com/projects/8re1948e-historia-de-la-estadistica>
- Sierra, J. J. (2022). Datos abiertos y fuentes de información en el sector público [Archivo de video no publicado].
- Soto, E. (2019). Aprender, entender y comprender matemáticas. <https://www.aprendematematicas.org.mx/aprender-entender-y-comprender-matematicas/>
- Superintendencia de Servicios Públicos Domiciliarios. (2019, diciembre). Disposición Final de Residuos Sólidos Informe Nacional- 2018. *Superintendencia de Servicios Públicos Domiciliarios*. https://www.superservicios.gov.co/sites/default/archivos/Publicaciones/Publicaciones/2020/Ene/informe_nacional_disposicion_final_2019_1.pdf
- Velazco, E. (2011). Metodología para determinar el índice de vocación turística de los municipios de Veracruz. https://www.academia.edu/1299495/Metodolog%C3%ADa_para_determinar_el_%C3%ADndice_de_vocaci%C3%B3n_tur%C3%ADstica_de_los_municipios_de_Veracruz
- Vives Brosa, J. (2002). *El diagnóstico de la sobredispersión en modelos de análisis de datos de recuento* [tesis doctoral, Universitat Autònoma de Barcelona]. <http://hdl.handle.net/10803/5422>

Bibliografía comentada

- Ramos, J., Águila, V. y Bazalar, A. (2020). *Estadística básica para negocios*. Fondo Editorial de la Universidad de Lima.

Este libro resulta ser un apoyo para el estudiante que incursiona en el aprendizaje de la estadística; su estructura se fundamenta en la

ejemplificación. A lo largo del texto, el lector podrá encontrar un buen compendio de ejemplos y de situaciones en el contexto de los negocios que le permitirán articular lo teórico con lo práctico. En particular, la sección 4 presenta una miscelánea de problemas donde se ponen a prueba los conocimientos básicos de clasificación de datos, lo que permite que el estudiante utilice las herramientas estadísticas adecuadas según el tipo de dato a analizar.

Dirección de Regulación, Planeación, Estandarización y

Normalización [Dirpen]. (2020). *Plan Nacional Estadístico. Actualización 2020-2022*. Departamento Administrativo Nacional de Estadística. <https://www.sen.gov.co/files/PEN/PEN%202020.pdf>

En Colombia, por medio del Artículo 160 de la Ley 1753 del 2016, se crea el Sistema Estadístico Nacional (SEN), con el objetivo de suministrar a la sociedad y al Estado territoriales de calidad, con lenguajes y procedimientos comunes respetuosos de los estándares estadísticos internacionales, que contribuyan a la transparencia, pertinencia, interoperabilidad, acceso, oportunidad y coherencia de las estadísticas producidas en el país. A la par, como principal instrumento del SEN, se gesta el Plan Nacional Estadístico-PNE, principal instrumento de planeación estadística del país; contiene los lineamientos estratégicos y acciones para lograr el propósito del SEN, y su objetivo es ser la hoja de ruta de la producción y difusión de información estadística en los próximos años.

Comisión Económica para América Latina [Cepal]. (2020, 30 de diciembre). *Estadísticas*. <https://www.cepal.org/es/fuentes-de-informacion-estadistica>

Es un recurso efectivo para reconocer el desarrollo de la estadística en otros países de la región; el sitio enlaza a las principales Oficinas Nacionales de Estadística (ONE), con lo cual se tiene un compendio amplio de fuentes de información oficiales. Así mismo, se destaca una lista y acceso abierto a diversos programas y agencias de Naciones Unidas de los países miembros, entre las que se encuentra el World Bank, donde podrá obtener incluso microdatos en diferentes campos de interés de la sociedad; por ejemplo, en temas de actualidad como la inclusión

financiera, datos que bien vale la pena tener en cuenta en el análisis de iniciativas legislativas en el contexto colombiano, como las que han causado polémica respecto al uso de saldos de cuentas bancarias inactivas. Estos y otros tantos conjuntos de datos le permiten al lector tener una perspectiva amplia de la información, no solo de la que se produce a nivel nacional sino más allá de sus fronteras.

Instituto Nacional de Salud, Observatorio Nacional de Salud, covid-19 en Colombia. (2020). *Covid-19 en Colombia, consecuencias de una pandemia en desarrollo*. <https://www.ins.gov.co/Direcciones/ONS/Informes/12COVID-19%20en%20Colombia,%20pandemia%20en%20desarrollo.pdf>

El Instituto Nacional de Salud presenta el decimosegundo informe técnico denominado “Covid-19 en Colombia: consecuencias de una pandemia en desarrollo”. En el capítulo 1 se expone la naturaleza del sars-cov-2 y las medidas sanitarias implementadas en el país para evitar el contagio; en el capítulo 2 se explica de manera detallada todo lo que en materia de investigación epidemiológica se había logrado a la fecha para contrarrestar el virus. Los capítulos 3 a 5 dan cuenta de las afectaciones directas, indirectas y sobre poblaciones especiales que trajo el covid-19; de este intervalo documental se destaca la elaboración de un índice denominado *índice de afectación económica*, a partir del cual se presentan estadísticas bastante importantes en la dinámica territorial.

Los capítulos 6 a 8 relacionan los impactos en poblaciones especiales como: grupos étnicos, personas privadas de la libertad (PPL), inmigrantes y comunidades inmersas en territorios de conflicto armado, entre otros. Las estadísticas presentadas en el informe resultan ser un insumo valioso para el análisis de problemáticas a partir de evidencias; sin embargo, vale la pena confrontar, a través de la sociedad civil organizada de tales grupos de interés, los datos mostrados, pues es muy frecuente subvalorar cifras, especialmente en contextos en los que las unidades de observación tienen previsiones, como es el caso de los inmigrantes venezolanos.

Distribución de frecuencias

EN ESTA UNIDAD SE PRETENDE QUE EL ESTUDIANTE RECONOZCA LAS diferentes fases de la estadística, como son: el proceso de recolección de datos, organización de la información, representación y análisis. Cada etapa del trabajo estadístico requiere del reconocimiento de los datos y su tipología. La identificación clara de esta taxonomía es clave para la selección apropiada de las herramientas y procedimientos según sea el caso.

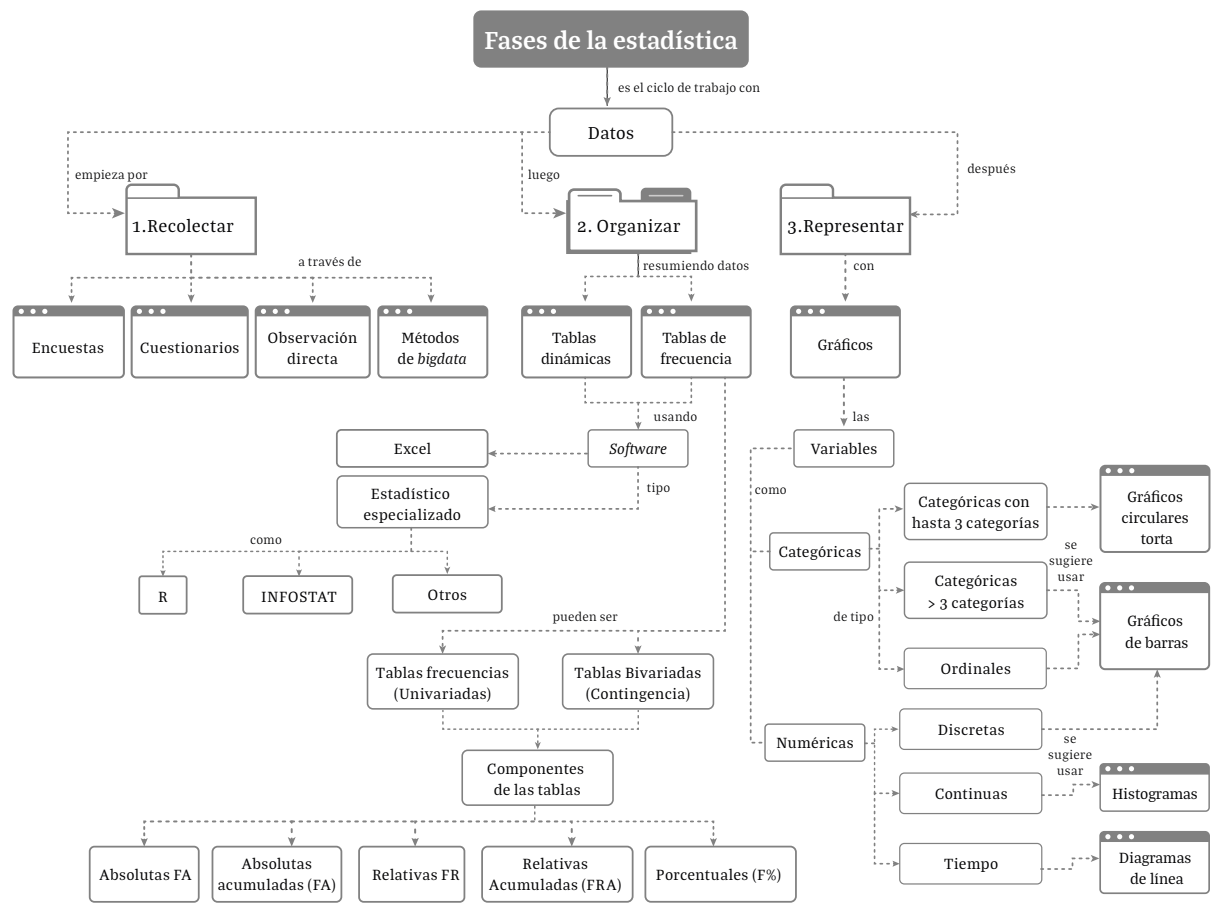
En el contexto de la cuarta revolución industrial, esa primera fase denominada recolección de datos es la puerta a la verdadera gestión del conocimiento. Los datos son un recurso tan valioso como el recurso humano, financiero y tecnológico, que dentro de un contexto tienen el potencial de llegar a ser conocimiento en la medida en que sean puestos para atender los óbices que surgen en la sociedad. En la actualidad existe información recolectada por otros, en la mayoría de los casos con otros intereses, que puede ser aprovechada; precisamente, el acceso abierto a la información se ha convertido en una política valiosa

y un propósito de los gobiernos. De ahí que cada vez más se suman esfuerzos con el fin de proporcionar al ciudadano la información importante para el desarrollo de la sociedad. En consecuencia, esta unidad se alinea con el propósito del *open government* o gobierno abierto, que es una forma de comunicación abierta, permanente y bidireccional entre la administración y los ciudadanos, basada en la transparencia por parte de la administración, la colaboración con la sociedad civil y las empresas y en la participación (Ferrer-Sapena *et al.*, 2011).

Las fuentes de información son múltiples. En el caso colombiano, el Dane es la entidad responsable de la planeación, el levantamiento, el procesamiento, el análisis y la difusión de las estadísticas oficiales de Colombia (Dane, 2021). El Dane –en articulación con entidades públicas y algunas privadas– produce operaciones estadísticas; las 10 entidades principales son: Banco de la República, Instituto de Hidrología, Meteorología y Estudios Ambientales (Ideam), Ministerio de Salud (Minsalud), Instituto Colombiano Agropecuario- ICA, Aerocivil, Ministerio de Comercio, Superintendencia de Servicios, Ministerio de Ciencias y el departamento Administrativo de la Función Pública (DAFP), aunque también otras 113 entidades producen información. Cabe resaltar que los datos se producen en múltiples escenarios.

Situación problema

La historia de la estadística evidencia el uso e importancia de los datos en la gestión pública; desde los censos realizados para la tributación en el antiguo Egipto hasta la actual investigación prospectiva aplicada a la economía de las naciones muestran la implementación de la estadística. Aunque la estadística ha sido aplicada desde tiempos remotos, ha evolucionado a la par que la tecnología; un claro ejemplo de ello son los paquetes estadísticos u hojas de cálculo, como Microsoft Excel, que han venido incorporando más herramientas que facilitan el tratamiento de datos. Así, el conocimiento y manejo de tales recursos resulta ser muy importante, dado que la sociedad produce constantemente una ingente cantidad de datos; entonces, la consecución y manejo de datos se tornan valiosos para la toma de decisiones en la administración.



Nota: para ver mejor el diagrama, escanear el código QR.

Figura 19. Ejes temáticos de la Unidad 2.
Fuente: elaboración propia. Este recurso lo puede encontrar en línea en Sierra (2021).

Lectura de motivación

La estadística y los datos abiertos

En su memorando sobre transparencia y gobierno abierto, el expresidente de los Estados Unidos de América, Barack Obama, señala tres puntos fundamentales sobre cómo debe ser el gobierno: (1) transparente, (2) participativo y (3) colaborador (The White House, 2009). Las nuevas tecnologías de la información permiten poner a disposición de la sociedad las operaciones y decisiones de los departamentos y organismos ejecutivos, lo que fomenta la transparencia. No obstante, se debe propender por que esas decisiones sean respaldadas por la comunidad –al incorporar activamente a los diferentes sectores de la sociedad– generando mayores oportunidades de participación para la formulación de políticas.

Entonces, los datos abiertos se han convertido en la revolución en materia de acceso a la información. Cuestiones como: ¿En qué hospital se atendieron más enfermos con síntomas de covid-19? ¿Cuáles son las instituciones educativas con más índice de ausentismo en una región? ¿Qué comunas de la ciudad presentan el mayor número de eventos delictivos?, son algunas de las que la administración pública está interesada en analizar. Y es que, precisamente, la gestión *per se* de datos, que va desde su preparación hasta pulirlos, reestructurarlos y documentarlos, pertenece a las buenas prácticas, pues entre otros beneficios se previene la pérdida de datos valiosos que en muchos casos se mudan cuando el funcionario se va.

Según la Cepal, algunas de las fuentes de información estadística son:

- Oficinas Nacionales de Estadística (ONE)
- Bancos Centrales
- Programas y Agencias de Naciones Unidas
- Organismos regionales e internacionales

En Colombia, quien ejerce el rol de oficina nacional de estadística es el Departamento Administrativo Nacional de Estadística (Dane); en la región latinoamericana, se reconocen ocho ONE adscritas a diferentes países, entre ellos: Brasil, Chile, Colombia, Cuba, Dominica, Ecuador, El Salvador y Guatemala.



Imagen de DANIEL DIAZ en Pixabay (2021). Puede ver licencia en: <https://pixabay.com/es/illustrations/big-data-base-de-datos-almacenar-2103091/>

Actividad de presaberes



¿Qué instrumentos conoces para recolectar información?

.....

.....

.....

.....

.....



¿Qué es una base de datos?

.....

.....

.....

.....

.....



¿Qué son datos anonimizados?

.....

.....

.....

.....

.....

Organización y análisis de la información

En el capítulo anterior se planteó el concepto de estadística y sus fases: recolectar, organizar, representar y analizar datos; en esta sección se abordan esas etapas presentando las herramientas y técnicas estadísticas usadas en cada una de ellas.

Recolectar datos

Entre los diferentes instrumentos para recolectar datos se encuentran:

- Sistematizadores de la observación
- Esquemas de bases de datos con diferentes fuentes
- Guion de entrevistas
- Cuestionario de encuestas

Sistematizadores de la observación

Estos instrumentos facilitan la recopilación de información producto de la observación: formatos de diario de campo, bitácoras, planillas e historias clínicas, etc. Un buen ejemplo es el proceso que realiza el

profesional de área de la salud en consulta cuando consigna las observaciones y datos del estado de salud de sus pacientes en un formato de historia clínica; algunos de esos datos son el valor de la presión arterial, peso y altura. Esos datos son obtenidos a través de la observación.

Esquemas de bases de datos

En este grupo se pueden ubicar las matrices con diferentes fuentes de información o bases de datos que se construyen a partir de estadísticas consolidadas por otras entidades, o simplemente datos ubicados por separado reportados por fuentes fidedignas como las entidades de estadísticas oficiales. Por ejemplo, una variable a consultar puede ser el Índice de Percepción de Corrupción (IPC) que publica la organización Transparencia Internacional, mientras que otra variable de interés quizá sea el número de condenas por corrupción en el país. Esta información se puede solicitar a las entidades judiciales con la competencia o en sus sitios web directamente, dependiendo de la apertura de datos que manejen. En estas matrices, las columnas representan las variables que se construyen y las filas las unidades de análisis de observación; para el caso referido, una columna se destinaría para el IPC, otra para el número de condenas y las filas serían cada uno de los países a los que se les analizan esas características.

Guion de entrevista

La entrevista es un proceso en el cual las unidades de observación son las personas. A través de la interlocución de un entrevistador y un entrevistado se formulan preguntas sin la obligatoriedad de dar opciones de respuesta al entrevistado; aunque existe la entrevista semiestructurada y estructurada, ambas permiten expresar la opinión sin limitar a unas opciones de respuesta específica. De un modo bastante sencillo se puede decir que la entrevista, a diferencia de la encuesta, es más flexible y le permite mayor espontaneidad al entrevistado que al encuestado. El instrumento usado para este proceso se conoce como guion de la entrevista.

Cuestionario de encuesta

Su etimología viene del latín *in quaerere*, que se traduce como indagar, transformándose al francés en *enquête* y posteriormente al español en *encuesta*. La encuesta es una técnica con la cual se recopila información de los sujetos de estudio a través de preguntas; esta información

pueden ser opiniones, actitudes y comportamientos. Las preguntas se consignan en un instrumento impreso o digital denominado *cuestionario*.

A modo de ejemplo, en la figura 20 se ilustra el proceso completo, que consta de seis pasos para llegar finalmente al cuestionario de la encuesta, que es el instrumento de recolección de información propiamente dicho.

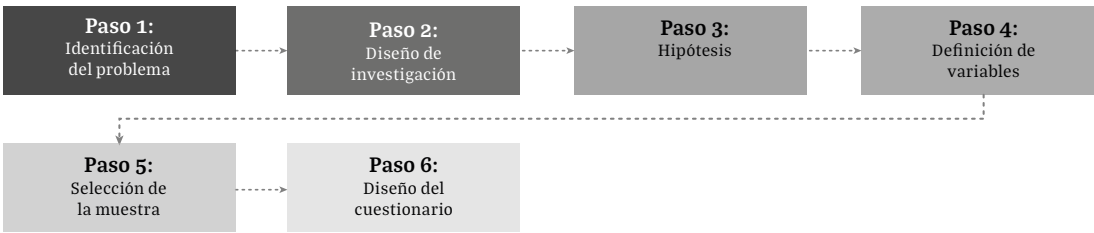


Figura 20. Planificación de la encuesta
Fuente: elaboración propia. Sierra (2022)

Paso 1: Identificación del problema

Aquí se debe exponer el problema que se va a trabajar, indicando el soporte (cifras, informes y fuente) que lo hace evidentemente un problema.

Ejemplo 2.1

Problema: Desempleo en Ibagué

Los indicadores históricos de desempleo para Colombia muestran que al cierre del 2017 el desempleo aumentó 0,2 % con relación al año inmediatamente anterior, pasando de 9,2 % a 9,4 %. (Departamento Nacional Administrativo de Estadística- Dane, 2018), mientras que en lo transcurrido del 2018 se posicionó en 10,8 %, con un total de 2 878 203 personas desempleadas. Estas cifras se han mantenido por los últimos 30 años, siendo del 12 % en 1997 y 9,8 % en 2007 (Banco de la República, 2017). Así, las estadísticas nacionales indican que una de cada 10 personas dispuestas a trabajar no encuentra donde emplearse.

A nivel regional, Ibagué presenta altas cifras de desempleo. Algunos reportes de revistas económicas señalan que las ciudades con altos índices de desempleo se ven afectadas significativamente en las condiciones sociales. Las ciudades con mayor porcentaje de desocupados frente a su población económicamente activa son: Quibdó (19,1 %); Armenia (19 %); Ibagué (17,9 %); Popayán (17,3 %) y Pereira (16,9 %) (Gutiérrez, 2011).

En Ibagué, 2 de cada 10 personas dispuestas a trabajar no encuentra donde emplearse. Entre las causas del desempleo los analistas han enumerado la llegada de población desplazada; las estadísticas indican que a partir del 2008 Ibagué se convierte en la segunda ciudad con más desplazados después de Bogotá (*Portafolio*, 2008). Otro aspecto de cómo el desempleo afecta a la economía es la disminución en el gasto de los consumidores, lo que a su vez conduce a la deflación provocada por la demanda mínima y la oferta de productos y servicios. El poco flujo de dinero incide en el progreso social, en ámbitos como vivienda, educación y salud, entre otros. Según el censo del 2005, en Ibagué sólo el 51 % de la población tenía vivienda propia. En el 2017 (por factores ya mencionados como el desplazamiento) la tenencia es de 23 %, y según el SUI de la Superintendencia de Servicios Públicos (2016) el 62,1 % de los hogares son de estrato 1 y 2, mientras que 25,6 % es de estrato 3, sumando así el 87,7 % en estos tres niveles.

Por otro lado, un detonante del desempleo en Ibagué es el débil tejido empresarial. Es evidente cómo la demanda disminuye paulatinamente cada año; la Cámara de Comercio reporta un alto número de empresas que se disuelven (Centro de estudios de la construcción y el desarrollo urbano y regional [Cenac], 2016)

Paso 2: Diseño de investigación

En función de lo que busca con su trabajo y los alcances que este tendrá, es necesario definir el método de investigación. Cuatro grandes grupos en los que puede ubicarse son:

Analítico experimental	Analítico observacio- nal o correlacional	Descriptivo	Explicativo
Si usted va a tener control sobre las variables o las va a manipular.	Las variables de interés son seleccionadas para conocer la relación que existe entre ellas, aprovechando su presencia o ausencia en grupos de sujetos escogidos cuidadosamente, de modo que sea posible el control sobre las variables identificadas por el investigador.	No se establecen relaciones de causa-efecto. Por lo general todo estudio comienza siendo descripti-vo. Acá sólo se describe el comportamiento o fenómeno, sin analizar la causa o consecuencia.	Busca una explicación y determinación de los fenómenos. Se pueden aplicar estudios de tipo predictivo en donde se pueda establecer una relación causal entre diversas variables.

Figura 21. Diseño de la investigación
Fuente: Hernández-Sampieri y Mendoza (2018).

Paso 3: Hipótesis

Una hipótesis es una conjetura o un supuesto de algo. En este caso, las hipótesis de la investigación son esas causas, consecuencias o factores asociados al problema. Para construir una hipótesis, se sugieren los siguientes pasos:

- **Una lluvia de ideas:** coloque en forma de lista todas las ideas que tenga del problema.
- **Paralelo entre causas y consecuencias:** realice una tabla con dos columnas, frente a cada causa del problema coloque sus respectivas consecuencias.
- **Preguntas problema:** redacte en forma interrogativa cada causa y consecuencia, y seleccione las que abarquen el problema de forma general; mínimo una, máximo tres.

Ejemplo 2.2

Lluvia de ideas

- ☐ Población desplazada. “Ese es un problema social grave”
- ☐ La desigualdad de género
- ☐ La inexperiencia laboral
- ☐ La falta de estudios (nivel de escolaridad)
- ☐ Por la inconformidad salarial
- ☐ La corrupción
- ☐ Los avances tecnológicos

Paralelo

Causa	Consecuencia
• Falta de formación académica	• Mano de obra calificada que se trae de otro lugar • Los mejores empleos son para personas de otras ciudades
• Inexperiencia laboral	• Inestabilidad laboral • Altas rotaciones en empleos
• Desplazamiento	• Aumento de índices de desempleo • Inseguridad • Disminución de salarios por competencia

Preguntas problema

- ¿La formación académica incide a la hora de conseguir empleo?
- ¿Los índices de desempleo en la juventud están creciendo por la falta de experiencia laboral?
- ¿La mayor parte de personas desempleadas son desplazadas?
- ¿En Ibagué se presenta la desigualdad de género a la hora de buscar empleo?
- ¿La corrupción está afectando la tasa del nivel del desempleo en la ciudad de Ibagué?
- ¿La aspiración salarial es un factor que incide en la tasa de desempleo?
- ¿En Ibagué se presenta desempleo por el rezago del capital humano frente a la automatización de procesos o la incursión tecnológica de las empresas?

Hipótesis

- Bajos niveles de formación académica están asociados al estado laboral “desocupado”
- A mayor experiencia laboral mayor probabilidad de conseguir empleo.
- La proporción de personas desempleadas es mayor en el grupo de desplazados frente al de no desplazados
- La proporción de personas desempleadas es mayor en las mujeres que en los hombres.

Paso 4: Definición de variables (tabla de variables)

Identificados los factores detonantes y consecuencias del problema, asócielos con una variable, que bien puede ser numérica o continua. Para ello, construya una tabla con cinco columnas, así:

Ejemplo 2.3

Tabla 5. Definición de variables

Variable	Descripción	Tipo	Categorías/intervalos	Códigos
Escolaridad	Corresponde al máximo nivel de escolaridad alcanzado por la persona	Nominal	• Ninguno	• V1C1
			• Preescolar	• V1C2
			• Básica Primaria (1º - 5º)	• V1C3
			• Básica secundaria (6º--9º)	• V1C4
			• Media (10º--13º)	• V1C5
			• Técnico sin título	• V1C6
			• Técnico con título	• V1C7
			• Tecnológico sin título	• V1C8
			• Tecnológico con título	• V1C9
			• Universitario sin título	• V1C10
			• Universitario con título	• V1C11
			• Postgrado sin título	• V1C12
			• Postgrado con título	• V1C13

Variable	Descripción	Tipo	Categorías/intervalos	Códigos
Vocación	Corresponde a la vocación económica del municipio	Nominal	• Turístico • Minero • Agrícola • Industrial	• V1C1 • V1C2 • V1C3 • V1C4
NBI	Corresponde al porcentaje de la población con necesidades básicas insatisfechas	Continua	0 a 100%	NA

Nota: solo la variable escolaridad sería adecuada cuando la unidad de análisis (UA) sean las personas. Vocación y NBI son variables que se pueden usar si la UA es municipio.

- **Variables:** es el nombre que se le da a la característica (usar una o máximo dos palabras).
- **Descripción o significado:** corresponde a la descripción de la característica.
- **Tipo:** según la clasificación del tipo de variables: numérica o categórica. De estas se desprenden unas subclasificaciones.
- **Categorías/o intervalo:** las diferentes opciones que puede tomar la variable categórica o los rangos de valores en los que se mueve la variable numérica. Es importante investigar los términos adecuados en cada caso.¹⁰
- **Códigos:** son propios de las variables categóricas, para facilitar el trabajo a la hora de sistematizar la información. Se asigna un número a cada categoría, empezando desde 0 o 1, de manera que

¹⁰ Según la variable que desee abordar es importante buscar fuentes oficiales que usen los términos técnicos y adecuados en cada caso. Algunos términos incorrectos son: discapacitados, desempleado-empleado, EPS-Sisbén, entre otros. En caso de que use la variable *situación laboral*, el Dane tiene definida dentro de la PEA (Población económicamente activa) las siguientes categorías: Ocupados, Desocupados, Desempleo abierto y otros (Lora, 2016).

al pasar los resultados del cuestionario aplicado, pueda colocar el número en lugar de la palabra. En el ejemplo de la tabla 5, en la variable vocación el código V1C1 indica que es la primera variable (vocación) y la primera categoría (turismo). Cuando se trabajan variables numéricas no se usa el término categoría, sino intervalo; en el caso de la segunda variable (NBI) este porcentaje puede oscilar entre 0 y 1, y por ello en la casilla códigos aparece que no aplica.

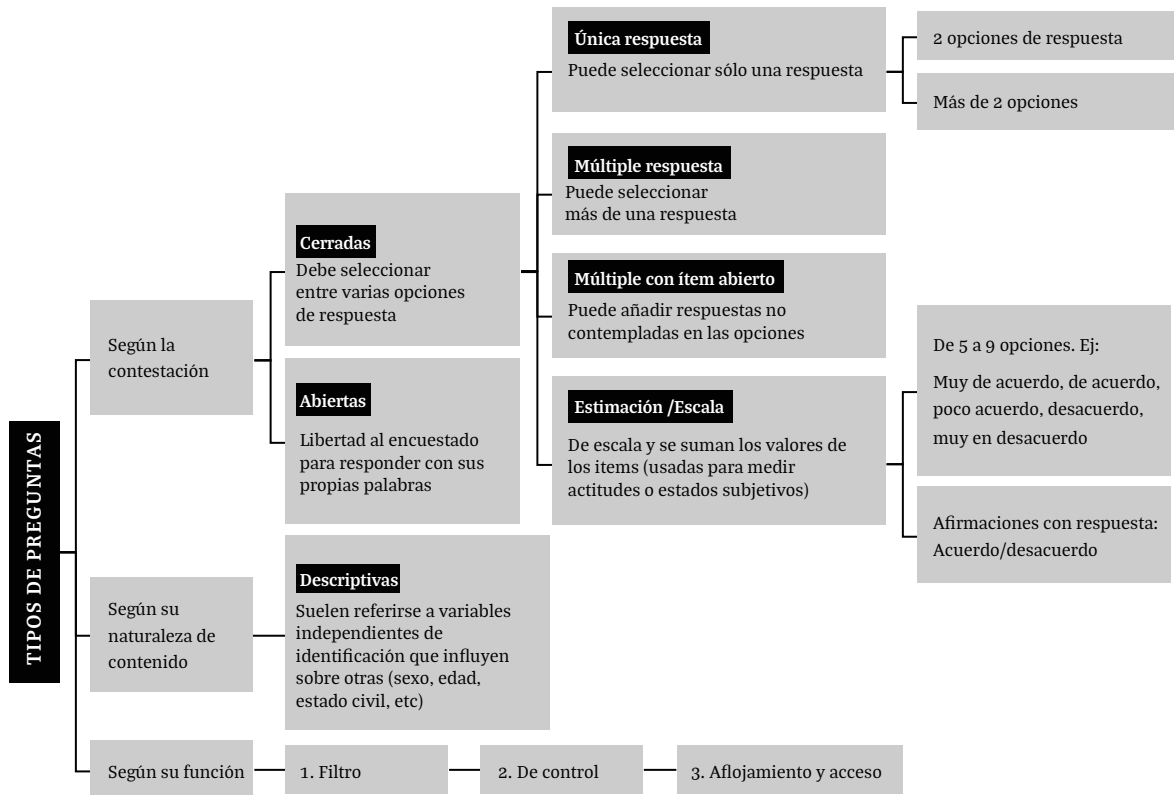
Paso 5: Muestra

Definir el tipo de muestreo y, a partir de este, el tamaño de la muestra.

Tipo de muestreo	Tamaño de muestra
Según el tipo de estudio y la facilidad para acceder a las unidades de observación dentro de la población puede escoger entre: • Muestreo probabilístico • Muestreo no probabilístico No tiene sentido hablar de fórmula para calcular el tamaño de la muestra si aplicó un muestreo no probabilístico.	Calcular cuántas personas o unidades de observación estudiará a través de una fórmula, para lo cual ya existen calculadoras en línea que lo hacen por usted, como la de Openepi (escanee el código para ir a la calculadora).

Paso 6: Diseño del cuestionario

El cuestionario es el instrumento básico utilizado en la investigación por encuesta. Se puede decir que es el “documento que recoge de forma organizada los indicadores de las variables implicadas en el objetivo de la encuesta”; por ello, las preguntas iniciales del cuestionario se construyen a partir de las hipótesis y las variables previamente establecidas (Casas Anguita *et al.*, 2003).



Nota: las preguntas filtro ayudan a seleccionar una parte de los encuestados para que contesten preguntas elaboradas solo para estos. Las preguntas de control corroboran la congruencia de las respuestas dadas por el encuestado. Y las preguntas de acceso buscan generar un ambiente para que el encuestado se acondicione y se le facilite responder.

Figura 22. Tipos de preguntas que se pueden incluir en un cuestionario

Fuente: elaboración propia a partir de Casas Anguita *et al.* (2003).

Formato definitivo

Después de realizar el estudio piloto y las correcciones oportunas, se podrá elaborar el formato definitivo del cuestionario, el cual debe incluir los siguientes apartados:

- Identificación del organismo que lleva a cabo la investigación.
- Título completo del estudio en el que se enmarca el cuestionario.
- Declaración explícita de que la información que se facilite va a ser tratada con máxima confidencialidad.
- Espacio para la fecha de cumplimentación del cuestionario (dato especialmente relevante en estudios longitudinales).

- Instrucciones para la adecuada cumplimentación.
- Frase de agradecimiento al encuestado por el esfuerzo realizado.

Práctica 3: cuestionario problemáticas de interés

Para poner en práctica el fundamento teórico que se expone en esta unidad, se desarrollará un cuestionario para describir las afectaciones de la pandemia en el sector económico de los municipios en Colombia.

Ejemplo 2.4

0.6 - Según los ingresos este negocio es...

En Colombia el segmento empresarial esta clasificado en micro, pequeñas, medianas y grandes empresas, esta clasificación está reglamentada en la Ley 590 del 2000, conocida como la Ley Mipymes y sus modificaciones (Ley 905 de 2004). Clasificación de la empresas, 2021.

Mano­factura:

- Micro­empresa (hasta \$839 007 741)
- Pequeña: (Superior a \$839 007 741 y hasta \$7 299 256 965)
- Mediana: (Superior a \$7 299 256 965 y hasta \$61 833 869 966)

Servicios:

- Micro­empresa (hasta \$1 174 603 716)
- Pequeña
- Mediana

Comercio

- Micro­empresa (hasta \$1 594 089 783)
- Pequeña
- Mediana

☐

 Micro­empresa

☐

 Pequeña empresa

☐

 Mediana empresa

Nota: los cuestionarios deben incluir siempre unas instrucciones sencillas para su correcta cumplimentación. En algunos casos puede ser oportuno incluir algunos ejemplos demostrativos para que los sujetos sepan exactamente qué tarea deben realizar. En la prueba piloto también debe comprobarse la calidad de estas instrucciones, solicitando a la persona que responde el cuestionario que lea en voz alta las preguntas; se debe tomar nota de los ítems que tuvo que releer o en los cuales pidió ayuda al investigador para responder, para replantearlos o eliminarlos, según sea el caso

Figura 23. Ejemplo de cuestionario

Fuente: elaboración propia. Puede explorar el formulario desde <https://arcg.is/108LOLO>

Organizar datos

Este apartado aborda de forma más detallada la fase de organización de la información. Aquí se procede a organizar la información recolectada en la anterior etapa. Los datos se deben presentar de forma sintética en un esquema matricial de filas y columnas denominado *tabla de frecuencias*.

Tablas de frecuencia para datos categóricos

Este tipo de tabla resume información de la base de datos, refiriendo en cada columna los títulos de: *variable*, *frecuencia absoluta*, *frecuencia relativa* y *frecuencia porcentual*. Las tablas de frecuencias tendrán un total máximo de 10 filas (sin incluir la fila de títulos), en donde las frecuencias absolutas indican el número de unidades de observación en esa fila, es decir, la cantidad total de unidades de observación que están en esa categoría. Las unidades de observación en la frecuencia absoluta pueden variar y pueden ser personas, instituciones, hogares o municipios, dependiendo del contexto de los datos. La frecuencia absoluta simplemente cuenta cuántas veces aparece cada unidad de observación en el conjunto de datos.

Por otro lado, la frecuencia relativa proporciona una perspectiva proporcional. Representa la proporción de cada categoría en relación con el total de unidades de observación en el conjunto de datos. Esto permite comprender mejor la distribución de las categorías en función del tamaño del conjunto de datos.

La frecuencia porcentual va un paso más allá y expresa la frecuencia relativa en términos de porcentaje. Es útil para comunicar de manera efectiva la proporción de cada categoría en relación con el total, ya que los porcentajes son una forma común y comprensible de representar proporciones en contextos numéricos. Observe los elementos de una tabla de frecuencias para datos categóricos en la figura 24.

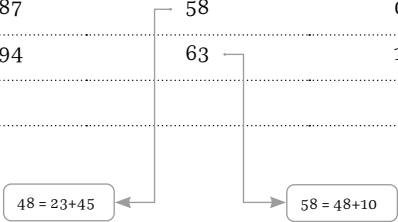
Tablas de frecuencia para datos numéricos

Además de las columnas presentadas en la tabla para datos categóricos, este tipo de tabla tiene otras columnas como: frecuencia absoluta acumulada y frecuencia relativa acumulada.

Tabla 7. Ingresos de clientes del Banco A

Variable: ingresos			Frecuencia absoluta (f _i)	Frecuencia relativa (fr _i)	Frecuencia absoluta acumulada (F _i)	Frecuencia relativa acumulada (FR _i)
i	Límite inferior	Límite superior	# personas	Proporción de personas	# personas que ganan menos de...	% personas que ganan menos de...
1	\$500 000	\$600 000	23	0,3700	23	0,37
2	\$ 600 001	\$ 700 000	25	0,4000	48	0,76
3	\$ 700 001	\$ 800 000	10	0,1587	58	0,92
4	\$800 001	\$ 900 000	5	0,0794	63	1,00
			63	1		

Fuente: elaboración propia. Sierra (2022)



En la tabla de frecuencias es importante diferenciar la variable de las unidades de observación. En la tabla 7, la variable es ingresos en pesos y las unidades de observación son personas. El contenido de las columnas límite inferior y superior siempre corresponde a valores de la variable analizada, mientras que los valores en las columnas de las frecuencias son el número de unidades de observación, es decir, personas con ingresos que oscilan en el intervalo al que se haga referencia. Veamos:

De la frecuencia absoluta

- $f_1 = 23$ Significa que hay 23 personas que tienen ingresos entre 500 mil y 600 mil pesos
- $f_2 = 25$ Significa que hay 25 personas que tienen ingresos mayores a 600 mil pesos, pero menores o iguales a 700 mil pesos
- $f_3 = 10$ Hay 10 personas que tienen ingresos mayores a 700 mil pesos, pero menores o iguales a 800 mil pesos
- $f_4 = 5$ Hay 10 personas que tienen ingresos mayores a 800 mil pesos, pero menores o iguales a 900 mil pesos

De la frecuencia relativa

- $fr_1 = 0,37$

El 37% tiene ingresos entre 500 mil y 600 mil pesos
- $fr_2 = 0,40$

Un 40% tiene ingresos mayores a 600 mil pesos, pero menores o iguales a 700 mil pesos
- $fr_3 = 0,1587$

Un 15,87% de la población tiene ingresos entre 700 mil y 800 mil pesos
- $fr_4 = 0,0794$

Un 7,94% tiene ingresos mayores a 800 mil pesos, pero menores o iguales a 900 mil pesos

De la frecuencia absoluta acumulada

- $F_1 = 23$

Hay 23 personas que tienen ingresos entre 500 mil y 600 mil pesos
- $F_2 = 48$

Hay 48 personas que tienen ingresos que van desde 500 mil hasta 700 mil pesos (o simplemente hay 48 personas que tienen ingresos por debajo de 700 mil pesos)
- $F_3 = 58$

Hay 58 personas que tienen ingresos que van desde 500 mil hasta 800 mil pesos
- $F_4 = 63$

Hay 63 personas que tienen ingresos que van desde 500 mil hasta 900 mil pesos

Tablas de doble entrada

En estas tablas se organizan las frecuencias de dos variables de forma simultánea.

Tabla 8. Tabla de doble entrada

Variable x	X1	X2	X3	Subtateles	
Variable y					
Y ₁	f ₁₁	f ₁₂	f ₁₃	$\sum_{j=1}^k f_{1j}$	Suma de elementos de las categorías i= 1,2,3...k de la variable x, solo en la categoría 1 de la variable y.
Y ₂	f ₂₁	f ₂₂	f ₂₃	$\sum_{j=1}^k f_{2j}$	
Y ₃	f ₃₁	f ₃₂	f ₃₃	$\sum_{j=1}^k f_{3j}$	
Y _n	f _{n1}	f _{n2}	f _{n3}	$\sum_{j=1}^k f_{nj}$	Suma de elementos de las categorías i= 1,2,3...k de la variable y, solo en la categoría 3 de la variable x.
	$\sum_{i=1}^k f_{i1}$	$\sum_{i=1}^k f_{i2}$	$\sum_{i=1}^k f_{i3}$		

Fuente: elaboración propia. (Sierra, 2022).

Tabla 9. Tabla de doble entrada (Tipo de sangre y género)

Suma de elementos de las categorías masculino y femenino de la variable sexo, solo en la categoría A de la variable tipo de sangre.

	Sexo	Masculino	Femenino	Total
Tipo de sangre				
A		18	33	51
B		22	17	39
AB		34	20	54
O		66	90	156
Total		140	160	300

Suma de elementos de las categorías A, B, AB, O de la variable tipo de sangre, solo la categoría Femenino de la variable sexo.

Fuente: elaboración propia.

Tablas de frecuencia en Excel

Para realizar una tabla de frecuencia en Excel debe verificar primero el tipo de dato que desea organizar en dicha tabla, pues de ello depende el tipo de herramienta a usar; para las variables categóricas se sugiere usar la herramienta de tablas dinámicas, y para las variables numéricas se sugiere obtener las tablas de frecuencia a través de la herramienta análisis de datos.

Práctica 4: tablas de frecuencia en Excel para variables categóricas. (Delitos sexuales)

La tabla 10 representa los reportes de delitos sexuales en 12 municipios del departamento de Antioquia. Realice una tabla de frecuencias usando la herramienta de tablas dinámicas en Excel.

Tabla 10. Delitos sexuales

MUNICIPIO	GÉNERO	*AGRUPA_EDAD_PERSONA
AMAGÁ	FEMENINO	ADOLESCENTES
AMAGÁ	FEMENINO	ADULTOS
AMAGÁ	MASCULINO	ADOLESCENTES
AMAGÁ	FEMENINO	MENORES

MUNICIPIO	GÉNERO	*AGRUPA_EDAD_PERSONA
ANDES	FEMENINO	ADULTOS
ANDES	FEMENINO	ADULTOS
APARTADÓ	FEMENINO	ADOLESCENTES
APARTADÓ	FEMENINO	MENORES
BELLO	FEMENINO	ADOLESCENTES
BELLO	FEMENINO	ADULTOS
BETANIA	MASCULINO	MENORES
CÁCERES	FEMENINO	ADOLESCENTES
CALDAS	FEMENINO	ADOLESCENTES
CALDAS	FEMENINO	ADULTOS
CARACOLÍ	FEMENINO	ADOLESCENTES
CAREPA	FEMENINO	ADULTOS
CAREPA	FEMENINO	ADULTOS
CAROLINA	MASCULINO	ADOLESCENTES
CAUCASIA	FEMENINO	ADULTOS
CAUCASIA	FEMENINO	ADOLESCENTES
CAUCASIA	FEMENINO	ADULTOS
CAUCASIA	FEMENINO	ADULTOS
CAUCASIA	FEMENINO	ADULTOS
COCORNÁ	FEMENINO	ADULTOS
CONCORDIA	FEMENINO	ADULTOS

Fuente: obtenido del sitio web de la Policía Nacional de Colombia, “Estadísticas delictivas” (2022).

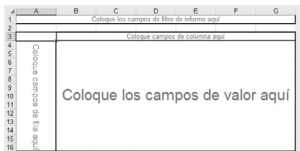
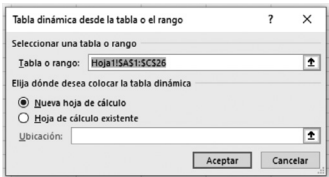
El siguiente es el paso a paso de la práctica:

1. Seleccione la tabla donde está la información.
2. En el menú principal, vaya a la pestaña “Insertar” y elija “Tabla dinámica”, después seleccione “De una tabla o rango”.

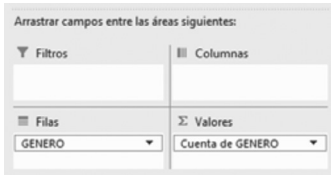
	A	B	C
1	MUNICIPIO	GÉNERO	*AGRUPO_EDAD_PERSONA
2	AMAGÁ	FEMENINO	ADOLESCENTES
3	AMAGÁ	FEMENINO	ADULTOS
4	AMAGÁ	MASCULINO	ADOLESCENTES
5	AMAGÁ	FEMENINO	MENORES
6	ANDES	FEMENINO	ADULTOS
7	ANDES	FEMENINO	ADULTOS
8	APARTADO	FEMENINO	ADOLESCENTES
9	APARTADO	FEMENINO	MENORES



3. Asegúrese de que la ventana emergente de “Tabla o rango” abarque toda la región que contiene sus datos, que en este ejemplo es la Hoja 1 desde A1 hasta C26. Si no es así, ajuste el área seleccionada utilizando la flecha hacia arriba y confirme haciendo clic en “Aceptar”.
4. Debe aparecer una hoja nueva y la siguiente imagen:



5. En el lado derecho de la pantalla, observe una serie de campos como se muestra a continuación. Con el cursor, arrastre la variable GÉNERO hacia el campo “Filas”.
6. Luego, repita el procedimiento llevando GÉNERO al campo “Valores”. Verifique que en valores diga cuenta de GÉNERO.



El resultado final debe ser:

	A	B
1	Coloque los campos de filtro de informe aquí	
2		
3	Cuenta de GÉNERO	
4	GÉNERO	Total
5	FEMENINO	22
6	MASCULINO	3
7	Total general	25

La salida final de este proceso es una tabla de frecuencias para la variable de género creada con la herramienta “Tabla dinámica” de Excel. A partir de esta tabla, se puede observar que en la base de datos de delitos sexuales de la Policía Nacional de Colombia hay un total de 22 registros correspondientes al sexo femenino y 3 al sexo masculino. Con estas frecuencias absolutas como punto de partida, es posible calcular las frecuencias relativas y los porcentajes para cada una de las categorías. Es importante destacar que en este caso se utilizó la tabla dinámica debido a que la variable “Género” es de naturaleza categórica.

Práctica 5: tablas de frecuencia en Excel, variables numéricas (NBI)

La tabla 11 representa el índice de necesidades básicas insatisfechas (NBI) de los residentes en los municipios de Colombia. La información la puede consultar en la sección de estadísticas por tema del Dane en Microdatos:

Sección estadística por tema:
Pobreza y condiciones de vida

Microdatos
para descargar en Excel:



La base de datos que descarga con el segundo de los anteriores códigos QR es un archivo Excel que tiene una pestaña con el nombre: “MUNICIPAL”, seleccione la fila 10 y aplique un filtro de modo que en la columna B “Departamento” filtre “Chocó”; en la columna D “Autorreconocimiento étnico” filtre “Ningún grupo étnico”. Obtendrá así la tabla 11.

Tabla 11. NBI en el departamento del Chocó (población sin ningún grupo étnico)

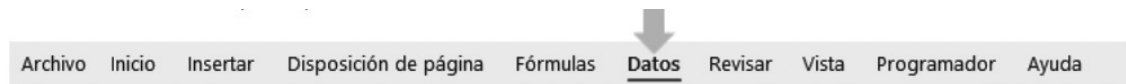
Municipio	NBI	Personas en miseria (%)	Componente Vivienda (%)	Componente Servicios (%)	Componente Hacinamiento (%)	Componente Inasistencia (%)	Componente Dependencia económica (%)	Total personas en hogares particulares
Quibdó	61,57	6,41	1,88	58,86	3,11	1,79	3,11	3245
Acandí	35,26	10,39	14,05	24,19	2,56	1,36	8,09	1174
Alto Baudó	38,10	9,09	7,79	31,60	3,46	5,63	0,43	231
Atrato	30,00	7,14	1,43	27,14	5,71	0,00	2,86	70
Bagadó	64,00	26,00	22,00	50,00	6,00	12,00	0,00	50
Bahia Solano	5,26	0,44	3,07	2,63	0,00	0,00	0,00	228
Bajo Baudó	47,54	14,75	6,56	44,26	8,20	1,64	9,84	61

Municipio	NBI	Personas en miseria (%)	Componente Vivienda (%)	Componente Servicios (%)	Componente Hacinamiento (%)	Componente Inasistencia (%)	Componente Dependencia económica (%)	Total personas en hogares particulares
Bojayá	93,47	47,77	6,23	88,72	41,84	5,64	13,35	337
El Cantón del San Pablo	70,45	22,73	9,09	45,45	15,91	0,00	22,73	44
Carmen del Darién	81,08	28,83	26,13	69,37	5,41	5,41	12,61	111
Cértegui	34,69	2,04	12,24	24,49	0,00	0,00	0,00	49
Condoto	53,51	16,23	4,39	51,75	9,21	2,63	10,09	228
El Carmen de Atrato	13,87	2,03	1,32	1,62	2,34	1,73	9,56	5134
El Litoral del San Juan	55,36	26,79	19,64	48,21	7,14	5,36	1,79	56
Istmina	85,12	7,96	1,17	83,16	4,70	0,39	3,66	766
Juradó	52,94	0,00	0,00	52,94	0,00	0,00	0,00	17
Lloró	65,22	36,23	5,80	62,32	37,68	0,00	18,84	69
Medio Atrato	94,23	18,75	12,50	91,59	0,00	5,29	7,21	416
Medio Baudó	82,30	33,01	29,67	64,11	13,88	3,83	19,14	209
Medio San Juan	38,24	13,24	0,00	27,94	17,65	2,94	4,41	68
Nóvita	29,47	3,16	1,05	29,47	0,00	0,00	2,11	95
Nuquí	17,14	2,86	2,86	0,00	11,43	2,86	2,86	35
Río Iró	18,75	0,00	6,25	12,50	0,00	0,00	0,00	16
Río Quito	75,00	68,75	9,38	71,88	21,88	21,88	46,88	32
Riosucio	67,19	33,48	41,23	44,57	7,63	7,75	12,47	3408
San José del Palmar	20,26	6,13	4,52	4,04	5,86	5,26	11,19	2967
Sipí	41,38	13,79	0,00	41,38	0,00	0,00	13,79	29
Tadó	43,79	3,79	3,79	33,45	2,41	2,07	6,90	290
Unguía	37,89	17,68	23,05	19,89	5,26	2,21	10,84	950
Unión Panamericana	21,41	2,49	1,04	8,73	6,03	2,91	5,20	481

Fuente: elaboración propia a partir de Microdatos del DANE.

Copia esta base de datos en un archivo de Excel; en ese archivo crea una hoja y nómbrala así: NBI-Choco. Para este caso se usará la opción “Análisis de datos” de Excel, pues la variable NBI es numérica, así como lo es % personas en miseria, Componente vivienda, etc.

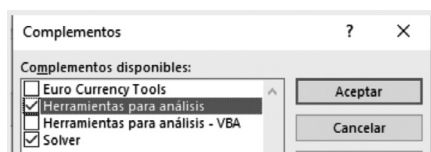
Localiza en el menú principal la ficha Datos



1. En Datos selecciona la opción Análisis de datos. Si no está activa, debes ir primero al menú principal a: Archivo → opciones → complementos → ir...

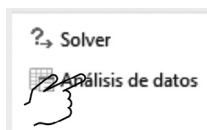


y activar la casilla **herramientas para análisis**

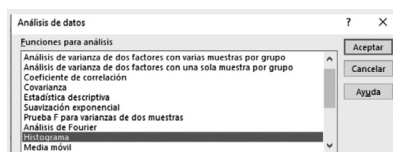


Ahora ya podrás ver que en el menú principal, en la ficha datos, se activará la opción **Análisis de datos**

2. Regresa al menú principal → DATOS → Análisis de datos:



En la ventana emergente busca la opción **Histograma**. Haz clic sobre la palabra, para que aparezca sombreada como se ve en la imagen, y luego en Aceptar.

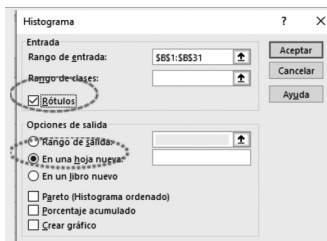


3. En la ventana emergente en Rango de entrada haz clic sobre la flecha que apunta hacia arriba y, luego de que esté minimizada esa ventana, selecciona toda la columna de NBI (incluyendo el título).



Dé clic de nuevo en la flecha que ahora apunta hacia abajo

4. Debe quedar ahora en Rango de entrada \$B\$1:\$B\$31, que indica que fueron seleccionadas las casillas B1 hasta B31. Selecciona la casilla de verificación de Rótulos hasta que quede la marca de verificación (o palomita) señalada. En opciones de salida selecciona “En una hoja nueva”; luego, haz clic en Aceptar.



El resultado aparecerá en una hoja nueva y debe ser:

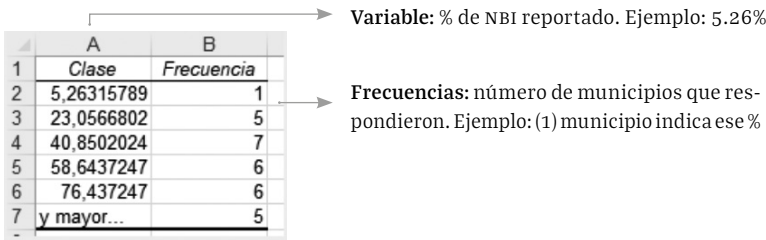


Figura 25. Salida análisis de datos NBI (%) municipios del Chocó según DANE (2018)

Aquí, la columna **Clase** es la variable (%NBI) y la columna **Frecuencia** es la cantidad de unidades de observación (municipios) que tienen ese valor de la variable. Así, se deduce que la variable %NBI es hasta 5,26% en un 1 municipio; desde 23,057% hasta 40,850% en 7 municipios; y hay 5 municipalidades del departamento del Chocó que tienen más de un 76,437% de su población con NBI. Recuerda que la columna **Clase** NO siempre será porcentaje; en este ejemplo la variable analizada era % NBI, pero en otro caso pueden ser los ingresos, situación en la cual se leería que se perciben ingresos hasta de 5,26 millones en un 1 municipio, que el ingreso oscila entre 23,057 y 40,850 millones en siete municipios, etc.

Lo anterior sugiere que cuando la variable se mide de forma porcentual, a veces se confunde con las frecuencias porcentuales (que es otra cosa), las cuales hacen referencia al porcentaje de unidades de observación que caen en cada intervalo. Para ver la diferencia entre una variable que mide porcentajes y las frecuencias porcentuales, se realizarán en la hoja de Excel de la salida anterior (figura 25) a partir de la columna C las frecuencias relativas (Fr) y porcentuales (F%). Empezé cambiando el nombre a la hoja de Excel donde quedó la tabla que obtuvo; llámela **Tabla**.

5. Completa la tabla con la información de las columnas C y D como se indica:

	A	B	C	D
1	Clase	Frecuencia	Fr	F%
2	5,26315789	1	0,033	3%
3	23,0566802	5	0,167	17%
4	40,8502024	7	0,233	23%
5	58,6437247	6	0,200	20%
6	76,437247	6	0,200	20%
7	y mayor...	5	0,167	17%
8		30	1	100%

5.2 Para obtener los valores de la columna C, divide cada frecuencia en el resultado de la suma de la columna B, por ejemplo, en C2 sería =B2/B8; en la casilla C3=B3/B8. Repite el proceso hasta llegar a la casilla C7.

Opcional: para no tener que repetir el proceso de forma manual en las casillas de la columna C, puedes dejar amarrada la casilla B8 realizando la operación C2=B2/\$B\$8 y luego de ello arrastrar para que Excel autocomplete la operación en las otras casillas desde la C3 hasta la C7. Nota que los signos \$ hacen que la casilla B8 donde se encuentra la suma 30 no se cambie (eso es amarrar o anclar).

5.1 Como se ve en la figura 26, suma todas las frecuencias absolutas; el resultado de la suma es 30, realizando en la casilla B8 la operación: =suma(B2:B7).

La columna D resulta de multiplicar por 100 los valores de la columna C. Otra forma es ubicarse en D2, colocar =C2 y oprimir la tecla **Enter**. Luego, cambia el formato de la casilla a % y arrastra hacia abajo.

Interpretación: hay 7 municipios del Chocó cuyo porcentaje de NBI se encuentra entre 23,057 y 40,850 por ciento; es decir que 23,33% (7/30= 0,233 ≈ 23,33%) de los municipios del Chocó indican que entre 23,057% y 49,850% de su población tienen NBI.

Figura 26. Tabla de frecuencias NBI (%) en Chocó. Fuente: DANE (2018)

Cambio de límites en la tabla

La tabla que se muestra en la figura 26 tiene seis filas (siete con los títulos), con los límites de la variable %NBI como son: 5,263; 23,057; 40,850; 58,644; 76437 y mayor, pero puedes configurar una tabla de frecuencias que tenga diferentes límites, lo que implica que puede tener más o menos filas, para ello debes volver a realizar el proceso de construcción de la tabla, como se indica a continuación:

- 1. A partir de la fila 10 columna A de la hoja *Tabla*, construye una tabla como se ve en la figura 27.
- 2. En esa tabla, en la casilla C12 realiza la siguiente operación: (A12 + B12) / 2. El resultado en C12 debe ser 7,5. Luego, arrastra la fórmula hasta la casilla C19. Las fórmulas en cada casilla y los resultados correspondientes deben verse como en la figura 28. La tabla resultante indica que la primera fila corresponde al rango de 5 a 10 (5, 10]; esto implica que el primer grupo estará compuesto por los municipios cuya población con NBI oscila entre el 5% y el 10%, incluyendo el 10%. En el segundo intervalo, se incluyen los municipios con un % NBI que va desde el 10% hasta el 20% (un intervalo abierto a la izquierda, lo que significa que no se incluyen aquellos con un 10%). Este mismo enfoque se aplica a las demás filas.

NBI		
Li	Ls	MC
5	=A13	=(A12+B12)/2
10	=A14	=(A13+B13)/2
20	=A15	=(A14+B14)/2
30	=A16	=(A15+B15)/2
50	=A17	=(A16+B16)/2
60	=A18	=(A17+B17)/2
70	=A19	=(A18+B18)/2
80	100	=(A19+B19)/2

	A	B	C
10	NBI		
11	Li	Ls	MC
12	5	10	7,5
13	10	20	15
14	20	30	25
15	30	50	40
16	50	60	55
17	60	70	65
18	70	80	75
19	80	100	90

Figura 28. Marca de clase en tabla de frecuencia de variable numérica
Fuente: elaboración propia.

	A	B	C	D	E	F
10	NBI					
11	Li	Ls	MC	FA	Fr	F%
12	5	10				
13	10	20				
14	20	30				
15	30	50				
16	50	60				
17	60	70				
18	70	80				
19	80	100				

Figura 27. Esquema para construir tabla de frecuencia de variable numérica
Fuente: elaboración propia.

- 3. Regresa a la hoja NBI-Chocó, en esa hoja repite los pasos 2, 3 y 4 del procedimiento de crear la tabla. Para ello, recuerda la siguiente ruta: menú principal → DATOS → *Análisis de datos* → *Histograma* y todo lo demás, sin hacer clic en *Aceptar* todavía. Antes de hacerlo, en el campo *Rango de clases*, haz clic en la flecha hacia



Figura 29. Configuración tabla de frecuencia de variable numérica con rango de clases.

Fuente: elaboración propia.

arriba y cambia a la hoja *Tabla*. Luego, con el cursor, selecciona la columna etiquetada como “LS” incluyendo el título, seguido del valor 10 y llega hasta 100. En el ejemplo, esta información se encuentra en las celdas B11 a B19.

Ahora en el campo *Rango de salida*, haz clic en la flecha hacia arriba y selecciona la celda B21 en la hoja *Tabla* (un poco más abajo de la tabla que copiaste). Luego, vuelve a dar clic en la flecha que ahora apunta hacia abajo. Finalmente, haz clic en *Aceptar*.

4. A la izquierda de la figura 30 se observa la tabla que se genera, copia esos datos de la Frecuencia que inician en 1 y terminan en 5; es decir, las casillas C22 hasta C29. Luego, pega esos datos en la *tabla* que estás construyendo desde la casilla D12 A D19. En la casilla D20 realiza la suma de la columna FA. Este proceso te ha permitido construir la frecuencia absoluta FA.

21		Ls	Frecuencia				
22		10	1	10			
23		20	3	20			
24		30	4	30			
25		50	8	50			
26		60	3	60			
27		70	4	70			
28		80	2	80			
29		100	5	100			
30		y mayor...	0				

	A	B	C	D	
10	NBI				
11	LI	LS	MC	FA	
12	5	10	7,5	1	
13	10	20	15	3	
14	20	30	25	4	
15	30	50	40	8	
16	50	60	55	3	
17	60	70	65	4	
18	70	80	75	2	
19	80	100	90	5	
20	=SUMA(D12:D19)				

Figura 30. Construcción de la frecuencia absoluta “Fa” de las NBI (%) con límites ajustados.

Fuente: elaboración propia.

5. Abre tu hoja de cálculo en Excel y dirígete a la celda E12. En esta celda, escribe la fórmula $D12/\$D\20 como se muestra en la figura 31. Luego, presiona la tecla *Enter* para confirmar la fórmula en la celda E12. Ahora, podrás observar el resultado de la división de la celda D12 entre el valor constante ubicado en la celda D20. Notarás que los signos de pesos (\$) en la fórmula indican que se

ha anclado la referencia a la celda D20, lo que significa que su valor no cambiará si copias la fórmula a otras celdas.

Coloca el cursor en la esquina inferior derecha de la celda E12. El cursor se transformará en una cruz pequeña. Haz clic y arrastra hacia abajo hasta la celda E19. A medida que arrastras, Excel aplicará automáticamente la fórmula en cada celda del rango E12:E19, manteniendo el divisor constante en la celda D20. Suelta el clic del ratón para aplicar la fórmula en las celdas E13 a E19. Ahora, cada celda E13, E14, ..., E19 tendrá la misma fórmula, pero ajustando automáticamente el numerador según la fila correspondiente.

- 6. Como se muestra en la figura 32, en la casilla F12 escribe =E12*100, confirma con *Enter*. Luego arrastra la fórmula hasta la casilla F19. (Solo usa 3 decimales, disminuyendo en el menú Inicio en el bloque de formato).

	A	B	C	D	Otra forma	
10	NBI					
11	Li	Ls	MC	FA	Fr	F%
12	5	10	7,5	1	0,033	=E12*100
13	10	20	15	3	0,100	
14	20	30	25	4	0,133	
15	30	50	40	8	0,267	
16	50	60	55	3	0,100	
17	60	70	65	4	0,133	
18	70	80	75	2	0,067	
19	80	100	90	5	0,167	

Nota: otra forma de obtener el % es ubicarse en la casilla F12 =E12 y colocarle el formato %. Luego, arrastrar la fórmula.

Figura 32. Construcción de la frecuencia porcentual “F%” de las NBI (%) con límites ajustados

Fuente: elaboración propia.

	A	B	C	D	E
10	NBI				
11	Li	Ls	FA	Fr	F%
12	5	10	1	0,03	3,33
13	10	20	3	0,10	10,00
14	20	30	4	0,13	13,33
15	30	50	8	0,27	26,67
16	50	60	3	0,10	10,00
17	60	70	4	0,13	13,33
18	70	80	2	0,07	6,67
19	80	100	5	0,17	16,67
20			30	1	100

Figura 33. Resultado final tabla de frecuencias NBI (%) con límites ajustados. Chocó

Fuente: elaboración a partir de Dane (2018).

	A	B	C	D	E
10	NBI				
11	Li	Ls	MC	FA	Fr
12	5	10	7,5	1	=D12/\$D\$20
13	10	20	15	3	
14	20	30	25	4	
15	30	50	40	8	
16	50	60	55	3	
17	60	70	65	4	
18	70	80	75	2	
19	80	100	90	5	
20				30	

Figura 31. Construcción de la frecuencia relativa Fr de las NBI (%) con límites ajustados

Fuente: elaboración propia.

Nota: para amarrar la casilla C20 ubique el cursor en medio de la C y el 20 y oprima la tecla F4 (así aparecen los signos pesos; si no le funcionan, insértelos de forma manual).

General	
\$	%
E	F
Fr	F%
0,033	=E12
0,100	

Acabas de crear una tabla de frecuencias para la variable NBI, usando el módulo análisis de datos del Excel, con límites definidos por ti mismo.

De esta tabla se deduce que hay en total 30 municipios, solo en uno de los cuales las NBI son de 5 %, 10 %; además, es común encontrar que las NBI sean entre 30 % y 50 %, como se reporta en 8 municipios. También puede verse que un 16,67 % de los municipios reportan entre 80 % y 100 %, inclusive con NBI.

Adicionalmente, la tabla de frecuencias de la variable numérica tendrá las frecuencias acumuladas: frecuencia absoluta acumulada, frecuencia relativa acumulada y frecuencia porcentual acumulada.

NBI					
Li	Ls	FA	Fr	F%	FA
5	10	1	0.03	3.33	1
10	20	3	0.10	10.00	=F12+C13

Frecuencia absoluta acumulada (FAA)

NBI							
Li	Ls	FA	Fr	F%	FAA	FrA	
5	10	1	0.03	3.33	1	0.03	
10	20	3	0.10	10.00	4	=G12+D13	

Frecuencia relativa acumulada (FrA)

NBI								
Li	Ls	FA	Fr	F%	FAA	FrA	F%A	
5	10	1	0.03	3.33	1	0.03	3.33	
10	20	3	0.10	10.00	4	0.13	=H12+E13	
20	30	4	0.13	13.33	8	0.27	26.67	
30	50	8	0.27	26.67	16	0.53	53.33	
50	60	3	0.10	10.00	19	0.63	63.33	
60	70	4	0.13	13.33	23	0.77	76.67	
70	80	2	0.07	6.67	25	0.83	83.33	
80	100	5	0.17	16.67	30	1.00	100.00	
		30	1	100				

Frecuencia % acumulada (F%A)

Figura 34. Tabla de frecuencias NBI (%) completa con límites ajustados
Fuente: elaboración propia.

Representar datos

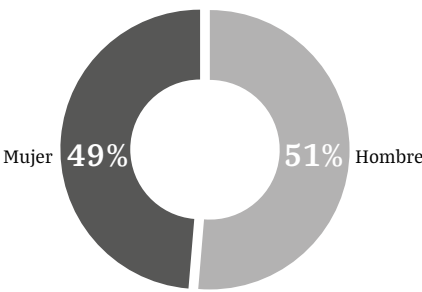
Según el tipo de dato –bien sea numérico o categórico–, se puede representar gráficamente mediante las herramientas que muestra la tabla 12.

Tabla 12. Gráficos sugeridos por tipo de variable

Datos categóricos	Datos numéricos	
	Discretos	Continuos
Diagramas circulares		Histogramas
Diagramas de barra	Diagramas de barra	
Pictogramas		Diagramas de línea

Fuente: elaboración propia.

Diagrama circular. Representación en forma de torta, donde cada sector circular representa una categoría de la población. Un gráfico circular es útil para comparar proporciones. Por ejemplo, se puede utilizar un gráfico circular para demostrar que una mayor proporción de mujeres se ha inscrito en una determinada clase.



Nota: género de población desplazada (diagrama circular)

Figura 35. Ejemplo de diagrama circular o tortas

Fuente: elaboración propia.

Diagrama de barra. Se emplean barras verticales u horizontales para representar los datos. La altura de una barra corresponde a la frecuencia absoluta (o relativa) de la cantidad de datos que hay en una clase; por lo que la barra más alta representa la frecuencia más alta de datos, y la barra más baja representa la frecuencia más baja de datos.

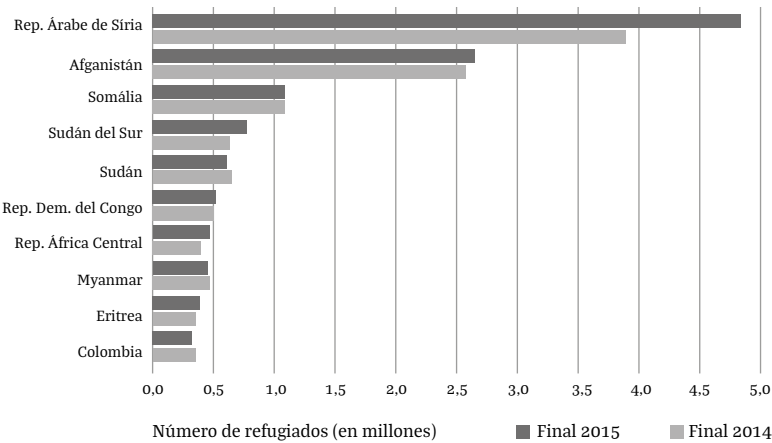


Figura 36. Ejemplo de diagrama de barra para datos categóricos

Fuente: Acnur (2015).

Nota: número de hijos Población desplazada

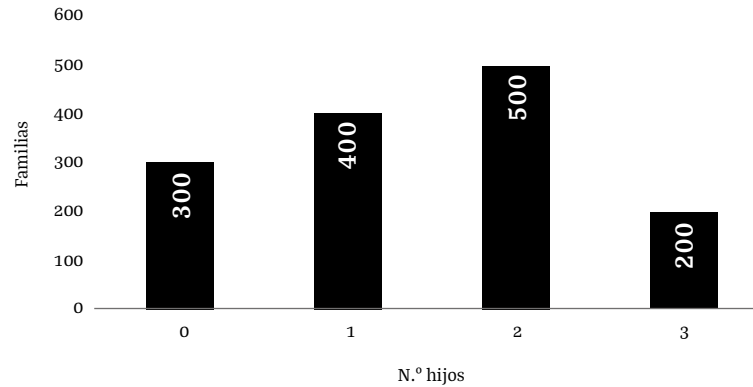


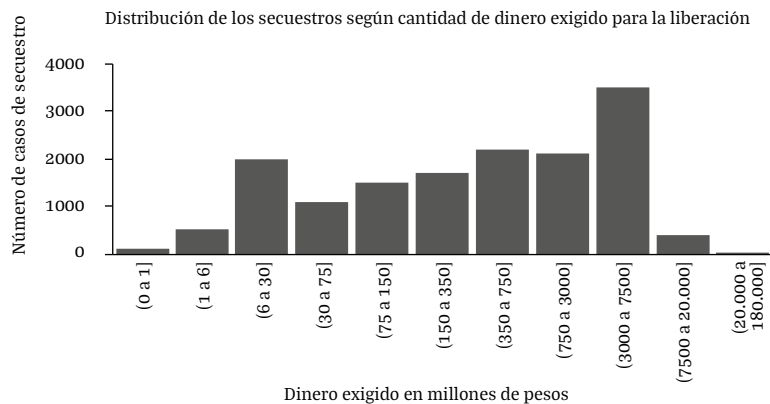
Figura 37. Ejemplo de diagrama de barra para datos numéricos

Fuente: elaboración propia. Sierra (2022)

Pictograma. Un pictograma es un tipo de gráfico cuya información se representa a través de dibujos.

Histogramas. Es un tipo de gráfico que cuenta con amplias aplicaciones en la estadística. Los histogramas permiten una interpretación visual de datos numéricos al indicar el número de puntos de datos que permanecen dentro del rango de valores, denominado *clase* o *compartimento*. La frecuencia de datos que entra en cada clase es representada mediante el uso de una barra.

Son similares a los diagramas de barra, pero tienen una diferencia sustancial. La razón de que estos tipos de gráficos sean diferentes tiene que ver con el nivel de medidas de los datos. Por una parte, los gráficos de barras son utilizados para los datos que están a un nivel nominal de medidas. Estos miden la frecuencia de los datos categóricos, y las clases para un gráfico de barras son estas categorías. Por otra parte, los histogramas son utilizados por datos que, en último lugar, cuentan con un nivel ordinal de medida. Las clases para un histograma son amplios rangos de valores.



Nota: en el eje horizontal se ubica la variable numérica a representar. El paréntesis indica que intervalo de valores es abierto por la izquierda, mientras que el paréntesis angular indica que es cerrado por la derecha.

Figura 38. Ejemplo de histograma de barra para datos numéricos continuos

Fuente: elaboración propia. Sierra (2022)

Diagramas de línea. Muestran los datos en forma de puntos, y todos los puntos de la misma serie se unen mediante una línea; de ahí su nombre. Cada valor aparece representado por un punto, que es la intersección entre los datos del eje horizontal y los del eje vertical (ITE, s. f.).

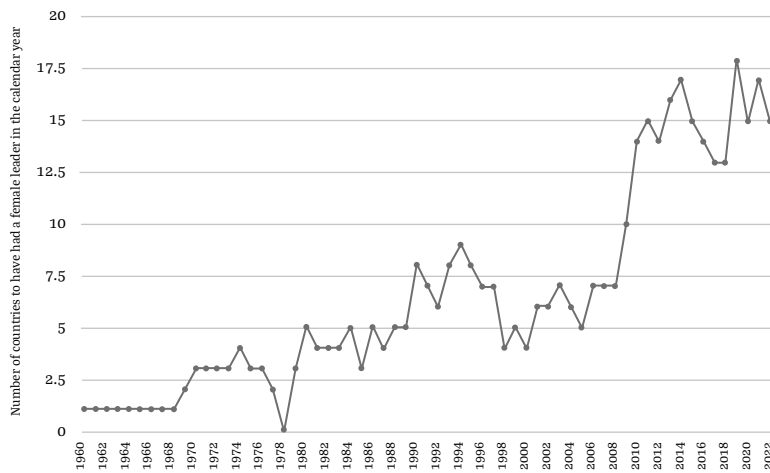


Figura 39. Ejemplo de diagrama de línea de número de países con primera posición del poder ejecutivo a cargo de mujeres (1960-2022).

Fuente: O’Neill (2022).

A propósito de la figura 39, en un análisis alrededor del mundo, desde 1960 hasta 2022, la primera posición del poder ejecutivo estuvo a cargo de una mujer en solo 59 países, siendo el 2019 el año que presentó el número más alto de mujeres líderes (19 en total). La primera ministra elegida democráticamente fue Sirimavo Bandaranaike, de Sri Lanka, quien asumió el liderazgo del Partido de la Libertad de Sri Lanka cuando su esposo fue asesinado, en 1959. Bandaranaike llevó con éxito a su partido a la victoria en tres elecciones: 1960, 1970 y 1994. Sin embargo, los cambios constitucionales en el decenio de 1980 significaron su último mandato como primera ministra. Sri Lanka es uno de los catorce países que ha tenido más de una mujer en el puesto más alto del poder ejecutivo, y la mayoría de estos países se encuentran en el subcontinente indio o en el norte de Europa. De estos catorce, Nueva Zelanda y Finlandia son los únicos países que han tenido tres mujeres líderes, incluidas las primeras ministras, Jacinda Ardern y Sanna Marin (O'Neill, 2022).

Gráficos de Excel. Un gráfico de Excel está formado por diferentes partes, que incluyen el área del gráfico, las series de datos, ejes, leyendas, rótulos del eje, entre otros, como muestra la figura 40.

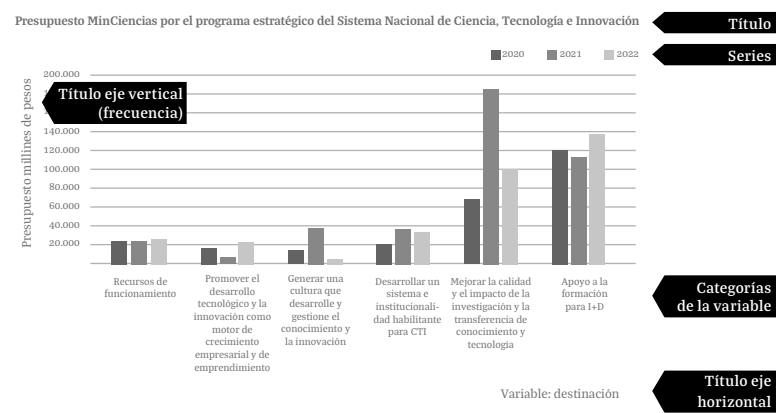


Figura 40. Elementos de un gráfico

Fuente: elaboración propia.

- Título del gráfico: corresponde a un texto que enuncia el contenido del gráfico
- Título del eje vertical (frecuencias): nombre de las frecuencias o unidades en las que se mide la variable
- Título del eje horizontal: nombre de la variable

Serie de videos y recursos

Podrás encontrar una serie de videos donde se explica paso a paso cómo realizar tablas de frecuencias y gráficos usando Excel. Así mismo, en los recursos de aprendizaje encontrarás los enlaces para la descarga de *software* estadístico.



Tablas de frecuencia “variables categóricas”

El Excel se convierte en un amigo indispensable, ya que hace parte de Office 365 para el cual se reporta solo en EE. UU. 731 663 empresas usuarias de la *suite* (Statista, 2021). La estadística descriptiva y la creación de tablas de frecuencia son de fácil desarrollo a través de esta herramienta tan popular. (Clic sobre el icono de la película)



Gráficos “variables categóricas”

¡Una imagen dice más que mil palabras! Por eso usar el recurso gráfico adecuado para representar nuestros datos puede facilitar la comprensión de la información. Se deberá optar por el gráfico que se ajuste el tipo de variable a representar. Los denominados gráficos de torta (o diagramas circulares) son ideales para representar variables categóricas. Y si bien los diagramas de barra también se usan para este tipo de variables, son ideales para aquellas variables con más de tres categorías, ya que una torta con muchas categorías es difícil de entender, al punto de parecer una incomprensible rueda de bicicleta.



Tablas de frecuencia “variables numéricas”

Las tablas dinámicas se convierten en un proceso sencillo para obtener tablas de frecuencia que permitan ver la distribución de los datos. Pero pierden su utilidad si de datos continuos se trata, más aún si son tantos los valores que toma nuestro conjunto de datos, limitando la obtención de tablas de información más resumidas (que es una de las finalidades de la estadística descriptiva), y para ello la opción de histograma en el módulo de análisis de datos ¡es perfecta!



Recursos de aprendizaje

Autor, año y título del recurso	Ubicación	Observaciones
Sierra, J.J. (2021). Clasificación de datos en el contexto agropecuario.	https://es.educaplay.com/juego/2961508-clasificacion_de_datos_en_el_contexto_agropecuario.html	Este es un recurso interactivo, tipo video <i>quiz</i> , con el cual el estudiante podrá poner a prueba sus conocimientos respecto a clasificación de variables, a través de preguntas guiadas con información que irá observando en un video.

Autor, año y título del recurso	Ubicación	Observaciones
Universidad Nacional de Córdoba. FCA-UNC. (30 de julio del 2021). Infostat. Software libre.	http://www.infostat.com.ar/	<i>Software</i> libre cuya licencia es proporcionada por la Universidad Nacional de Córdoba. FCA-UNC (Argentina) (2021). El usuario solo debe ingresar al sitio oficial del <i>software</i> , diligenciar un pequeño formulario y a su correo será enviada la clave de activación del producto, la cual le permitirá usar la versión estudiantil durante un periodo de tres (3) años sin costo alguno.
Rodríguez, O. A. [Oscar Alonso Rodríguez Gracias] (25 de octubre del 2021). <i>Instalación de Infostat versión estudiantil</i> . [Video]. YouTube. https://www.youtube.com/watch?v=-vQUtLdoS8o	https://www.youtube.com/watch?v=-vQUtLdoS8o	Video tutorial en el cual se indica cómo instalar el <i>software</i> estadístico Infostat. Duración: 8:28 min
Sierra, J. J. [Jeammy Sierra] (8 de abril del 2019). <i>Tablas de frecuencia para datos categóricos usando Excel</i> [Video]. YouTube. Ibagué, Tolima, Colombia. https://youtu.be/x4Yg-VrdFXA	https://youtu.be/x4Yg-VrdFXA	Video realizado por la autora en el que se explica cómo realizar tablas de frecuencia para datos categóricos usando Excel. Duración: 5:30 min (Sierra, 2019)
Sierra, J. J. [Jeammy Sierra] (8 de abril del 2019). <i>Tabla de frecuencias para datos numéricos usando Análisis de datos de Excel</i> [Video]. YouTube. Ibagué, Tolima, Colombia. https://youtu.be/XsPk31CKYjg	https://youtu.be/XsPk31CKYjg	Video realizado por la autora en el que se explica cómo realizar tablas de frecuencia para datos numéricos usando Excel. Duración: 9:15 min (Sierra, 2019)
Sierra, J. J. [Jeammy Sierra] (8 de abril del 2019). <i>Gráficos circulares (tortas) y barras</i> [Video]. YouTube. Ibagué, Tolima, Colombia. https://youtu.be/04vLYgQoyoA	https://youtu.be/04vLYgQoyoA	Video realizado por la autora en el que se explica cómo realizar gráficos para datos categóricos usando Excel. Duración: 6:06 min (Sierra, 2019)
Sierra, J. J. [Jeammy Sierra] (11 de febrero del 2019). <i>Medidas de Resumen con Función Análisis de datos y tabla dinámica en Excel</i> [Video]. YouTube. Ibagué, Tolima, Colombia. https://www.youtube.com/watch?v=TjMGSfPPEZM	https://www.youtube.com/watch?v=TjMGSfPPEZM	Video realizado por la autora en el que se explica cómo obtener las medidas de resumen (media, mediana, moda, varianza, desviación, etc.) usando el módulo de Excel denominado Análisis de Datos. Duración: 5:35 min (Sierra, 2019)

Autor, año y título del recurso	Ubicación	Observaciones
Ministerio de Tecnologías de la Información y las Comunicaciones- MinTic. (10 de febrero del 2021). Open Data Colombia. Datos abiertos: https://www.datos.gov.co/	https://www.datos.gov.co/	Este es un sitio web desarrollado por el Ministerio de Tecnologías de la Información y las Comunicaciones -MinTic en el cual se puede acceder a datos abiertos de diferentes ámbitos de interés para el administrador público; para poder aprovechar al máximo este recurso se debe registrar en el sitio. En el menú principal, en la opción “Visualiza”, podrá consultar y descargar bases de datos; estas se encuentran clasificadas por categorías (entre otras, Agricultura y Desarrollo, Vivienda, Ciudad y Territorio)
Asociación Nacional de Movilidad Sostenible- Andemos (2021). Informe interactivo sector automotor	https://datastudio.google.com/u/o/reporting/ceb8deeb-3b00-4e08-8536-5a0f2ebb5cf2/page/PPgdB	Sitio de datos abiertos sobre el sector automotor en Colombia (datos de matrículas de vehículos por departamentos, ciudades, tipos de vehículos, etc.)
Facultad de Economía- Universidad de los Andes (Colombia) (2021). Catálogo de datos -Cede.	https://datoscede.uniandes.edu.co/es/catalogo-de-microdata	Datos Cede recopila, organiza y sistematiza todas las bases de datos disponibles en la Facultad de Economía de la Universidad de los Andes, con el propósito de que los usuarios de la plataforma tengan acceso a la información; solo necesita dar clic sobre el nombre de la base de datos que desea descargar.

Referencias

Almanzor, J. R., Murias, P. G. y Rodríguez, M. T. (1997, octubre). Resumen histórico de la evolución de la estadística. *Estudios de Economía Aplicada*, 8(3), 141. <https://dialnet.unirioja.es/servlet/articulo?codigo=176008>

Alto Comisionado de las Naciones Unidas para los Refugiados [Acnur] (2015). *Tendencias Globales Desplazamiento Forzado en 2015 -Forzados a Huir*. <https://acnur.org/fileadmin/Documentos/Publicaciones/2016/10627.pdf>

American Statistical Association. (04 de Julio del 2021). ASA. *History of Statistics*. <https://www.amstat.org/ASA/History-of-Statistics.aspx>

Azcona, M., Manzini, F. y Dorati, J. (2013). Precisiones metodológicas sobre la unidad de análisis y la unidad de observación. IV Congreso Internacional de Investigación de la Facultad de Psicología. [Memorias]. Universidad Nacional de La Plata. <http://sedici.unlp.edu.ar/handle/10915/45291>

- BBC. (2018). Pansexual, intersexual, 'queer', homosexual, transexual... ¿cuántas formas hay de definir la identidad sexual? <https://www.bbc.com/mundo/noticias-38995644>
- Casas Anguita, J., Repullo Labrador, J. R. y Donado Campos, J. (2003). La encuesta como técnica de investigación. Elaboración de cuestionarios y tratamiento estadístico de los datos (I). *Atención Primaria*, 31(8). [https://doi.org/10.1016/s0212-6567\(03\)70728-8](https://doi.org/10.1016/s0212-6567(03)70728-8)
- Comisión Económica para América Latina [Cepal] (2020, 30 de diciembre). Estadísticas. <https://www.cepal.org/es/fuentes-de-informacion-estadistica>
- Departamento Administrativo Nacional de Estadística [Dane]. (2019, 16 de septiembre). Enfoque diferencial e intersectorial. <https://www.dane.gov.co/index.php/estadisticas-por-tema/enfoque-diferencial-e-interseccional/autorreconocimiento-etnico>
- Departamento Administrativo Nacional de Estadística [Dane] (2021, julio). Glosario Estadístico. *Dane*. <https://www.dane.gov.co/files/Glosario.pdf>
- Departamento Administrativo Nacional de Estadística [Dane] (2021, 24 de octubre). Entidades adscritas al Dane. https://www.dane.gov.co/index.php?option=com_content&view=article&id=780
- Departamento Nacional de Planeación [DNP] (2019). *Medición del desempeño municipal*. <https://www.dnp.gov.co/programas/desarrollo-territorial/Estudios-Territoriales/Indicadores-y-Mediciones/Paginas/desempeno-integral.aspx>
- Dirección de Regulación, Planeación, Estandarización y Normalización (Dirpen). (2020). Plan Nacional Estadístico. Actualización 2020-2022. Departamento Administrativo Nacional de Estadística. <https://www.sen.gov.co/files/PEN/PEN%202020.pdf>
- Ferrer-Sapena, A., Peset, F. y Aleixandre-Benavent, R. (2011). Acceso a los datos públicos y su reutilización: *open data* y *open government*. *El profesional de la información*, 20, 260-269. <https://doi.org/10.3145/epi.2011.may.03>
- Hernández-Sampieri, R. y Mendoza, C. (2018). *Metodología de la investigación. Las rutas cuantitativa, cualitativa y mixta*. Mc Graw Hill Education.

- Instituto Nacional de Salud, Observatorio Nacional de Salud, covid-19 en Colombia. (2020). Covid-19 en Colombia, consecuencias de una pandemia en desarrollo. <https://www.ins.gov.co/Direcciones/ONS/Informes/12COVID-19%20en%20Colombia,%20pandemia%20en%20desarrollo.pdf>
- La Agencia de Renovación del Territorio. (2021, 10 de agosto). PDET. https://www.renovacionterritorio.gov.co/especiales/especial_PDET/#descargables
- Ministerio de Tecnologías de la Información y las Comunicaciones [MinTic] (2021, 10 de febrero). Open Data Colombia. *Datos abiertos*. <https://www.datos.gov.co/>
- Mode, E. (2021). Introducción. En E. Mode, *Elementos de probabilidad y estadística* (p. 2). Reverté, S. A.
- O'Neill, A. (2021, 8 de septiembre). Number of countries where the highest position of executive power was held by a woman, in each year from 1960 to 2021. *Statista*. <https://www-statista.com/statistics/1058345/countries-with-women-highest-position-executive-power-since-1960/>
- Ramos, J., Águila, V. y Bazalar, A. (2020). *Estadística básica para negocios*. Fondo Editorial de la Universidad de Lima. <https://books.google.com.co/books?id=6ZcLEAAQBAJ&printsec=frontcover&dq=como+hacer+tablas+de+frecuencia+libro&hl=es&sa=X&ved=2ahUKEwiog8e48-jzAhUpQzABHT-JMAykQ6AF6BAGJEA#v=onepage&q&f=false>
- Sierra, J. J. (2019, 22 de abril). Tabla de frecuencias para datos numéricos usando Análisis de datos de Excel. Ibagué, Tolima, Colombia. <https://www.youtube.com/watch?v=XsPk31CKYjg>
- Sierra, J. J. (2019, 8 de abril). Gráficos circulares (tortas) y barras. Ibagué, Tolima, Colombia. <https://www.youtube.com/watch?v=04vLYgQoyoA>
- Sierra, J. J. (2019, 11 de febrero). Medidas de resumen con función análisis de datos y tabla dinámica en Excel. Ibagué, Tolima, Colombia. <https://www.youtube.com/watch?v=TjMGSfPPEZM>
- Sierra, J. J. (2019, 8 de abril). Tablas de frecuencia para datos categóricos usando Excel. Ibagué, Tolima, Colombia. <https://youtu.be/x4Yg-VrdFXA>

- Sierra, J. J. (2021, 30 de agosto). Mapa conceptual: “Fases de la estadística”. <https://caco.com/diagrams/BClP1khP3gPOx8E1/36C7E>
- Superintendencia de Servicios Públicos Domiciliarios. (2019). Disposición Final de Residuos Sólidos Informe Nacional 2018. *Superintendencia de Servicios Públicos Domiciliarios*. https://www.superservicios.gov.co/sites/default/archivos/Publicaciones/Publicaciones/2020/Ene/informe_nacional_disposicion_final_2019_1.pdf
- The White House. (2009). Transparency and Open Government. *The WhiteHouse President Barack Obama*. <https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government>
- Universidad Nacional de Córdoba. FCA-UNC. (2021, 30 de julio). *Infostat*. <https://www.infostat.com.ar/>

Bibliografía comentada

- Comisión Económica para América Latina [Cepal] (2020, 30 de diciembre). Estadísticas. <https://www.cepal.org/es/fuentes-de-informacion-estadistica>

Es un recurso efectivo para reconocer el desarrollo de la estadística en otros países de la región. El sitio enlaza a las principales Oficinas Nacionales de Estadística (ONE), con lo cual se tiene un compendio amplio de fuentes de información oficiales. Así mismo, se destaca una lista y acceso abierto a diversos programas y agencias de Naciones Unidas de los países integrantes, entre las que se encuentra el World Bank, donde podrá obtener incluso microdatos en diferentes campos de interés de la sociedad; por ejemplo, en temas de actualidad, como inclusión financiera, hay datos que bien vale la pena auscultar en el análisis de iniciativas legislativas en el contexto colombiano, como las que han causado polémica respecto al uso de saldos de cuentas bancarias inactivas. Estos y otros tantos conjuntos de datos le permiten al lector tener una perspectiva amplia de la información, no solo de la que se produce a nivel nacional, sino más allá de sus fronteras.

Medidas de tendencia central

EN ESTA UNIDAD SE ESTUDIARÁN LAS MEDIDAS DE TENDENCIA CENTRAL, como: media aritmética, mediana, moda y cuantiles. Además, la diversidad de variables a estudiar incita el uso de otros tipos de medidas de tendencia central, en algunos casos poco implementadas, como media armónica, media geométrica y media recortada (o *trimmed mean*), cada una con su particularidad.

Este apartado presenta de forma clara el tratamiento de variables de tipo numérico. Se utiliza como herramienta de apoyo Excel, aplicando diferentes funciones tales como promedio, desviación, rango y cuantiles, entre otras; también se aprovecha el complemento de Excel “análisis de datos”, con el cual los cálculos se tornan sencillos, lo que permite al aprendiz enfocarse en aspectos relevantes, como la interpretación de los resultados obtenidos.

También, en esta unidad se presenta una serie de ejemplos que te ayudarán a comprender cómo operacionalizar el proceso para calcular las medias de resumen. Así mismo, se dispone de cinco prácticas, todas estas con enfoque en problemáticas sociales y el oficio

del administrador público; es el caso del déficit habitacional en cada departamento; el fenómeno de desempleo; problemáticas de salud pública, como los embarazos en adolescentes; o la asignación de recursos provenientes del Fonpet. Las bases de datos o información necesaria se pueden descargar escaneando los códigos QR al inicio de cada práctica.

Por ejemplo, la práctica 6 “Medidas de tendencia central usando funciones de Excel (Déficit habitacional)” corresponde a la problemática de hogares por déficit habitacional según tipo y componentes (miles hogares / participación %); el déficit cuantitativo identifica a los hogares que habitan en viviendas con deficiencias estructurales y de espacio, esto es, para los cuales es necesario adicionar nuevas viviendas al *stock* total de viviendas del país, de tal forma que exista una relación uno a uno entre la cantidad de viviendas adecuadas y los hogares que requieren alojamiento. La información para esta práctica es obtenida de los microdatos del Dane. Aquí se busca calcular el promedio, desviación, mediana y moda del número de hogares con déficit cuantitativo por departamento; se destaca un déficit habitacional medio departamental de 38 000 hogares aproximadamente. Es de esta manera, precisamente, como se debe aprovechar la estadística para soportar las decisiones en análisis sólidos y rigurosos que permiten optimizar los recursos públicos.

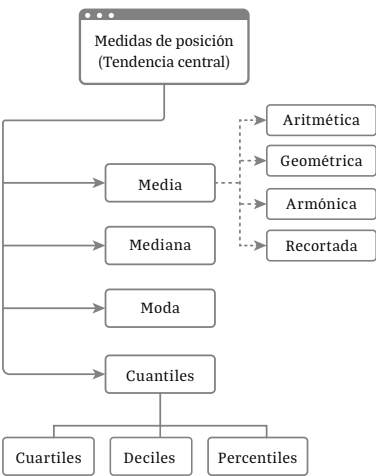
Situación problema

La información es un recurso valioso para la toma de decisiones, y ante un conjunto amplio de datos es habitual tratar de resumirlos, siendo el promedio o media aritmética bastante popular para este fin. Sin embargo, se le ha dado un uso indiscriminado, sin reparar en el tipo de variable a analizar y otros aspectos que determinan qué tan razonables serán los resultados obtenidos a partir de estos cálculos.

Así, algunas situaciones que evidencian esta situación problema se plasman en cuestiones como: ¿Qué tan razonable es promediar

números de cédulas que se encuentran en una base de datos? ¿Qué puede representar el promedio de cédulas? ¡Absolutamente nada! En este sentido, es importante reconocer la naturaleza de los datos para poder utilizar de forma adecuada las herramientas que provee la estadística para su análisis. Otra práctica común en el análisis de datos a través de medidas de tendencia central es utilizar la media aritmética indiscriminadamente, desconociendo lo que representan los datos, pues no basta con aludir a que son números, sino que es necesario reconocer la característica que representa ese número; en muchos casos se trabaja con porcentajes, como con el porcentaje de participación del PIB de los departamentos al PIB nacional, o con tasas de ocupación, entre otros. En estos casos existen otros métodos más respetuosos de las diferencias intrínsecas entre las dimensiones que un promedio aritmético, como la media geométrica (Turchyn, 2021), así como la media armónica y media recortada (o *trimmed mean*), las cuales, lamentablemente, son poco familiares.

Tal desconocimiento es un común denominador que incide en la toma de decisiones, cálculos inadecuados e interpretaciones erradas, no solo en el diagnóstico de problemáticas, sino en el manejo de recursos. El estudiante de administración pública debe tener un conocimiento amplio de este tipo de métodos, pero es convencional limitarse al conocimiento y manejo de algunos estadísticos como el promedio, desconociendo el amplio uso que se hace de otras medidas estadísticas en la administración pública.



Nota: para ver mejor el diagrama, escanear el código QR.

Figura 41. Ejes temáticos de la Unidad 3
Fuente: elaboración propia. Este recurso lo puede encontrar en línea en Sierra (2021).

Lectura de motivación

Promedios en los procesos de contratación pública

En Colombia, en aras de fortalecer la transparencia de los procesos contractuales en la administración pública, la Ley 80 de 1983 contempla el estatuto general para la contratación de la administración pública, estableciendo cuáles entidades estatales deben solicitar más de una oferta en el proceso de adjudicación de contratos y, en clave la Ley 489 de 1988, señala las entidades que no están obligadas a ello. La selección de contratistas se realiza a través de licitaciones, convocatorias o concurso público, o mediante cualquier modalidad contractual, según el régimen jurídico de contratación que sea aplicable, como sugiere la Ley 816 del 2003 (Zambrano *et al.*, 2018).

El factor económico es uno de los criterios de evaluación para la selección de oferentes en procesos de contratación pública. La entidad estatal contratante debe asignar un máximo de puntos de acuerdo con el método seleccionado de ponderación de oferta económica.

Para ello, se selecciona aleatoriamente un método entre los cuatro presentados en la tabla 13. La entidad tomará los centavos de la Tasa de Cambio Representativa del Mercado (TRM), certificada por la Superintendencia Financiera de Colombia (s. f.) para determinar el método de ponderación. Si el proceso se realiza a través del Secop II¹¹, se usará la TRM vigente del día, es decir, la publicada el día anterior. En caso de ser lunes se tomará la publicada el viernes.

Tabla 13. Métodos de ponderación de la oferta económica en el proceso de contratación pública

Rango TRM (inclusive)	Núm.	Método
De 0,00 a 0,24	1	Media con valor absoluto
De 0,25 a 0,49	2	Media geométrica
De 0,50 a 0,74	3	Media aritmética baja
De 0,75 a 0,99	4	Menor valor

Fuente: Colombia compra eficiente.

De lo anterior se tiene que el método con el que se valoran las ofertas desde el factor económico no corresponde siempre a las ofertas con menor valor, es decir, no se puntúa mejor aquel oferente que presente su propuesta con valores financieros por debajo de las demás.

11 La primera versión (SECOP I) dispone cinco métodos de ponderación, la segunda versión (SECOP II) solo establece cuatro: “Los procedimientos de licitación que antes del 10 de marzo de 2020 hayan publicado aviso de convocatoria continuarán rigiéndose por los Documentos Tipo-Versión 1 incluidos en la Resolución 1 798 de 2019”. Una causal de rechazo económico de las propuestas es que la propuesta económica se diligencie de forma incompleta, de tal modo que no permita su verificación aritmética (necesaria para calcular el valor corregido de cada una de las propuestas).

Actividad de presaberes



**¿Qué significa
distribución de datos?**

.....

.....

.....

.....

.....



¿Qué es la media?

.....

.....

.....

.....

.....



**¿Qué es la
mediana?**

.....

.....

.....

.....

.....

Medidas de tendencia central o medidas de posición

En la actualidad, cuando las tecnologías de la información juegan un papel tan importante para su captura y distribución, es cada vez más sencillo obtener grandes conjuntos de datos. La preocupación recae en comprender la estructura de los datos y tratar de resumir información; el problema no subyace en su consecución, sino en su traducción y comprensión de forma sencilla, a través de indicadores o de medidas que resuman esos cúmulos de datos. Es por ello que estadísticas como las medidas de resumen revisten gran valor.

Media (o promedio)

Existen diferentes tipos de medias, media aritmética, media geométrica, media armónica y media cuadrática. A continuación se presentan las formas de calcular cada una de estas medias.

Si X es una variable numérica donde $X=\{ x_1 + x_2 + x_3 + \cdots + x_n \}$, entonces:

Media aritmética:

$$MA_n(X) = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

Media geométrica:

$$MG_n(X) = \sqrt[n]{x_1 + x_2 + x_3 + \dots + x_n}$$

Media armónica:

$$MH_n(X) = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \dots + \frac{1}{x_n}}$$

Media cuadrática:

$$MC_n(X) = \sqrt{\frac{x_1^2 + x_2^2 + x_3^2 + \dots + x_n^2}{n}}$$

Media aritmética (o promedio). Sumatoria de los valores de la variable dividida en el total de datos. Simbólicamente es:

$$Media\ poblacional = \mu = \sum_{i=1}^N \frac{X_i}{N}$$

$$Media\ muestral = \bar{x} = \sum_{i=1}^n \frac{X_i}{n}$$

La diferencia entre la media poblacional y la media muestral consiste básicamente en que la primera se calcula usando todos los N datos de una población, mientras que la segunda se calcula usando los n datos de una muestra.

Ejemplo 3.1

En la tabla 14 se encuentra la demanda total del agua en Colombia desde 2017.

Tabla 14. Demanda de agua en Colombia

Año	Demanda agua Millones m³	$Promedio = \frac{39,17 + 41,13 + 43,19 + 45,35 + 47,62}{5}$ $= \frac{216,46}{5} = 43,29$ En el último quinquenio, desde el 2017, en promedio al año se han demandado en Colombia 43,29 millones de metros cúbicos de agua.
2017	39,17	
2018	41,13	
2019	43,19	
2020	45,35	
2021	47,62	

En Excel:

Copia y pega los datos desde la casilla A1. Ubica en la casilla B7 la sumatoria, así:

	A	B
1	Año	Demanda agua Millones m3
2	2017	39,17
3	2018	41,13
4	2019	43,19
5	2020	45,35
6	2021	47,62
7	Total=	=SUMA(B2:B6)

En Excel:

En la casilla B8 coloca =PROMEDIO y entre paréntesis señala el rango de valores a los que les calcularás el promedio, así:

	A	B
1	Año	Demanda agua Millones m3
2	2017	39,17
3	2018	41,13
4	2019	43,19
5	2020	45,35
6	2021	47,62
7	Total=	216,46
8	Promedio=	=PROMEDIO(B2:B6)

Fuente: estimación a partir del ENA (Ideam, 2018)

Propiedades de la media aritmética:

- i. Sensible a valores extremos. Valores muy pequeños o grandes afectan el resultado del promedio.
- ii. En un conjunto de datos solo hay una media aritmética o promedio.

Media geométrica. Es la raíz enésima (*n*) del producto de los valores de la variable numérica analizada.

Media
geométrica:

$$MG_n(X) = \sqrt[n]{x_1 + x_2 + x_3 + \dots + x_n}$$

Propiedades de la media geométrica:

- iii. Menos sensible a valores extremos que la media aritmética.
- iv. No se puede usar cuando un valor de la variable toma valor cero o negativo.
- v. Es adecuado cuando los datos parecen tomar forma de progresión geométrica.
- vi. Adecuado para calcular el promedio de tasas de crecimiento¹².

12 Tasas de crecimiento como la de contagios del covid-19 o variables que presentan cambios a través del tiempo; por ejemplo, para determinar la tasa promedio de una inversión que genera rendimientos de 8 % el primer año y de 52 % el segundo año, calcule la media geométrica $(1,08 * 1,52)^{1/2} \approx 1,28$ (un rendimiento promedio de 28 %) (Minitab, 2022).

Cuando hay decremento de las variables es posible encontrar tasas que sean negativas; en este caso se utilizarán los factores de crecimiento para calcular el promedio geométrico.

Ejemplo 3.2

Los siguientes son los datos reportados del crecimiento del PIB anual en Colombia desde el 2017 hasta el 2021.

Tabla 15. Variación anual del PIB Colombia, 2017-2021

Año	Tasa de variación del PIB (% anual)-Colombia
2017	1,4
2018	2,6
2019	3,2
2020	-7
2021	10,6

Nota: datos obtenidos de cuentas nacionales anuales del Dane y estadísticas del Banco Mundial (s. f.).

Como se puede observar en la tabla 15, en el 2020 hubo una disminución de 7 % en el PIB; variación significativa en el año de la pandemia ocasionada por la covid-19.

En Excel:
Para calcular el promedio de la variación del PIB, se calcula primero el factor de crecimiento a partir de la fórmula:

$$FC = \frac{Ti}{100}$$

$Ti = \text{Tasa de variación}$

En una hoja nueva de Excel en la casilla C2 debes realizar la operación =1+B2 y luego presionar la tecla *Enter*. Arrastra la fórmula hasta la casilla C6. En la casilla C7 coloca =MEDIA.GEOM y selecciona desde C2 hasta C6. El resultado es 1,020, entonces, la variación promedio es de 0,20 es decir, 20 % aprox.

Figura 42. Media geométrica PIB (Excel)

	A	B	C
	Año	Crecimiento del PIB (% anual) - Colombia	Factor de crecimiento
1			
2	2017	1,4%	=1+B2
3	2018	2,6%	1,026
4	2019	3,2%	1,032
5	2020	-7,0%	0,930
6	2021	10,6%	1,106
7		Promedio	1,020
7		Promedio	=MEDIA.GEOM(C2:C6)

Fuente: elaboración propia.

Caso de estudio

En el municipio de Acacías (Meta) se abre oferta para contratar el mantenimiento del cementerio. El día de la licitación pública es un lunes. El viernes anterior a la apertura de sobres de ofertas la Superintendencia Financiera de Colombia publica que la TRM que registrá es de \$ 4324,27 Además, establecen que el factor económico corresponde máximo a 600 puntos. Se presentan tres oferentes para la licitación: A, B y C.

Las ofertas son respectivamente:

- A= 500 millones
- B= 580 millones
- C= 636 millones

Solo analizando el factor económico, ¿cuál es la oferta que se debe seleccionar como ganadora?

Solución: la TRM es \$ 4324,27 y el valor de los centavos es de 27, es decir, 0,27 pesos, por lo cual el método a usar para la asignación de puntaje será la media geométrica:

$$MG_n(X) = \sqrt[n]{x_1 + x_2 + x_3 + \dots + x_n}$$
$$MG_n(X) = \sqrt[3]{(500) \cdot (580) \cdot (636)}$$
$$MG_n(X) = \sqrt[3]{184\,440\,000}$$
$$MG_n(X) = 569,226$$

La oferta más cercana a la media geométrica será seleccionada, en este caso la B:

A: 500 - 572 = -72

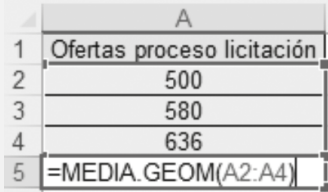
B: 580 - 572 = 8

C: 636 - 572 = 64

Media geométrica en Excel

En una hoja nueva de Excel copia los valores de las tres ofertas en la columna A (desde la casilla A2 hasta A4). Luego, en la casilla A5 coloca =MEDIA.GEOM y selecciona los tres datos. Al dar *Enter* debes obtener el valor: 569, 226.

Figura 43. Media geométrica en Excel (caso licitación)



Fuente: elaboración propia.

Media armónica. Este promedio se define estableciendo que el recíproco de la media armónica es igual a la media aritmética del recíproco de los valores de la variable (Martínez, 2019). Se calcula como el número total de observaciones dividido por la suma de los recíprocos, así:

Media armónica:

$$MH_n(X) = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \dots + \frac{1}{x_n}}$$

Es muy utilizada cuando se quiere promediar variables asociadas a velocidades o tiempos, así como para promediar múltiplos o cocientes.

Propiedades de la media armónica

- i. No se puede usar cuando un valor de la variable toma valor de cero
- ii. Tiene menos sensibilidad o presenta un menor impacto ante números altos, sucediendo lo opuesto con valores pequeños
- iii. Hay una relación entre las tres medias, de forma que la media armónica es siempre inferior a la media geométrica, que a su vez es inferior a la media aritmética.

Ejemplo 3.3

El número de nacidos vivos en Colombia desde el 2018 hasta el 2021, cuya madre se ubica en el grupo de edad de 10 a 14 años, se muestra en la tabla 16.

Tabla 16. Nacidos vivos madres de 10-14 años

Año	Nacidos vivos de madres de 10 a 14 años
2018	5442
2019	4795
2020	4301
2021	4726

Fuente: elaboración propia.

El promedio armónico de dos años consecutivos t y $t+1$ es usado como factor de corrección para la tasa de mortalidad infantil del año $t+1$ (Lombardo, 1981)

En Excel:

	A	B	C
	Año	Nacidos vivos de madres de 10 a 14 años	Media armónica de nacidos en dos años consecutivos
1			
2	2018	5442	
3	2019	<code>=MEDIA.ARMO(B2:B3)</code>	
4	2020	4301	4534,585532
5	2021	4726	4503,495292
6			

En una hoja nueva de Excel, desde la casilla A1 hasta B5, copia los datos. Luego, en la casilla C3 debes realizar la operación `=MEDIA.ARMO` selecciona las casillas B2 hasta B3 (nacidos vivos del 2018 y 2019) y luego oprime *Enter*. El resultado será 5098,054. Arrastra la fórmula hasta la casilla C5.

Media recortada. Se conoce también como media truncada. A diferencia del promedio aritmético, que tiene en cuenta todos los valores para su cálculo, la media truncada recorta los valores atípicos. Los valores atípicos pueden afectar la media (especialmente si solo hay uno o dos valores muy grandes), y en este sentido es preferible usar una media recortada (o truncada) cuando se trabaja con un conjunto de datos con valores altos o bajos erráticos, así como para distribuciones extremadamente sesgadas (Benites, 2021).

Se utiliza un valor de porcentaje como referencia para estas medias, ese número sugiere el porcentaje de datos a eliminar. Por ejemplo, una media recortada del 20 % indica que no se deben tomar ni el 20 % más bajo ni el 20 % más alto de los datos. La media se calcula a partir del 60 % restante de los puntos de datos. Veamos el siguiente ejemplo.

Ejemplo 3.4

Encuentra la media recortada del 20 % para los siguientes puntajes de prueba: 60, 81, 83, 91, 99.

No.	Dato	Dato		
1	60,00	Eliminado	n=	5,00
2	81,00	81,00	Recorte	20 %
3	83,00	83,00	Cantidad datos	1,00
4	91,00	91,00		
5	99,00	Eliminado		
Suma	414,00	255,00		
Promedio	82,80	85,00		

Mediana

Es el valor de la variable que parte a la población en dos, de tal forma que por debajo de ese valor se encuentra la mitad de los datos, y por encima de ese valor la otra mitad. El procedimiento para calcular la mediana es:

- 1. Determinar la cantidad de datos (tamaño muestra o población).
- 2. Organizar los datos de menor a mayor; el dato que se ubique en la posición de la mediana.
- 3. Calcular la posición de la mediana usando la fórmula:

$$\text{Posición de la mediana} = \frac{n+1}{2} \text{ (n es la cantidad de datos)}$$

En algunos casos la mediana no es un valor de la serie de datos, esto pasa por ejemplo cuando el número de datos “n” es par.

Ejemplo 3.5

Notas: 5, 4, 2, 3, 4, 5

- 1. Tamaño de la muestra n= 5
- 2. Organizar de menor a mayor: 3, 4, 4, 2, 5, 5

3.
$$\text{Posición de la mediana} = \frac{5+1}{2} = 3 \text{ (tercera posición)}$$

4. La nota 4,2 está en la tercera posición y es la mediana

Es de anotar que la mediana no se deja afectar por datos extremos, como sí es el caso de la media.

Ejemplo 3.6

Notas: 5, 4,2, 3, 4

1. Tamaño de la muestra $n = 4$
2. Organizar de menor a mayor: 3, 4, 4,2, 5
3. $\text{Posición de la mediana} = \frac{4+1}{2} = 2,5$ (La mediana esta entre el segundo y tercer dato)
4. por ello se promedian los dos $(4+4,2) / 2 = 4,1$.
La nota mediana es 4,1.

En los dos ejemplos anteriores la nota mediana deja por debajo al 50% de los datos; dicho de otra forma, el 50% de los datos es inferior a esa nota y el otro 50% superior a la misma.

Moda

Es el valor que más se repite en un conjunto de datos. Es posible que haya dos valores que se repitan más y en la misma cantidad, en ese caso el conjunto será bimodal (dos modas); trimodal si hay tres valores con la misma frecuencia, que a su vez es la más alta; o polimodal si son más de tres; así como es posible que no haya moda.

Ejemplo 3.7

En el conjunto de datos a continuación se repiten dos veces en cada caso los números 7, 9, 34, 74 y 121, por lo cual no hay una única moda, sino que es una distribución polimodal.

Tabla 17. Déficit cuantitativo habitacional Colombia 2021¹³

Déficit cuantitativo	1	7	7	9	9	10	11	11	12	13	14	18	18	23	27	30	32	34	34	35	36	39	45	46	47	49	58	74	74	89	109	121	121
----------------------	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----

Fuente: elaboración propia.

Cuantiles

Un cuantil es aquel punto que divide la función de distribución de una variable en intervalos regulares. Los cuantiles son valores de la variable, es decir, si la variable es ingresos (en pesos), un cuantil será un valor de ingreso. Los cuantiles se clasifican en: cuartiles, deciles y percentiles.

Cuartiles. Dividen el conjunto de datos en 4 partes iguales, y hay 3 cuartiles, denominados cuartil 1 (Q1), cuartil 2 (Q2) y cuartil 3 (Q3). En la figura 44 se observa que el Q1 deja por debajo un 25 % de los datos, el Q2 un 50 % y el Q3 un 75 %.

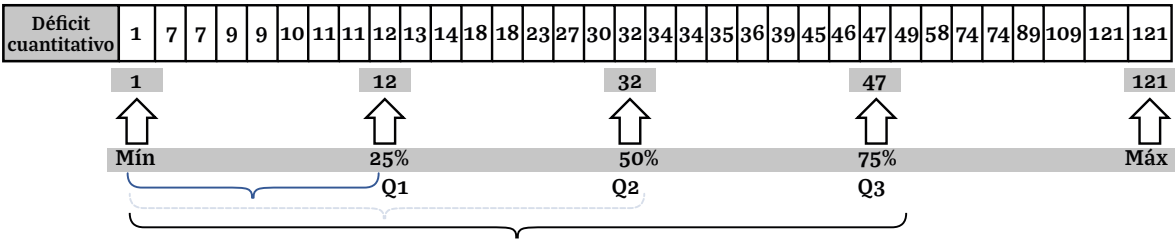


Figura 44. Distribución de datos en cuantiles

Fuente: elaboración propia.

Deciles. Dividen el conjunto de datos en 10 partes iguales, y hay 9 deciles: D1, D2, hasta D9. En la figura 45 se observa que el D1 deja por

13 La tabla corresponde a hogares por déficit habitacional según tipo y componentes (miles /participación %). Total nacional, departamentos y área. El déficit cuantitativo identifica a los hogares que habitan en viviendas con deficiencias estructurales y de espacio, esto es, para los cuales es necesario adicionar nuevas viviendas al *stock* total de viviendas del país de tal forma que exista una relación uno a uno entre la cantidad de viviendas adecuadas y los hogares que requieren alojamiento. Este concepto reconoce que las deficiencias estructurales y de espacio de las viviendas en las que habitan estos hogares no son susceptibles de ser mejoradas para superar la condición de déficit. Puede consultar los microdatos en Dane (2021b).

debajo un 10 % de los datos, el D2 un 20 %, el D3 un 30 % y así sucesivamente hasta el D9, que deja por debajo al 90 % de los valores analizados.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)	(29)	(30)	(31)	(32)	(33)
Déficit cuantitativo	1	7	7	9	9	10	11	11	12	13	14	18	18	23	27	30	32	34	34	35	36	39	45	46	47	49	58	74	74	89	109	121	121
	10%			20%			30%			40%			50%			60%			70%			80%			90%								
Posición	3.40			6.80			10.20			13.60			17.00			20.40			23.80			27.20			30.60								
Vlr déficit decil	7.8			10.8			13.2			21.0			32.0			35.4			45.8			61.2			101.0								

Figura 45. Distribución de datos en deciles
Fuente: elaboración propia.

Decil 1 (D1): El primer paso es calcular la posición (o rango) del decil primero:

$$\text{Posición del decil 1} = \frac{(n+1)+P}{100} = \frac{(33+1)\times 0,01}{100} = 3,4$$

Dado que la posición encontrada no es un número entero, es necesario utilizar el método de interpolación. El decil primero se encuentra entre los valores en las posiciones 3 y 4. Esos valores, basados en los datos organizados en orden ascendente, son 7 y 9. El valor de 3,4-3 = 0,4 corresponde a la proporción de la distancia entre 7 y 9, donde se encuentra el cuartil que se busca:

$$D_1 = 7 + 0,4 \times (9-7) = 7,8$$

Esto completa el cálculo y se concluye que el decil primero es igual a D₁ = 7,8. Igualmente, se puede calcular el D2, D3, etc. A continuación se presentará el cálculo del D2 y se deja al lector que calcule los demás deciles.

Decil 2 (D2): El primer paso es calcular la posición (o rango) del segundo decil:

$$\text{Posición del decil 2} = \frac{(n+1)+P}{100} = \frac{(33+1)\times 0,02}{100} = 6,8$$

Dado que la posición encontrada no es un número entero, es necesario utilizar el método de interpolación. El decil segundo se encuentra entre los valores en las posiciones 6 y 7. Esos valores, basados en los datos organizados en orden ascendente, son 10 y 11. El valor de 6,8-6 = 0,8 corresponde a la proporción de la distancia entre 10 y 11, donde se encuentra el cuartil que se busca:

$$D_2 = 10 + 0,8 \times (11-10) = 10,8$$

Esto completa el cálculo y se concluye que el decil primero es igual a $D_2=10,8$

Percentiles. Dividen el conjunto de datos en 100 partes iguales, y hay 99 percentiles.

Se deja al lector la práctica con la función de Excel denominada: PERCENTIL.EXC, para calcular los percentiles de conjuntos de datos, los percentiles se identifican de la siguiente manera:

Tabla 18. Función percentil en Excel

Percentil	Ingreso en Excel ¹⁴
10	0,10
33	0,33
98	0,98

Fuente: elaboración propia.

Práctica 6: medidas de tendencia central usando funciones de Excel (déficit habitacional)

Las siguientes funciones de Excel permiten calcular las medidas de tendencia central vistas hasta ahora. Para ello, usa los datos de la tabla 17. En una hoja nueva de Excel ubica la tabla de datos desde la casilla A1:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH
1	Déficit cuantitativo	1	7	7	9	9	10	11	11	12	13	14	18	18	23	27	30	32	34	34	35	36	39	45	46	47	49	58	74	74	89	109	121	121

Medida de tendencia central (o posición)	Función de Excel	Resultado
Media	=PROMEDIO(B1:AH1)	38,27
Mediana	=MEDIANA(B1:AH1)	32
Moda	=MODA.VARIOS(B1:AH1)	7
		9
		11
		18
		34
		74
		121
Cuantiles	=PERCENTIL.EXC(B1:AH1;0,25)	12
Deciles	=PERCENTIL.EXC(B1:AH1;0,2)	10,8
Percentiles	=PERCENTIL.EXC(B1:AH1;0,60)	35,4

Figura 46. Medidas de tendencia central (o posición) en Excel

Fuente: elaboración propia.

14 El separador decimal puede variar según esté configurado el paquete de Office para Excel; en algunos casos el separador es punto (.) y en otros es coma (,).

Práctica 7: medidas de resumen con el módulo de análisis de datos de Excel (ECV)

Fuerza de trabajo (Dane)

La fuerza de trabajo es uno de los componentes analizados dentro de la Encuesta de Calidad de Vida (ECV). Esta es una operación estadística que el Dane realiza con el objeto de recoger información sobre diferentes aspectos y dimensiones del bienestar y las condiciones de vida de los hogares, incluyendo temas como el acceso a bienes y servicios públicos, privados o comunales, salud, educación, atención integral a niños menores de 5 años, entre otros. La consideración de estos aspectos hace posible realizar posteriores análisis a los factores que explican los diferentes niveles de vida existentes en la sociedad. La ECV es, además, la fuente de información del cálculo del Índice de Pobreza Multidimensional (IPM). El universo para la Encuesta Nacional de Calidad de Vida está conformado por la población civil no institucional residente en todo el territorio nacional (Dane, 2022). En la descripción de la operación estadística que se encuentra en la página del Dane se establece dentro de las variables del tema de Fuerza de trabajo (personas de 12 años y más): población económicamente activa (PEA), población económicamente inactiva (PEI), ocupados, rama de actividad, posición ocupacional, sitio de trabajo, tamaño de la empresa, tipo de transporte utilizado para desplazarse al trabajo e ingresos, entre otros.

H. FUERZA DE TRABAJO (Para todas las personas de 12 años y más) (continuación)						
17. ¿Qué tipo de contrato tiene, verbal o escrito? Verbal 1 Escrito 2	18. ¿El contrato de trabajo es a término indefinido o a término fijo? Si no sabe el número de meses del contrato, escriba 99; si la duración del contrato es inferior a 15 días, escriba 00; de 15 días a un mes, escriba 01; si la duración del contrato es de 98 meses o más, escriba 98. A término indefinido 1 A término fijo 2 → Número de meses El número de meses se refiere al contrato vigente.		19. ¿Está afiliado/a por una empresa o individualmente a una Administradora de Riesgos profesionales (ARL) (por accidentes de trabajo, enfermedad profesional, etc.)? Si 1 No 2 Si en pregunta H 15 = 4 o 5 pase a la pregunta 32.	20. ¿Cuántos meses lleva trabajando en esta empresa, negocio, industria, oficina, firma o finca de manera continua? Si es menos de un mes, escriba 00.	21. ¿La empresa donde trabaja le suministra elementos de protección personal para la realización de su trabajo? Si 1 No 2 No necesita 3	22. Antes de descuentos, ¿cuánto ganó... el MES PASADO en este empleo? (incluya propinas y comisiones y excluya horas extras, viáticos y pagos en especie.) Si no recibe salario en dinero, escriba 00.
	C.R.	Número de meses		Meses		Valor mensual (\$)
01						

Figura 47. Cuestionario Encuesta Calidad de Vida-Fuerza laboral

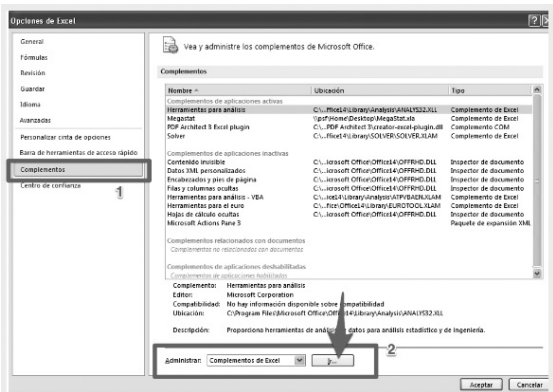
Fuente: Cuestionario Encuesta Nacional de Calidad de Vida (ECV) (Dane, 2022), tópico H. Fuerza de trabajo, preguntas 17 a 22.

1. Descarga el archivo de datos denominado “Fuerza de trabajo”, el cual está comprimido y alojado en:

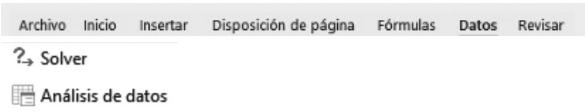


Allí encontrarás tres archivos, uno de estos es un archivo tipo comas (.CSV). Haz clic derecho sobre el nombre del archivo y extrae en la carpeta “Documentos” del computador. Ábrelo y guárdalo como archivo de Excel. En la hoja de cálculo ubica la variable *P6426* (se encuentra en la columna Y); esta corresponde a la pregunta “20. ¿Cuántos meses lleva trabajando en esta empresa, negocio, industria, oficina, firma o finca de manera continua?”. Y ubica la variable *P8624*, esta corresponde al ingreso en pesos \$.

3. En la ventana que se abre, haz clic en *Complementos*. Luego haz clic en “ir”.



5. Ahora ya puedes ir al menú principal → Datos → Análisis de datos



Como ya está activa la opción → Análisis de datos, puedes seleccionar la columna de datos a la cual le quieres calcular el promedio.
Nota: puedes calcular el promedio de varias variables (columnas) a la vez.

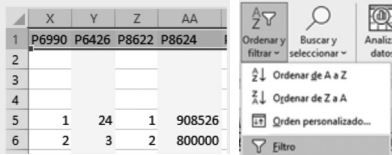
2. *Calcula las medidas de resumen con el módulo análisis de datos de Excel*
En el menú principal en la pestaña Datos ubica la opción análisis de datos.
¿Qué pasa si no encuentro la opción Análisis de datos?
Realiza el siguiente procedimiento (esto se hace solo si no apareció Análisis de datos, de lo contrario salta al paso 2)
En el menú principal de Excel, en la ficha “Archivo” selecciona: *opciones*



4. En la ventana que se abre debes activar con un clic la casilla ☒ *Herramientas para análisis* y luego hacer clic en “Aceptar”.



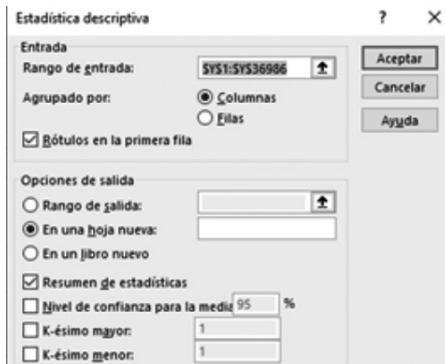
6. Esta base de datos de la encuesta de calidad de vida tiene algunas de las casillas vacías para la variable *P6426*, pues si la persona no está trabajando, no contesta; en total hay 7576 valores (respuestas). Filtra la columna. Para ello selecciona toda la primera fila y en el menú Inicio → Ordenar y Filtrar → Filtro



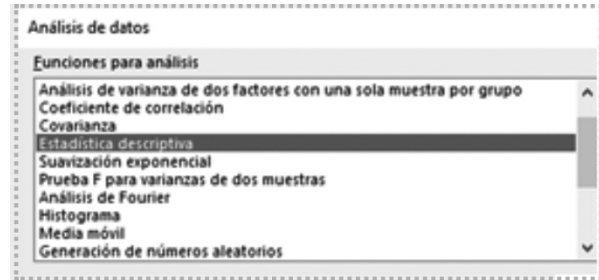
7. En la casilla Y1 despliega la flecha del filtro y deselecciona la casilla de verificación (Vacías); así solo verás las casillas con valores.



9. En *Rango de entrada* selecciona la columna Y, desde el nombre de la variable hasta la última casilla que tiene un valor. Deja marcada la opción *rótulos* y selecciona en una *hoja nueva*. Luego Aceptar.



8. En el menú principal → Datos → Análisis de datos selecciona la opción *Estadística descriptiva* y luego haz clic en Aceptar



10. En una nueva hoja del Excel debe aparecer de forma automática la siguiente tabla:

	A	B
1	P6426	
2		
3	Media	46,6502112
4	Error típico	0,91313035
5	Mediana	12
6	Moda	12
7	Desviación estándar	79,4790672
8	Varianza de la muestra	6316,92212
9	Curtosis	10,9162909
10	Coefficiente de asimetría	3,03459793
11	Rango	648
12	Mínimo	-48
13	Máximo	600
14	Suma	353422
15	Cuenta	7576

¡Listo! Has calculado las medidas de resumen para la variable P6426 que corresponde al número de meses que la persona analizada lleva trabajando en la empresa.

De los resultados obtenidos con la opción análisis de datos se observa que la antigüedad promedio en el trabajo actual de las personas es de aproximadamente 46,650 meses. No obstante, 1 de cada 2 personas solo llevan a lo sumo 12 meses vinculados; el equivalente sería que el 50% de las personas máximo han acumulado hasta 12 meses de trabajo. Además, se evidencia que la desviación con respecto a la antigüedad media es de 79,479 meses. En cuanto a la concentración de los datos (curtosis) se puede inferir que es leptocúrtica, con un valor de curtosis de 10,916, lo cual es mayor a 3.

Se deja al lector el cálculo de las medidas de resumen para la variable P8624.

Medidas de resumen para datos agrupados

Las tablas de datos agrupados contienen la información de forma resumida. A partir de las tablas de frecuencia o simplemente la agrupación de los datos en grupos de intervalos se pueden calcular las medidas de resumen como media aritmética, mediana, moda, desviación estándar. Veamos un ejemplo en la tabla 19.

Tabla 19. Nacimientos según edad de la madre

Grupo de edad de la madre	Total
Total nacional	611 669
De 10-14 años	4726
De 15-19 años	106 695
De 20-24 años	174 078
De 25-29 años	152 946
De 30-34 años	102 870
De 35-39 años	54 424
De 40-44 años	14 533
De 45-49 años	958
De 50-54 años	132

Fuente: Dane (s. f.)

La tabla 19 indica 4726 nacimientos en los cuales la madre tiene entre 10 y 14 años, así mismo se pueden consultar las frecuencias absolutas en cada intervalo de edades. De modo que es una tabla con la estadística consolidada y no el conjunto completo de observaciones; se desconoce cuántos nacimientos corresponden a madres de una edad determinada, solo se puede afirmar que esa cantidad de nacimientos se ubica en un rango de valores. Entonces, sin tener datos específicos de la edad de cada madre, se usa una edad que representa a cada intervalo, como es la marca de clase, la cual se multiplica por la frecuencia absoluta de cada grupo. Luego, la sumatoria de cada uno de estos productos se divide en la cantidad de datos, y este resultado es la “media para datos agrupados” de la edad de las madres adolescentes.

Media para datos agrupados

$$\bar{x} = \frac{\sum_{i=1}^n (x_i^*) \cdot FA_i}{n}$$

$\bar{x}$: Media

x_i^* : Marca de clase

FA_i : Frecuencia absoluta del intervalo i

Mediana para datos agrupados

$$Me = L_i + \frac{\frac{N}{2} - F_{i-1}}{f_i} \cdot a$$

Me : Mediana

L_i : Límite inferior del intervalo de la mediana

a : Amplitud del intervalo de la mediana

F_{i-1} : Frecuencia absoluta acumulada anterior al intervalo de la mediana

f_i : Frecuencia absoluta del intervalo de la mediana

N : Total de datos

El intervalo de la clase mediana es aquel donde la frecuencia absoluta acumulada sea de mínimo el 50 % de los datos.

Moda para datos agrupados

$$Mo = L_i + i \left(\frac{d_1}{d_1 + d_2} \right)$$

Me : Mediana

L_i : Límite inferior de la clase modal

i : intervalo o amplitud

a : Amplitud del intervalo de la mediana

d_1 : Frecuencia absoluta del intervalo modal menos frecuencia absoluta del intervalo anterior

d_2 : Frecuencia absoluta del intervalo modal menos frecuencia absoluta del intervalo siguiente.

El intervalo de la clase modal es aquel donde la frecuencia absoluta sea la máxima.

Práctica 8: medidas de resumen en Excel para datos agrupados (nacimientos)

El Dane, en estadísticas por tema en la sección salud, reporta los nacimientos según grupo de edad de la madre, lo cual es importante dada la problemática de embarazos tempranos no deseados. Puedes descargar y explorar los datos a continuación:

Sección estadística por tema: Salud

Microdatos para descargar en Excel:



1. Descarga la base de datos de los nacimientos según: grupos de edad de la madre; esto para el nivel nacional en Colombia en el 2021.

Tabla 20. Nacimientos según edad de la madre

Grupo de edad de la madre	Total
Total nacional	611 669
De 10-14 años	4726
De 15-19 años	106 695
De 20-24 años	174 078
De 25-29 años	152 946
De 30-34 años	102 870
De 35-39 años	54 424
De 40-44 años	14 533
De 45-49 años	958
De 50-54 años	132
Sin información	307

Nota: no se trabajará el grupo sin información

El total corresponde a la frecuencia absoluta acumulada

2. Ubicándose en una hoja nueva copia la tabla siguiente:

	A	B	C
1	Intervalo	Grupo de edad de la madre	
2	1	10	14
3	2	15	19
4	3	20	24
5	4	25	29
6	5	30	34
7	6	35	39
8	7	40	44
9	8	45	49
10	9	50	54
11			

Se calcula el punto medio o marca de clase de cada intervalo así:

	A	B	C	D
1	Intervalo	Grupo de edad de la madre		MC
2	1	10	14	12
3	2	15	19	$=(B3+C3)/2$
4	3	20	24	22

Se suman las dos casillas de las edades en el intervalo y se divide por 2 $(B3+C3)/2$

	A	B	C	D
1	Intervalo	Grupo de edad de la madre		MC
2	1	10	14	12
3	2	15	19	17
4	3	20	24	22
5	4	25	29	27
6	5	30	34	32
7	6	35	39	37
8	7	40	44	42
9	8	45	49	47
10	9	50	54	52
11				

3. Media: multiplica la marca de clase (mc) por la frecuencia absoluta:

	A	B	C	D	E	F
1	Intervalo	Grupo de edad de la madre		MC	FA	MC*FA
2	1	10	14	12	4.726	=D2*E2
3	2	15	19	17	106.695	1813815

Obtendrás:

	A	B	C	D	E	F
1	Intervalo	Grupo de edad de la madre		MC	FA	MC*FA
2	1	10	14	12	4.726	56712
3	2	15	19	17	106.695	1813815
4	3	20	24	22	174.078	3829716
5	4	25	29	27	152.946	4129542
6	5	30	34	32	102.870	3291840
7	6	35	39	37	54.424	2013688
8	7	40	44	42	14.533	610386
9	8	45	49	47	958	45026
10	9	50	54	52	132	6864
11					611.362	15.797.589
12						
13					Promedio=	25,84

La tabla muestra que la edad promedio de la madre es de 25,84 años.

4. Ahora calcularás la mediana y la moda de los datos agrupados.

Mediana:

	A	B	C	D	E	F
1	Intervalo	Grupo de edad de la madre		MC	FA	FAA
2	1	10	14	12	4.726	4.726
3	2	15	19	17	106.695	111.421
4	3	20	24	22	174.078	285.499
5	4	25	29	27	152.946	438.445
6	5	30	34	32	102.870	541.315
7	6	35	39	37	54.424	595.739
8	7	40	44	42	14.533	610.272
9	8	45	49	47	958	611.230
10	9	50	54	52	132	611.362
11					611.362	
12						
13						
14					Promedio=	25,840
15					Mediana=	23,045

	A	B	C	D	E	F	G
1	Intervalo	Grupo de edad de la madre		MC	FA	FAA	Fr
2	1	10	14	12	4.726	4.726	0,77%
3	2	15	19	17	106.695	111.421	18,23%
4	3	20	24	22	174.078	285.499	46,70%
5	4	25	29	27	152.946	438.445	71,72%
6	5	30	34	32	102.870	541.315	88,54%
7	6	35	39	37	54.424	595.739	97,44%
8	7	40	44	42	14.533	610.272	99,82%
9	8	45	49	47	958	611.230	99,98%
10	9	50	54	52	132	611.362	100,00%
11					611.362		
12							
13							
14					Promedio=	25,840	
15					Mediana=	25,528	
16					Moda=	23,045	

Moda:

	A	B	C	D	E	F	G
1	Intervalo	Grupo de edad de la madre		MC	FA	FAA	Fr
2	1	10	14	12	4.726	4.726	0,77%
3	2	15	19	17	106.695	111.421	18,23%
4	3	20	24	22	174.078	285.499	46,70%
5	4	25	29	27	152.946	438.445	71,72%
6	5	30	34	32	102.870	541.315	88,54%
7	6	35	39	37	54.424	595.739	97,44%
8	7	40	44	42	14.533	610.272	99,82%
9	8	45	49	47	958	611.230	99,98%
10	9	50	54	52	132	611.362	100,00%
11					611.362		
12							
13							
14					Promedio=	25,840	
15					Mediana=	25,528	
16					Moda=	23,045	

Diagrama de caja y bigotes (o *boxplot*)

Un gráfico de cajas y bigotes muestra la distribución de datos en cuartiles, resaltando el promedio y los valores atípicos. Las cajas podrán tener líneas que se extienden verticalmente llamadas *bigotes*. Estas líneas indican la variabilidad fuera de los cuartiles superior e inferior y cualquier punto fuera de esas líneas o bigotes se considera un valor atípico.

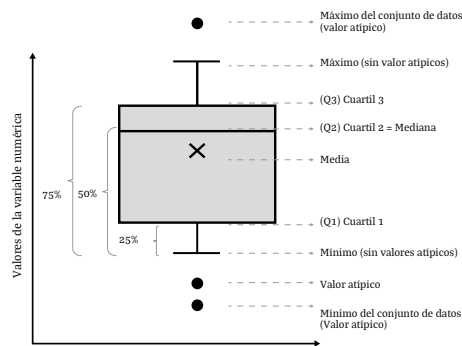


Figura 48. *Boxplot* o diagrama de caja y bigotes
Fuente: elaboración propia.

En la figura 48 se observa que el cuartil 1 deja por debajo al 25 % de los datos, mientras que el cuartil 2 al 50 % y el cuartil 3 al 75 %. Además, los puntos representan datos atípicos denominados *outliers*, los cuales son valores extremos que se comportan diferente al resto de datos.

Ejemplo 3.8

En la figura 49 se presenta el diagrama de *boxplot* correspondiente a los ingresos corrientes tributarios de los 64 municipios de Nariño (Colombia). Se observa que el valor mínimo es aproximadamente 380 millones de pesos; el 25 % de los municipios recaudan hasta 500 millones de pesos; un 50 % de estos municipios, menos de 800 millones de pesos y el 75 % de ellos hasta 880 millones de pesos; además, se identifica que el máximo valor obtenido por tributaciones es de 1000 millones de pesos, y se presenta un valor atípico de 300 millones de pesos por debajo del mínimo esperado.

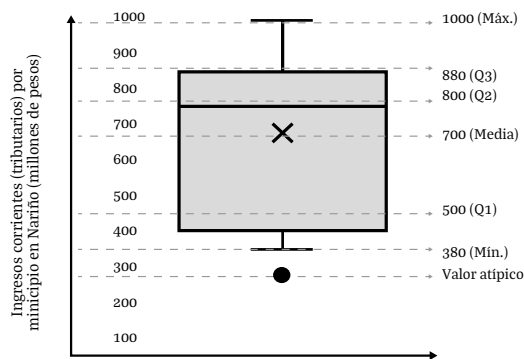


Figura 49. *Boxplot* de ingresos tributarios en los municipios de Nariño (Colombia), 2021
Fuente: elaboración propia.

Ejemplo 3.8

En la figura 50 se presenta el diagrama de *boxplot* correspondiente a los ingresos corrientes tributarios de los departamentos de Nariño y Santander (Colombia). Se evidencia en el ejemplo que Nariño reporta ingresos superiores a los de Santander, pues en este último el 75 % de los datos se encuentra por debajo de 700 millones de pesos, mientras que el primero alcanza hasta los 880 millones aproximadamente.

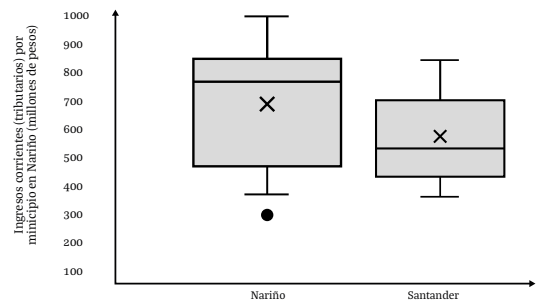


Figura 50. *Boxplot* de ingresos tributarios Nariño y Santander (Colombia), 2021
Fuente: elaboración propia.

Práctica 9: *boxplot* en Excel (recursos Fonpet)
Los recursos girados del Fonpet¹⁵ a los departamentos de Antioquia, Boyacá y Cundinamarca durante el primer semestre del 2022 se muestran en la tabla 21.

Tabla 21. Recursos Fonpet de tres departamentos

Departamento	Valor Giro (billones de pesos)
Antioquia	\$ 157 229,33
Antioquia	\$ 190 107,09
Antioquia	\$ 140 762,47
Antioquia	\$ 192 950,70
Antioquia	\$ 215 710,44
Antioquia	\$ 188 297,73
Boyacá	\$ 71 918,49
Boyacá	\$ 86 878,84
Boyacá	\$ 64 328,38
Boyacá	\$ 91 810,82
Boyacá	\$ 98 579,56
Boyacá	\$ 86 051,97
Cundinamarca	\$ 94 984,29
Cundinamarca	\$ 115 026,12
Cundinamarca	\$ 85 169,69
Cundinamarca	\$ 118 346,89
Cundinamarca	\$ 130 517,67
Cundinamarca	\$ 113 931,35



Puedes descargar la información de cada departamento desde el mapa de inversiones¹⁶. Debes seleccionar la entidad territorial de la cual quieres obtener el reporte; allí debes seleccionar el municipio, vigencia y tipo de recurso que deseas visualizar.

15 Las entidades territoriales reconocían y pagaban las pensiones de sus servidores públicos directamente o a través de Cajas o Entidades de Previsión. Social, pero debido a varios problemas, era insostenible el pago de las obligaciones pensionales a su cargo. La Ley 100 incorporó al nuevo Sistema General de Seguridad Social en Pensiones a estos empleados, quienes pudieron elegir entre el Régimen Solidario de Prima Media con Prestación Definida o el Régimen de Ahorro Individual con Solidaridad. Sin embargo, persistieron los problemas, y se expidió la Ley 549 de 1999 mediante la cual se creó el Fondo Nacional de Pensiones de las Entidades Territoriales (Fonpet) como un fondo sin personería jurídica administrado por el Ministerio de Hacienda y Crédito Público, a través de patrimonios autónomos, constituidos en sociedades fiduciarias, compañías de seguros y fondos de pensiones obligatorias (MinHacienda, s. f.).

16 El mapa de inversiones del Departamento Nacional de Planeación (DNP) es una herramienta con la cual puede identificar y descargar los valores girados a cada una de las entidades territoriales, según diferentes tipos de recursos.

1. Coloca la información de cada uno de los tres departamentos en un archivo nuevo de Excel. Selecciona la columna Departamento y Valor Giro, así:

	A	B
	Departamento	Valor Giro (Billones pesos)
1		
2	Antioquia	\$157,229.33
3	Antioquia	\$190,107.09
4	Antioquia	\$140,762.47
5	Antioquia	\$192,950.70
6	Antioquia	\$215,710.44
7	Antioquia	\$188,297.73
8	Boyacá	\$71,918.49
9	Boyacá	\$86,878.84
10	Boyacá	\$64,328.38
11	Boyacá	\$91,810.82
12	Boyacá	\$98,579.56
13	Boyacá	\$86,051.97
14	Cundinamarca	\$94,984.29
15	Cundinamarca	\$115,026.12
16	Cundinamarca	\$85,169.69
17	Cundinamarca	\$118,346.89
18	Cundinamarca	\$130,517.67
19	Cundinamarca	\$113,931.35

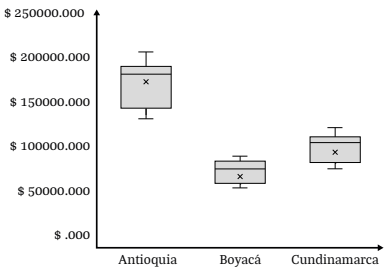
3. En el asistente para gráficos recomendados, ve a la pestaña *Todos los gráficos* y allí selecciona *Cajas y bigotes*



2. Ve al menú principal y en el menú Insertar seleccione el comando *Gráficos recomendados*



4. Finalmente, en aceptar, debe aparecer una gráfica como esta: puedes cambiar el título y agregar las unidades al eje vertical.



Autoevaluación: ¡Practica tus conocimientos!

Práctica 10: media truncada

Usa los datos correspondientes al PIB a precios corrientes base 2015 reportados por el Dane de los departamentos de Colombia para el 2021. Calcula una media truncada del 20 %.

Departamento	PIB (miles de millones)
Amazonas	228
Antioquia	51 007
Arauca	3 555
Atlántico	13 507
Bogotá D. C.	89 832
Bolívar	11 169
Boyacá	8226
Caldas	6015
Caquetá	1254
Casanare	7842
Cauca	5059
Cesar	6462
Chocó	1232
Córdoba	6023
Cundinamarca	19 546
Guainia	127
Guaviare	349
Huila	6076
La Guajira	4199
Magdalena	4323
Meta	7953
Nariño	5169
Norte de Santander	5098
Putumayo	1116
Quindío	2750
Risaralda	5512
San Andrés, Providencia y Santa Catalina (Archipiélago)	489
Santander	18 733
Sucre	2386
Tolima	7702
Valle del Cauca	34 639
Vaupés	80
Vichada	302
Colombia	337 958

Recursos de aprendizaje

Autor, año y título del recurso.	Ubicación	Observaciones
Sierra, J. J. (2019, 11 de febrero). Medidas de Resumen con Función Análisis de datos y tabla dinámica en Excel [Video]. YouTube. Ibagué, Tolima, Colombia. https://www.youtube.com/watch?v=TjMGSfPPEZM	https://www.youtube.com/watch?v=TjMGSfPPEZM	Video realizado por la autora en el que se explica cómo obtener las medidas de resumen (media, mediana, moda, varianza, desviación, etc.) usando el módulo de Excel denominado Análisis de Datos. Duración: 5:35 min. (Sierra, 2019)
León, A. y Pérez, C. (2019). Análisis estadístico en investigaciones positivistas: Medidas de tendencia central. <i>Orbis Revista Científica Electrónica de Ciencias Humanas / Scientific e-journal of Human Sciences</i> , 1(43), 50-60.	http://bdigital2.ula.ve:8080/xmlui/bitstream/handle/654321/3770/art5.pdf?sequence=1&isAllowed=y	Artículo en revista que presenta de forma resumida las medidas de tendencia a utilizar según el nivel de medición de la variable (León y Pérez, 2019)
Unidad de Servicios Penitenciarios y Carcelarios de Colombia. Uspec Colombia. (2020, 21 de diciembre). Audiencia Pública de Adjudicación Licitación Pública Upsec-LP-032	https://www.facebook.com/unidad.uspec/videos/1469991159862106	Video de audiencia de adjudicación de licitación. A partir del minuto 34:20 se puede reconocer los diferentes métodos de evaluación económica entre los que se destaca: • Mediana con valor absoluto • Media geométrica con Presupuesto Oficial • Media aritmética alta • Media aritmética baja • Menor valor
Universidad Nacional de Córdoba. FCA-UNC. (2021, 30 de julio). Infostat. https://www.infostat.com.ar/	http://www.infostat.com.ar/	<i>Software</i> libre cuya licencia es proporcionada por la Universidad Nacional de Córdoba. FCA-UNC (Argentina) (2021). El usuario solo debe ingresar al sitio oficial del <i>software</i> , diligenciar un pequeño formulario y a su correo será enviada la clave de activación del producto, la cual le permitirá usar la versión estudiantil durante un periodo de tres (3) años sin costo alguno.
Turchyn, I. (2021). Weighted Geometric Mean and its Properties. <i>Applications and Applied Mathematics: An International Journal (AAM)</i> , 16(1), Article 3. https://digitalcommons.pvamu.edu/aam/vol16/iss1/3	https://digitalcommons.pvamu.edu/aam/vol16/iss1/3	Este es un artículo en Inglés del <i>Journal AAM (Applications and Applied Mathematics: An International Journal)</i> en donde se exponen las propiedades de un promedio alternativo al promedio aritmético, como es la “Media ponderada”, presentando simulaciones que comprueban que esta media logra un cuadrado medio del error (CME) más pequeño al de otras medidas de tendencia central en ciertas distribuciones como Nakagami y Log-logística. Esta lectura invita al estudiante a implementar otras medidas de tendencia central (Turchyn, 2021).

Departamento Administrativo Nacional de Estadística [Dane]. (2017, 21 de junio). Tercer censo Nacional Agropecuario. Anda; Microdatos del Dane. http://microdatos.dane.gov.co/index.php/catalog/513/get_microdata	http://microdatos.dane.gov.co/index.php/catalog/513/get_microdata	Microdatos de la página del Dane del Tercer Censo Nacional agropecuario realizado en el 2014 en Colombia. Los datos se encuentran divididos en archivos por cada departamento en una carpeta comprimida, dentro de la cual se encuentra el archivo tipo comas extensión CSV. De la misma manera, se encuentran archivos extensión SPSS. Igualmente hay un archivo consolidado de todos los departamentos. Para usar estos datos, el estudiante debe descargar la información de un departamento de Colombia sobre el cual trabajará y presentará un informe estadístico del diagnóstico de ese departamento. Proyecto: “Diagnóstico” en el Plan de Desarrollo Territorial. Dimensión: económica (ver dimensiones Ley 152 y Ley 388). Línea estratégica: por definir Sector: agropecuario Se deja a consulta del estudiante el objeto virtual de aprendizaje “Kit de Planeación Territorial”, con el cual podrá comprender el proceso de diagnóstico en la elaboración del Plan de Desarrollo Territorial con todas sus etapas, ubicado en el portal territorial del DNP (2019).
Instituto Nacional de Medicina Legal. (2021). Observatorio de Violencia. https://www.medicinalegal.gov.co/observatorio	https://www.medicinalegal.gov.co/observatorio	A través del módulo “Cifras estadísticas” encontrará acceso a los siguientes recursos: Forensís <i>Masatugó</i> Boletines Epidemiológicas Boletines Estadísticos Mensuales Alertas Epidemiológicas Forenses Por ejemplo, <i>Masatugó</i> es una publicación quinquenal sobre lesiones de causa externa en mujeres en Colombia. Es un aporte significativo que presenta las formas de violencia contra la mujer, con estadísticas detalladas por municipio, tipologías y demás, que resultan ser un insumo importante en el desarrollo no solo de monográficos en la descripción del problema y antecedentes de proyectos, sino para propuestas sobre equidad de género.

Colombia Compra Eficiente (2021). Documentos tipo para licitación de obra pública de infraestructura de transporte- versión x.	https://www.colombiacompra.gov.co/content/01-documentos-tipo-para-licitacion-de-obra-publica-de-infraestructura-de-transporte	Documento tipo. Es una guía dentro de la página de Colombia Compra Eficiente que orienta sobre el proceso de compra en las diferentes modalidades. Allí se presenta un grupo de cinco documentos tipo, de los cuales sugiere consultar al estudiante: 01.Documentos tipo de licitación de obra pública de infraestructura de transporte Se pide leer de este grupo la última versión. Si es necesario ingresar en la última versión con modificaciones y revisar la vigente en el año en curso con el mes más cercano a la fecha actual. Documento Base o pliegos tipo Licitación.docx Esta guía permite al estudiante reconocer ámbitos de la administración pública en donde puede implementar otras medidas de tendencia central, como la media geométrica; para el caso, en procesos de contratación y licitación.
Unidad de Servicios Penitenciarios y Carcelarios de Colombia. Uspec Colombia. (21 de diciembre del 2020). Audiencia Pública de Adjudicación Licitación Pública Upsec-LP-032	https://www.facebook.com/unidad.uspec/videos/1469991159862106	Este es un video de audiencia de adjudicación de licitación. A partir del minuto 34:20 se pueden reconocer los diferentes métodos de evaluación económica, entre los que se destacan: • Mediana con valor absoluto • Media geométrica con Presupuesto Oficial • Media aritmética alta • Media aritmética baja • Menor Valor

Referencias

- Banco Mundial. (s. f.). Crecimiento del PIB (% anual) Colombia.
<https://datos.bancomundial.org/indicador/NY.GDP.MKTP.KD.ZG?locations=CO>
- Banco Mundial (2020). Esperanza de vida al nacer, total (años).
<https://datos.bancomundial.org/indicador/SP.DYN.LE00.IN?view=map>
- Benites, L. (2021, 31 de diciembre). Media recortada / Media truncada: definición, ejemplos. <https://statologos.com/medio-recortado/>
- Carvajal, R. (2021). Indicios de calentamiento global en temperatura media y mínima para la ciudad de Bogotá, Colombia 1981-2015. *Geográfica Digital*, 17(34), 19-30. <http://dx.doi.org/10.30972/geo.17344421>
- Cuenca, M. E. (2020, 14 de septiembre). *Índice de Desarrollo Humano y distribución del ingreso*. <https://www.uninorte.edu.co/web/deptoeconomia/home/-/blogs/indice-de-desarrollo-humano-y-distribucion-del-ingre-1>
- Departamento Administrativo Nacional de Estadística [Dane] (s. f.). Nacimientos. <https://www.dane.gov.co/index.php/estadisticas-por-tema/salud/nacimientos-y-defunciones/nacimientos>
- Departamento Administrativo Nacional de Estadística [Dane] (2017). *Tercer censo Nacional Agropecuario*. http://microdatos.dane.gov.co/index.php/catalog/513/get_microdata
- Departamento Administrativo Nacional de Estadística [Dane] (2021a). *Comunicado de prensa “Pobreza monetaria”*. https://www.dane.gov.co/files/investigaciones/condiciones_vida/pobreza/2020/Comunicado-pobreza-monetaria_2020.pdf
- Departamento Administrativo Nacional de Estadística [Dane] (2021b). Déficit habitacional. <https://www.dane.gov.co/files/investigaciones/deficit-habitacional/anexo-tec-deficit-hab-2021.xlsx>
- Departamento Administrativo Nacional de Estadística [Dane] (2022). *Colombia-Encuesta Nacional de Calidad de Vida-ecv 2021*. <https://microdatos.dane.gov.co/catalog/734/study-description>

- Departamento Nacional de Planeación [DNP]. (2019). *Kit de Planeación Territorial- Unidad 2: Diagnóstico*. <https://portalterritorial.dnp.gov.co/kpt/>
- Departamento Nacional de Planeación [DNP]. (2021). *Metadatos y Microdatos*. <https://anda.dnp.gov.co/index.php/catalog>
- Instituto de Hidrología, Meteorología y Estudios Ambientales [Ideam] (2018). Estudio Nacional del Agua. http://www.ideam.gov.co/web/agua/estudio-nacional-del-agua/-/document_library_display/hWSQikOLFPrw/view/125678471
- Instituto Vasco de Estadística [Eustat] (2020b). *Índice de Desarrollo Humano por indicadores según países. 2020*. Euskal Estatistika Erakundea, Instituto Vasco de Estadística. https://www.eustat.eus/elementos/ele0013500/ti_indice-de-desarrollo-humano-por-indicadores-segun-paises-2019/tblo013566_c.html
- Gómez, D. (2018). Francisco Porras Sánchez, Gobernanza: Propuestas, límites y perspectivas. *Gestión y Política Pública*, 27(1), 285-287. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1405-10792018000100285&lng=es&tlng=es
- Instituto Nacional de Medicina Legal y Ciencias Forenses / Colombia. (2021). *Masatugó: mujer que recibe lo malo, para entregar lo bueno. 2015-2019*. (G. C. Violencia, Ed). <https://www.medicinalegal.gov.co/cifras-estadisticas/masatugo>
- Instituto Nacional de Salud de Colombia. (2021). *Herramientas de visualización de datos de salud pública*. <https://www.ins.gov.co/Direcciones/ONS/visor-de-indicadores-en-salud-p%C3%BAblica>
- Instituto Nacional de Salud, Observatorio Nacional de Salud, covid-19 en Colombia. (2020). *Covid-19 en Colombia, consecuencias de una pandemia en desarrollo*. <https://www.ins.gov.co/Direcciones/ONS/Informes/12COVID-19%20en%20Colombia,%20pandemia%20en%20desarrollo.pdf>
- León, A. y Pérez, C. (2019). Análisis estadístico en investigaciones positivistas: Medidas de tendencia central. *Orbis Revista Científica Electrónica de Ciencias Humanas / Scientific e-journal of Human Sciences*, 1(43), 50-60.

- Lombardo, E. (1981). ¿Media aritmética o armónica Nella Stima Della Mortalita'infantile? *Genus*, 37(1/2), 159-161. <http://www.jstor.org/stable/29788316>
- Martínez, C. B. (2019). *Estadística básica aplicada*. Ecoe Ediciones.
- Ministerio de Agricultura de Colombia [MinAgricultura] (2021). *Agronet: Estadísticas*. Red de información y comunicación del sector Agropecuario Colombiano. <https://www.agronet.gov.co/Paginas/inicio.aspx>
- Ministerio de Educación Nacional de Colombia [MEN] (2021). *Información y Estadísticas Sectoriales*. Datos abiertos MEN. https://www.mineduacion.gov.co/1759/w3-propertyvalue-57277.html?_noredirect=1
- Ministerio de Hacienda y Crédito Público [MinHacienda] (s. f.). Marco General. https://minhacienda.gov.co/webcenter/portal/FONPET/pages_marcogeneral
- Ministerio de Tecnologías de la Información y las Comunicaciones-MinTic. (10 de febrero del 2021). *Open Data Colombia*. Datos abiertos. <https://www.datos.gov.co/>
- Organización Mundial del Turismo [Unwto]. (2021). *Unwto tourism data dashboard*. Unwto. <https://www.unwto.org/es/unwto-tourism-dashboard>
- Minitab (2022). Función Media geométrica. <https://support.minitab.com/es-mx/minitab/18/help-and-how-to/calculations-data-generation-and-matrices/calculator/calculator-functions/statistics-calculator-functions/geometric-mean-function/>
- Pérez, J. (15 de junio del 2010). Cocientes demográficos: tasas, probabilidades, razones y proporciones. <https://apuntesdedemografia.com/curso-de-demografia/cocientes-demograficos-tasas-probabilidades-razones-y-proporciones/>
- Programa de Naciones Unidas para el Desarrollo [PNUD]. (2020). *Informe sobre Desarrollo Humano 2020: La próxima frontera. El desarrollo humano y el Antropoceno*. Programa de las Naciones Unidas para el Desarrollo-PNUD. http://hdr.undp.org/sites/default/files/hdr_2020_overview_spanish.pdf

- Programa Naciones Unidas para el Desarrollo- PNUD. (2020). *Nota informativa para los países acerca del Informe sobre Desarrollo Humano-Colombia*. Programa Naciones Unidas para el Desarrollo- PNUD. http://hdr.undp.org/sites/all/themes/hdr_theme/country-notes/es/COL.pdf
- Ramos, J., Águila, V. y Bazalar, A. (2020). *Estadística básica para negocios*. Fondo editorial de la Universidad de Lima. <https://books.google.com.co/books?id=6ZcLEAAQBAJ&printsec=frontcover&dq=como+hacer+tablas+de+frecuencia+libro&hl=es&sa=X&ved=2ahUKEwiog8e48-jzAhUpQzABHT-JMAykQ6AF6BAGJEAI#v=onepage&q&f=false>
- Sierra, J. J. (2019, 11 de febrero). Medidas de Resumen con Función Análisis de datos y tabla dinámica en Excel. Ibagué, Tolima, Colombia. <https://www.youtube.com/watch?v=TjMGSfPPEZM>
- Sierra, J. J. (2021, 30 de agosto). Mapa conceptual: “Fases de la estadística”. <https://cacao.com/diagrams/BCIP1khP3gPOx8E1/36C7E>
- Silva, R. (2009). Reflexiones acerca de la Esperanza de Vida y una implementación en una hoja de cálculo. *Punto de encuentro para el gremio actuarial*, 1. <http://www.colegioactuarial.org/Revista%20Actuario/Publicaciones/023.pdf>
- Superintendencia Financera de Colombia (s. f.). Tasa de Cambio Representativa del Mercado- trm. <https://www.superfinanciera.gov.co/inicio/informes-y-cifras/cifras/establecimientos-de-credito/informacion-periodica/diaria/tasa-de-cambio-representativa-del-mercado-trm-60819>
- Turchyn, L. (2021). Weighted Geometric Mean and its Properties. *Applications and Applied Mathematics: An International Journal (AAM)*, 16(1), 43-54. <https://digitalcommons.pvamu.edu/cgi/viewcontent.cgi?article=1120&context=aam>
- United Nations Development Programme [UNDP] (2020a). *Data Center*. Human Development Reports: <http://hdr.undp.org/en/data>
- United Nations Development Programme [UNDP] (2020b). *Human Development Index trends, 1990-2019*. <http://hdr.undp.org/en/composite/trends>

- United Nations Development Programme [UNDP] (2020c). *Latest Human Development Index Ranking*. <http://hdr.undp.org/en/content/latest-human-development-index-ranking>
- United Nations Development Programme [UNDP] (2020d). *Technical notes*. http://hdr.undp.org/sites/default/files/hdr2020_technical_notes.pdf
- Universidad Nacional de Córdoba. FCA-UNC. (2021, junio 30). *Infostat*. <https://www.infostat.com.ar/>
- Zambrano, N.A., Jaramillo, M.A., Pérez, D. D. y Serrano, M. F. (2018). Fórmulas de selección económica de contratistas en adjudicación de obras de infraestructura vial: Estudio de caso Valle del Cauca, Colombia. *Entre ciencia e ingeniería*, 12(24). http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1909-83672018000200060

Bibliografía comentada

- Ministerio de Agricultura de Colombia (2021). Agronet: Estadísticas. <https://www.agronet.gov.co/Paginas/inicio.aspx>

Es la Red de Información y Comunicación del Sector Agropecuario de Colombia, liderada por el Ministerio de Agricultura y Desarrollo Rural y con el apoyo de la Organización de las Naciones Unidas para la Agricultura y la Alimentación (FAO). Centraliza y difunde información del sector para apoyar la toma de decisiones. Por ello, establece sinergias con otras unidades de gestión de información sectorial, nacional y regional, tales como instituciones descentralizadas, entidades gubernamentales, centros de investigación, universidades y proyectos de campo en ejecución a cargo de diferentes organizaciones nacionales o internacionales.

- Instituto Nacional de Medicina Legal y Ciencias Forenses (2021). *Masatugó*. <https://www.medicinalegal.gov.co/cifras-estadisticas/masatugo>

El Instituto Nacional de Salud, Grupo Centro de Referencia Nacional sobre Violencia (CRNV) del Instituto Nacional de Medicina Legal y Ciencias

Forenses a través del Sistema Red de Desaparecidos y Cadáveres (Sirdec), sistema encargado de recopilar la información relacionada con el ejercicio médico legal, reportó que durante el quinquenio comprendido entre 2015 y 2019 se realizaron 5013 necropsias en el contexto de la violencia homicida contra las mujeres, con una tasa promedio por cada 100 000 mujeres de 4,12. Dentro de su misión, el Instituto Nacional de Medicina Legal y Ciencias Forenses no está llamado a la calificación del delito de feminicidio, ya que esto es competencia exclusiva de la Fiscalía General de la Nación; no obstante, brinda herramientas a los operadores de justicia (investigadores, fiscales y jueces) para el esclarecimiento de posibles casos de feminicidio. Esta publicación quinquenal sobre lesiones de causa externa en mujeres es un aporte significativo en el marco de la relación que existe entre el desequilibrio de poder entre los géneros, la discriminación, la violación de los derechos humanos y las libertades fundamentales de las mujeres adultas, niñas y adolescentes, donde se aborda:

- Homicidio en mujeres
- Violencia interpersonal contra mujeres
- Violencia intrafamiliar contra niñas y adolescentes mujeres
- Violencia contra la mujer adulta joven por parte de familiares diferentes a la pareja
- Violencia intrafamiliar contra la mujer adulta mayor
- Lesiones no fatales en violencia de pareja contra la mujer
- Exámenes medicolegales en mujeres por presunto delito sexual
- Suicidios en mujeres
- Comportamiento de la desaparición en mujeres

El reconocimiento de este flagelo permite resignificar los derechos de las mujeres. A través de las políticas públicas territoriales se pueden gestar y enfocar esfuerzos para seguir trabajando en el fomento de la equidad de género, no solo desde la prevención de hechos tan lamentables como la violencia contra la mujer, sino en el propósito de consolidarla en el desarrollo de la sociedad a través de su posicionamiento en el ámbito público.

Se invita al estudiante a complementar esta lectura con la serie documental *Habla por todas*, dirigida por Carmen Figueroa y Lina Pérez.

Son 24 capítulos grabados en diferentes municipios antioqueños, en los que, con varias invitadas y una experta principal por capítulo, se abordan temas como:

- La educación. https://www.youtube.com/watch?v=uLdv_zJuhOU
- Estereotipos. <https://www.youtube.com/watch?v=t4Jf7qtvW6A>
- Otros. <https://play.teleantioquia.co/habla-por-todas/>

Variación y forma

EN ESTE APARTADO SE EXPONE OTRO DE LOS GRUPOS DE MEDIDAS QUE se pueden obtener a partir de las variables de tipo numérico, como rango, desviación estándar, desviación media, varianza y coeficiente de variación. Este tipo de mediciones son bastante pertinentes a la hora de analizar datos, al aceptar la tesis de que el mundo que nos rodea no es estático y por el contrario sufre continuas metamorfosis. En este sentido, los cambios pueden darse a través del tiempo, como el número de casos positivos por un determinado virus en el planeta entre el 2020 y el 2022 (figura 52); también pueden presentarse al comparar el comportamiento de ciertas unidades de observación frente a otras, como observar en diferentes personas el tiempo que llevan desocupados y buscando empleo. Estas fracciones de tiempo pueden ser en algunos casos largas y en otros pequeñas; así, la dispersión se hace más grande cuando estos datos se comportan muy diferentes entre sí.

Entonces, después de haber estudiado la unidad anterior de medidas de tendencia central, es claro que estas medidas permiten ver hacia qué valor o lado de la escala se mueven los datos del conjunto analizado, pero en muchos casos pueden darse “tendencias ficticias”

dada la dispersión significativa de los valores observados. Piense sencillamente en la Tasa Representativa del Mercado (TRM); esta es una tasa promedio que mide el valor pagado en pesos por unidad de moneda americana, asociada a otro concepto muy aplicado en el mercado financiero, a saber, *volatilidad* (medición de la variabilidad o dispersión en la rentabilidad de un activo). El comportamiento de la TRM entre el 2019 y el 2020 mostró altas fluctuaciones, lo que genera gran incertidumbre en los mercados, por ejemplo, la TRM era de 3432,82 COP (pesos colombianos) el 31 de diciembre del 2020 y menos de un mes después, el 28 de enero del 2021, se ubicaba en 3636,91 COP, es decir, presentó un aumento de casi 6%; si calcula la desviación de la TRM en ese lapso, encontrará que es de 49 pesos, lo cual es muy significativo si de inversiones se habla.



Foto tomada de https://www.freepik.es/fotos-premium/pesos-colombianos-trasfondo-empresarial_29575761.htm

Autor: Johan111

Situación problema

A las entidades observables o medibles que se reconocen como variables propensas al cambio o a las fluctuaciones, pueden medírseles esas variaciones; como bien refieren Reading y Shaughnessy

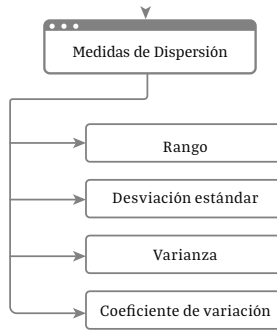


Figura 51. Ejes temáticos de la Unidad 4

Fuente: elaboración propia.

(2004), no es lo mismo variabilidad que variación. Tampoco son equivalentes las medidas de posición con las de dispersión; cada una, aunque trate de dar cuenta del comportamiento de los datos, tiene un fin diferente. Si bien con la aplicación de la estadística se logran reconocer las diferencias latentes entre las medidas de posición y dispersión, es común que estos conceptos no sean claros para la mayoría de las personas.

Precisamente el interés principal al estudiar el comportamiento de los fenómenos a través de la estadística es reconocer y tratar de medir la variación; pues “la variación es el corazón de toda investigación estadística. Si no hubiera variación en los conjuntos de datos, no habría necesidad de estadística” (Watson y Kelly, 2002, p. 1).

Actividad de presaberes



¿Qué significa que un conjunto de datos tiene alta variación?



¿Qué es el rango de los datos?



¿Qué representa la desviación estándar?

Lectura de motivación

Estadísticas cívicas

Los ciudadanos deben conocer y comprender las estadísticas relativas al pasado, tendencias, situaciones actuales y posibles cambios futuros en la demografía, el empleo, los salarios, migración, salud, delincuencia, pobreza, acceso a servicios, energía, educación, derechos humanos y otros dominios para participar plenamente en el ejercicio de la democracia. Las estadísticas y los datos sobre estos y otros temas relacionados se denominan estadísticas cívicas. La información sobre estadísticas cívicas es proporcionada por agencias oficiales de estadísticas y otros proveedores de estadísticas públicos y privados sin fines de lucro, y algunos de ellos se transmiten al público a través de medios impresos y visuales. Se requiere comprensión de las estadísticas cívicas para participar en sociedades democráticas, pero esto involucra datos que a menudo son abiertos, a gran escala, oficiales, multivariados en la naturaleza o presentada dinámicamente; esas estadísticas no suelen ser el núcleo de los informes regulares. Estos materiales llenan un vacío importante no abordado en la educación general a nivel escolar o universitario.

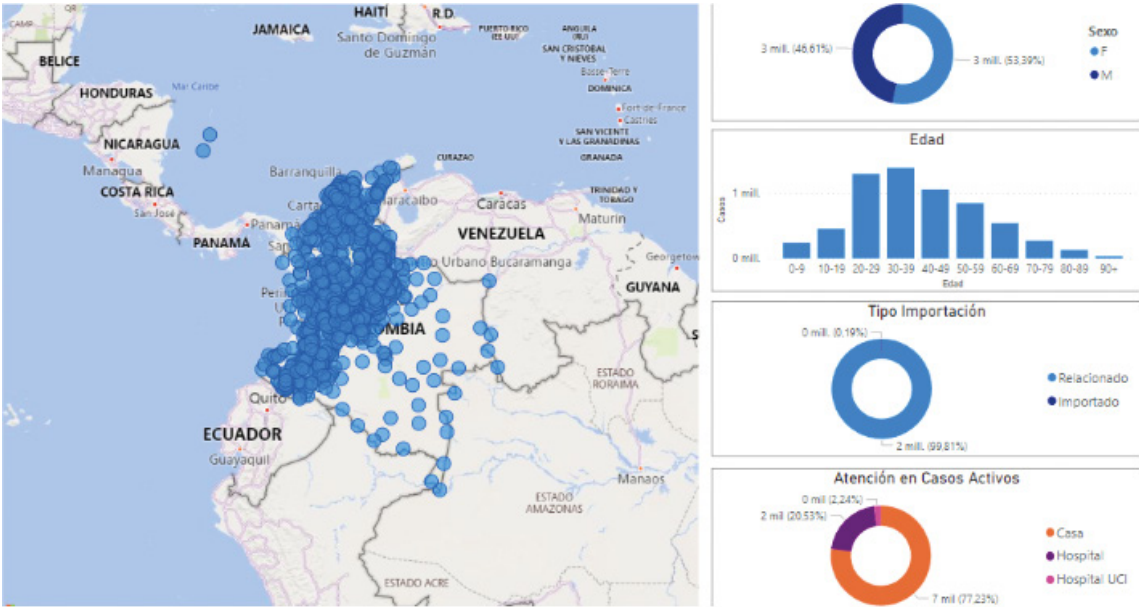


Figura 52. Casos de covid-19 en Colombia durante el 2022

Fuente: Instituto Nacional de Salud de Colombia (2022).

Medidas de dispersión

La variabilidad es uno de los comportamientos evidentes en nuestra sociedad. Bien lo plantean Oddi *et al.* (2020) en su texto *Cuando la variabilidad varía: Heterocedasticidad y funciones de varianza*, en el que refieren que en el ámbito social, en particular en la economía, hay interés en la medición de desigualdad, y en este orden, en conocer por qué la desigualdad del ingreso varía entre países (Li *et al.* 1998) o a qué se debe la volatilidad (variabilidad) de los precios en el mercado financiero (Shiller, 1989). En consecuencia, esta unidad se enfoca en el cálculo de medidas de dispersión como el rango, varianza, desviación estándar y el coeficiente de variación.

Rango

Corresponde a la distancia entre los datos mayor y menor del conjunto analizado. Esto muestra que cuanto más alejados estén estos dos datos, mayor será el valor numérico que corresponda al rango. El rango se calcula así:

$$R = X_{max} - X_{min}$$

Ejemplo 4.1

Dados los valores: 60, 81, 83, 91, 99 calcular el rango.

$$\text{Máximo} = 99$$

$$\text{Mínimo} = 60$$

$$\text{Rango (R)} = 99 - 60 = 39$$

Desviación media

Dado un conjunto de observaciones como $x_1, x_2, \dots, x_N$, la desviación media (DM) se define como el promedio aritmético de las respectivas desviaciones absolutas con respecto a la media. La desviación media se calcula así:

$$DM = \frac{\sum_{i=1}^N |x_i - \bar{x}|}{N}$$

Ejemplo 4.2

El presupuesto ejecutado entre junio y octubre del 2022 en un municipio en Colombia es de 4, 2, 3, 7 y 9 miles de millones de pesos, respectivamente. Calcula la desviación media.

Primero se debe calcular la media:

$$\bar{x} = \frac{4+2+3+7+9}{5} = \frac{25}{5} = 5$$

Segundo se debe calcular la desviación media:

$$DM = \frac{|4 - 5| + |2 - 5| + |3 - 5| + |7 - 5| + |9 - 5|}{5} = \frac{1 + 3 + 2 + 2 + 4}{5} = 2,4$$

El presupuesto ejecutado cada mes difiere en promedio 2 400 000 000 COP de la media aritmética.

Desviación media de datos agrupados

Si en una tabla de frecuencias, $x_1, x_2, ..., x_N$ corresponden a los puntos medios o marcas de clase de los intervalos (o clases), y $f_1, f_2, ..., f_N$ son las frecuencias absolutas, entonces la desviación media se calcula así:

$$DM = \sum_{i=1}^N \frac{|x_i - \bar{x}| \cdot f_i}{N}$$

Ejemplo 4.3

Calcule la desviación media del número de meses buscando empleo que se muestra en la tabla 22.

Tabla 22. Meses buscando empleo

Intervalo	Meses buscando empleo		MC	FA
1	1	4	3	3
2	4	8	6	4
3	8	9	9	5
4	9	10	10	4
5	10	11	11	6
6	11	12	12	4
7	12	13	13	6
8	13	14	14	4
9	14	15	15	4
				40

La tabla 23 presenta los cálculos para obtener la desviación media.

Tabla 23. Desviación media del tiempo buscando empleo (meses)

Intervalo	Meses buscando empleo		MC	FA	MC-Media	MC-Media
1	1	4	3	3	(3-10,20)*FA	23,100
2	4	8	6	4	(6-10,20)*FA	16,800
3	8	9	9	5	(9-10,20)*FA	8,500
4	9	10	10	4	(10-10,20)*FA	2,800
5	10	11	11	6	(11-10,20)*FA	1,800
6	11	12	12	4	(12-10,20)*FA	5,200
7	12	13	13	6	(13-10,20)*FA	13,800
8	13	14	14	4	(14-10,20)*FA	13,200
9	14	15	15	4	(15-10,20)*FA	17,200
				40		102,400

$$DM = \frac{102,400}{40} = 2,560$$

Varianza

Esta es una medida de resumen clasificada como medida de dispersión por cuanto indica qué tanto el nivel de la concentración de los datos realmente representa a la mayoría de estos, o si el comportamiento es tan variable que esas estimaciones a través de valores medios resultan poco fiables; como plantea Quintela (2021) “si hay poca variabilidad, la información de los valores medios será muy precisa”.

Dado un conjunto de observaciones $x_1, x_2, ..., x_N$, la notación σ^2 (basada en el alfabeto griego, y que se lee “sigma cuadrada”) representa la varianza de tales datos, y se calcula así:

$$\sigma^2 = \sum_{i=1}^N \frac{(x_i - \bar{x})^2}{N}$$

Ejemplo 4.4

Usando los datos del ejemplo 4.2, del presupuesto de 4, 2, 3, 7 ejecutado de junio a octubre en el 2022 en un municipio en Colombia se tiene:

$$\sigma^2 = \frac{(4-5)^2+(2-5)^2+(3-5)^2+(7-5)^2+(9-5)^2}{5} = \frac{1^2+3^2+2^2+2^2+4^2}{5}$$

$$\sigma^2 = \frac{34}{5} = 6,8$$

La necesidad de elevar al cuadrado los datos es para evitar la anulación entre los términos como sucede con 7-5= -2 y 9-5 =2; estos dos resultados al sumarse se anularían (darian cero).

Desviación estándar

La varianza se presenta en unidades cuadradas respecto a la variable analizada, pero para mejor interpretación se puede usar la desviación estándar, que representa la raíz cuadrada de la varianza y, en consecuencia, se encuentra en unidades lineales. Se denota con la letra griega sigma , sin el cuadrado, y su cálculo es:

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N}}$$

Ejemplo 4.5

Usando los datos del ejemplo 4.2, del presupuesto ejecutado de junio a octubre en el año 2022 en un municipio en Colombia de 4, 2, 3, 7 se tiene:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{34}{5}} = \sqrt{6,8} = 2,61$$

Ejemplo 4.6

En la tabla 24 se ofrecen tres grupos de datos que tienen la misma media (5).

Tabla 24. Observaciones con igual media

Municipio 1	Municipio 2	Municipio 3
4	5	12
2	5	0
3	5	0
7	5	2
9	5	11

Nota: la tabla contiene el presupuesto ejecutado cada mes (miles de millones de pesos) de junio a octubre en el año 2022 de 3 municipios en Colombia.
Fuente: elaboración propia.

En la tabla 25 sus varianzas son diferentes: 6,8, 0 y 28,8, respectivamente.

Tabla 25. Medidas de resumen datos ejemplo 4.6

	Municipio 1	Municipio 2	Municipio 3
Promedio=	5	5	5
Rango=	7	0	12
Desviación media=	2,4	0	5,2
Varianza=	6,8	0	28,8
Desviación estándar=	2,61	0,00	5,37

Fuente: elaboración propia.

La tabla 25 muestra que el comportamiento del presupuesto en el municipio 2 en los diferentes meses presenta menor dispersión, y cada uno de sus puntos (montos ejecutados del presupuesto) está muy cerca a la media, mientras que los municipios 1 y 3 tienen mayor variabilidad, en cuanto tienen varios valores más alejados al promedio que el municipio 2. El municipio 1 tiene valores extremos 2 y 9; el municipio 3 tiene valores extremos de 0 y 12, y por esto tiene varianza (28,8) y desviación (5,37), la más alta entre los tres municipios.

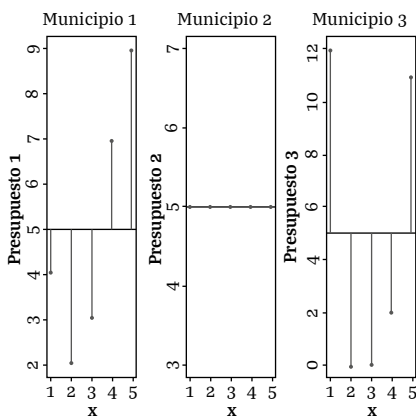


Figura 53. Comparación presupuesto ejemplo 4.5

Fuente: elaboración propia.

Varianzas y desviación estándar de datos agrupados

Si en una tabla las frecuencias, $x_1, x_2, \dots, x_N$ corresponden a los puntos medios o marcas de clase de los intervalos (o clases), y $f_1, f_2, \dots, f_N$ son las frecuencias absolutas, entonces la varianza (σ^2) y (σ) desviación se calculan así:

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \bar{x})^2 \cdot f_i}{N}$$
$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2 \cdot f_i}{N}}$$

Ejemplo 4.7

Usando el ejemplo 4.3, calcula la varianza y desviación estándar del número de meses buscando empleo.

Tabla 26. Medidas de resumen datos agrupados (tiempo desempleado)

Intervalo	Meses buscando empleo		MC	FA	$(x_i - \bar{x})^2 \cdot f_i$	$(x_i - \bar{x}) \cdot f_i$
1	1	4	3	3	$(3-10,20)^2 \cdot 3$	177,870
2	4	8	6	4	$(6-10,20)^2 \cdot 4$	70,560
3	8	9	9	5	$(9-10,20)^2 \cdot 5$	14,450
4	9	10	10	4	$(10-10,20)^2 \cdot 4$	1,960
5	10	11	11	6	$(11-10,20)^2 \cdot 6$	0,540
6	11	12	12	4	$(12-10,20)^2 \cdot 4$	6,760
7	12	13	13	6	$(13-10,20)^2 \cdot 6$	31,740
8	13	14	14	4	$(14-10,20)^2 \cdot 4$	43,560
9	14	15	15	4	$(15-10,20)^2 \cdot 4$	73,960
40					421,400	

Fuente: elaboración propia.

$$\sigma^2 = \frac{421,400}{40} = 10,535$$
$$\sigma = \sqrt{\frac{421,400}{40}} = \sqrt{10,535} = 3,246$$

El número de meses buscando empleo se distancia en promedio 3,246 meses de la media aritmética de ese tiempo.

Coefficiente de variación

Esta medida de la variabilidad relativa es utilizada para comparar la variabilidad de dos (o más) conjuntos de datos, principalmente si estos se encuentran en diferentes unidades de medida. Se obtiene de la relación entre la desviación estándar y la media (promedio), y la fórmula está dada por:

$$CV = \frac{s}{\bar{x}}$$

S: Desviación estándar de los datos

$\bar{x}$: Promedio de los datos

Ejemplo 4.8

Supón que los siguientes datos corresponden al número de pacientes que se atienden por urgencias en un mes en una muestra de 15 hospitales en la ciudad de Bogotá¹⁷: 616, 720, 650, 475, 890, 1250, 3500, 428, 750, 680, 2500, 845, 1500, 620, 900.

Por lo tanto, el número de pacientes que se atienden en las salas de urgencias en estos 15 hospitales de Bogotá es en promedio 1088 pacientes al mes, con una varianza 666 815 (verifica este cálculo). Por otro lado, en Cali se seleccionaron 10 hospitales, obteniendo los siguientes resultados: 425, 780, 569, 1340, 875, 3200, 415, 890, 876, 900.

Así, el número de pacientes que se atiende en las salas de urgencias en estos 10 hospitales de Cali es en promedio de 1027 al mes, con varianza 590 842 (verifica este cálculo).

Los coeficientes de variación de cada ciudad están dados por:

Coeficiente de Variación en Bogotá	Coeficiente de Variación en Cali
$CV = \frac{s}{\bar{x}}$ $= \sqrt{\frac{666815}{1088}}$ $= 0,7505$ $= 75.05\%$	$CV = \frac{s}{\bar{x}}$ $= \sqrt{\frac{590842}{1027}}$ $= 0,7484$ $= 74.84\%$

Conclusión: Con base en estos porcentajes obtenidos, se puede concluir que la variación al interior de cada grupo es bastante grande: 75,05 % y 74,84 %; por otro lado, la variación entre ellos es muy similar.

Práctica 11: medidas de dispersión usando funciones de Excel (TRM)

Las siguientes funciones de Excel permiten calcular las medidas de tendencia central vistas hasta ahora. Para ello, usa los datos de la tabla 27 y luego en una hoja nueva de Excel, ubica la tabla de datos desde la casilla A1:

¹⁷ Ejemplo del sitio web de Estadística Fundamental de la red Unal, Unidad 2: Estadística descriptiva.

Tabla 27. TRM, enero del 2022

Fecha (dd/mm/aaaa)	TRM (COP/USD)
27/01/2022	3947,83
26/01/2022	3987,32
25/01/2022	3977,51
24/01/2022	3964,30
23/01/2022	3964,30
22/01/2022	3964,30
21/01/2022	3980,80
20/01/2022	4003,95
19/01/2022	4033,37
18/01/2022	3993,65
17/01/2022	3993,65
16/01/2022	3993,65
15/01/2022	3993,65
14/01/2022	3950,40
13/01/2022	3970,08
12/01/2022	4011,65
11/01/2022	4043,46
10/01/2022	4043,46
9/01/2022	4043,46
8/01/2022	4043,46
7/01/2022	4039,31
6/01/2022	4042,36
5/01/2022	4084,11
4/01/2022	4082,75
3/01/2022	3981,16
2/01/2022	3981,16
1/01/2022	3981,16

Fuente: Banco de la República de Colombia.

Ahora, ubícate en la casilla B30 y, a partir de allí, empieza a realizar las operaciones que se indican en la tabla 28.

Tabla 28. Medidas de dispersión en Excel

Medida de tendencia central (o posición)	Función de Excel	Resultado
Varianza población=	=VAR.P(B2:B28)	1403,72
Varianza muestral=	=VAR.S(B2:B28)	1457,71
Desviación población=	=DESVEST.P(B2:B28)	37,47
Desviación muestral=	=DESVEST.M(B2:B28)	38,18
Coefficiente de variación poblacional=	=B32/PROMEDIO(B2:B28)	0,00936
Coefficiente de variación muestral=	=B33/PROMEDIO(B2:B28)	0,00954

Fuente: elaboración propia.

	A	B	C
27	2/01/2022	3981,16	
28	1/01/2022	3981,16	
29			
30	Varianza Población=	1403,72	
31	Varianza Muestral=	1457,71	
32	Desviación Población=	37,47	
33	Desviación Muestral=	38,18	
34	Coefficiente de variación Poblacional=	0,00936	
35	Coefficiente de variación muestral=	=B33/PROMEDIO(B2:B28)	

Figura 54. Salida de Excel medidas de dispersión (TRM enero 2022)

Fuente: elaboración propia.

Autoevaluación: ¡Practica tus conocimientos!

Práctica 12: medidas de dispersión en Excel (TRM).

Descarga los datos históricos de la TRM en el sitio web del Banco de la República de Colombia:



De la página del Banco de la República descarga el histórico de la TRM desde el 1 de enero del 2020 hasta el 1 de enero del 2022, luego calcula el rango, varianza, desviación del 2020; repite esto para el 2021 y finalmente obtén el coeficiente de variación del 2020 frente al 2021. Puedes usar las funciones de Excel:

Práctica 13: otros softwares para estadística

Excel es la aplicación más usada a nivel empresarial por su sencillo manejo y su disposición automática en los sistemas operativos que

más utiliza la población. Ahora bien, en cuanto a herramientas para hacer estadística, aunque Excel es bastante práctico, existe *software* especializado, como SPSS, Minitab, SAS y otros de tipo licenciado. No obstante, ahora es muy común el uso de *software* libre, como Infostat en su versión estudiantil; o R, por ser un lenguaje en el que millones de personas contribuyen a través del denominado proyecto R, y que brinda resultados de gran alcance y calidad.

Introducción al Rstudio

R es un lenguaje y entorno de programación para análisis estadístico y gráfico. Es un dialecto libre del lenguaje S, desarrollado por Robert Gentleman y Ross Ihaka del Departamento de Estadística de la Universidad de Auckland en 1993.

- Ver el video tutorial para realizar instalación de R y RStudio: <https://youtu.be/TeXeDzmlQiY>
- Descargar R: <https://cran.r-project.org/bin/windows/base/>
- Descargar RStudio: <https://www.rstudio.com/products/rstudio/download/>

Reconoce la interfaz de RStudio

Para empezar, la interfaz de RStudio se divide en cuatro paneles y un menú superior.

- (1) El menú superior, (2) el editor, (3) el entorno de variables, (4) la consola, y (5) las utilidades y outputs (salidas de gráficos).

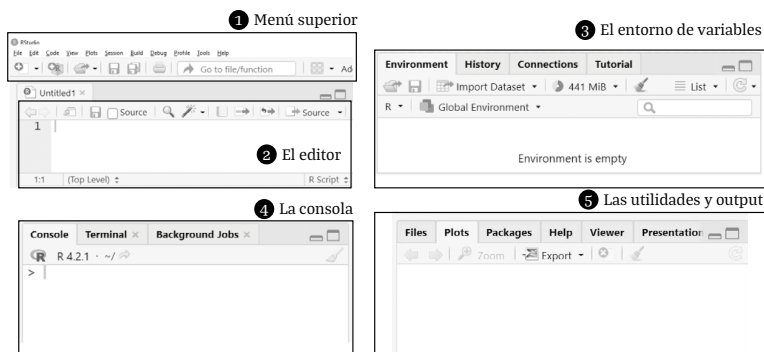


Figura 55. Interfaz inicial de Rstudio (a)

Fuente: elaboración propia.

Script: es un documento de texto en el que escribimos líneas de código de R para estructurar nuestro análisis de datos. Así, estas líneas de código son funciones que sirven para leer datos en R, limpiarlos y visualizarlos. Un *script* es como un archivo de Word donde escribe y puede crear varios archivos y tenerlos abiertos al mismo tiempo. El *script* se ve en el editor.

Para ello, ubícate en el menú superior. Luego: File→New File→ R Script



Comentarios: cualquier línea de texto o secuencia que se escriba en el *script* e inicie con el signo numeral será entendida por R como un comentario y no se ejecutará; el texto se verá de color verde. Ejemplo: Paso 1: Cargar la base de datos

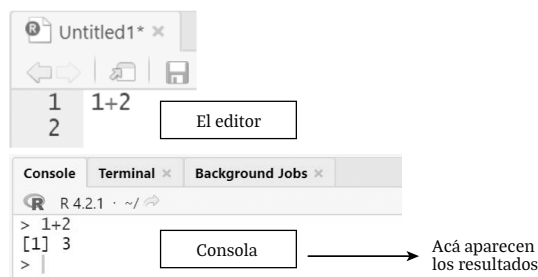
Línea o secuencia de código: es cada una de las filas en el *script*. Ubicados al final de la línea y dando clic en el botón run (flecha verde) se ejecutará el código. La forma rápida es clic + Enter:



Figura 56. Interfaz inicial de Rstudio (b)

Fuente: elaboración propia.

Por ejemplo, si en la primera línea escribes 1+2 y luego al final haces clic + Enter, aparecerá el resultado 3.



Vectores: para crear un vector se usa la función `c()`. Ejemplo: `c(1,2,3,4,5)`
Resultado de correr la línea de código:



Figura 57. Creación de un vector
Fuente: elaboración propia.

Variable: es un espacio en la computadora donde guardamos un objeto, algo así como una caja con un nombre para guardar cosas. Puedes usar cualquier letra o palabra para designar una variable seguida de un igual = o con `<-`. Ejemplo: `Data<-c(1,2,3,4,5)`

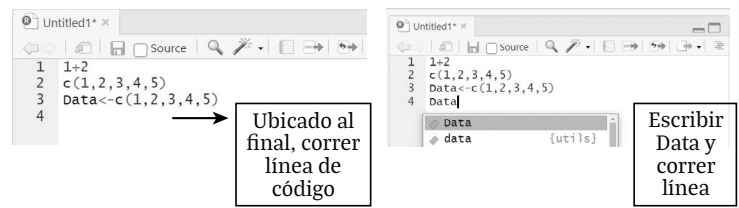


Figura 58. Asignación de objetos y variables en Rstudio
Fuente: elaboración propia.

Práctica calidad de aire en Rstudio

Paso 1: Cargar base de datos. Se carga la base de datos calidad de aire; esta es una de las tantas que están dispuestas en la aplicación de forma

libre para que los usuarios trabajen. Para llamar la base de datos escribiremos en el editor la siguiente línea de código: `Datos.aire<-airquality`

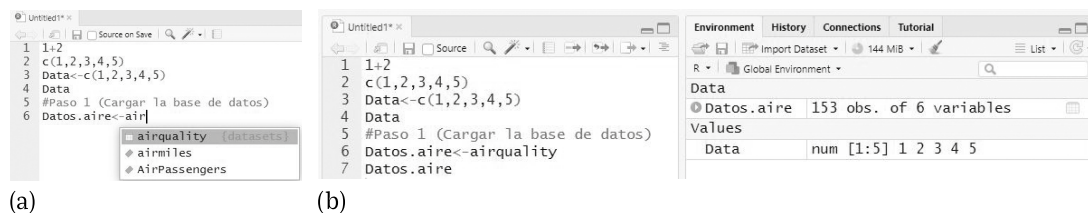


Figura 59. Cargar datos en Rstudio

Fuente: elaboración propia.

`Datos.aire` es como se nombró al objeto que guardará la base de datos; puede llamarlo como desee, sin dejar espacios. En la figura 59, el entorno de variables indica que hay 153 observaciones y 6 variables.

Paso 2: Medidas de resumen. Calcular las medidas de resumen de las seis variables: media, mínimo, primer cuartil, mediana, tercer cuartil y máximo, con la función: `summary(Datos.aire)`

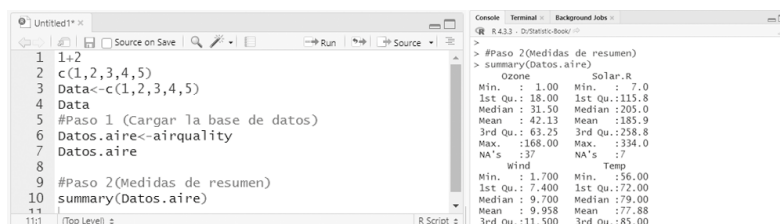


Figura 60. Medidas de resumen en Rstudio

Fuente: elaboración propia.

Paso 3: Gráficos. Para realizar los gráficos de cada una de las seis variables se usa la función: `plot()`¹⁸.

`plot(Datos.aire)`

¹⁸ Si la Ventana de utilidades y output es muy pequeña, puede aparecer un error así: `Error in plot.new() : figure margins too large`. Para solucionarlo simplemente amplie un poco halando los márgenes de la ventana



Fuente: elaboración propia

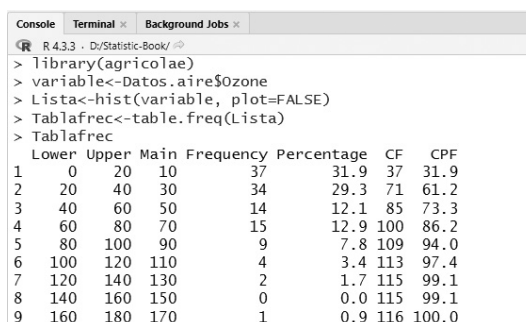
```
hist(Datos.aire$Ozone)
```



Fuente: elaboración propia

```
19
20 #Paso 5 (Tabla de frecuencias)
21 install.packages("agricolae")
22 library(agricolae)
23 variable<-Datos.aire$Ozone
24 Lista<-hist(variable, plot=FALSE)
25 Tablafrec<-table.freq(Lista)
26 Tablafrec
```

28 (Top Level)



```

R 4.3.3 - D:/Statistic-Book/
> library(agricolae)
> variable<-Datos.aire$Ozone
> Lista<-hist(variable, plot=FALSE)
> TablaFreq<-table.freq(Lista)
> TablaFreq

```

	Lower	Upper	Main	Frequency	Percentage	CF	CPF
1	0	20	10	37	31.9	37	31.9
2	20	40	30	34	29.3	71	61.2
3	40	60	50	14	12.1	85	73.3
4	60	80	70	15	12.9	100	86.2
5	80	100	90	9	7.8	109	94.0
6	100	120	110	4	3.4	113	97.4
7	120	140	130	2	1.7	115	99.1
8	140	160	150	0	0.0	115	99.1
9	160	180	170	1	0.9	116	100.0

Figura 63. Tabla de frecuencias en Rstudio

Fuente: elaboración propia

Finalmente puedes guardar el archivo así:

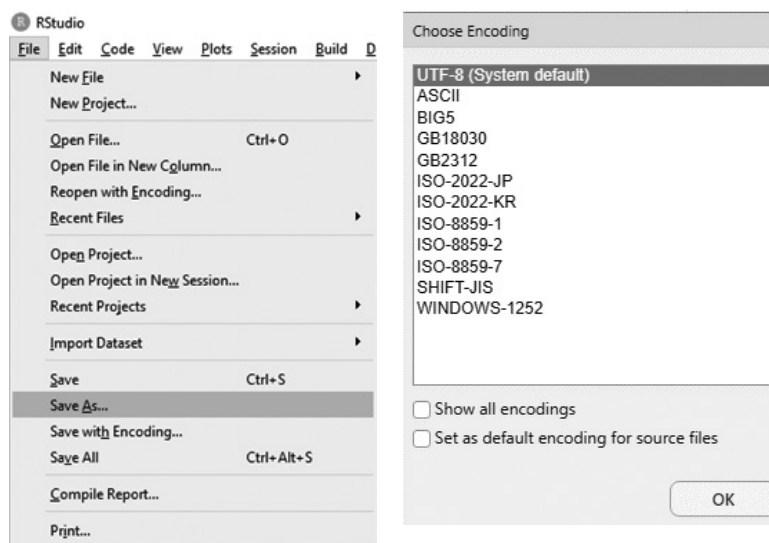


Figura 64. Salvar archivo en Rstudio

Fuente: elaboración propia.

Cargar bases de datos desde formatos Excel u otros formatos

Dirígete al menú superior y allí en la pestaña File→Import Dataset→From Excel...

Luego, en la ventana que aparece, haz clic en *Browse* y selecciona, desde la carpeta correspondiente, el archivo que quieres cargar; para evitar problemas el archivo de Excel debe estar cerrado.



Previewing first 50 entries.

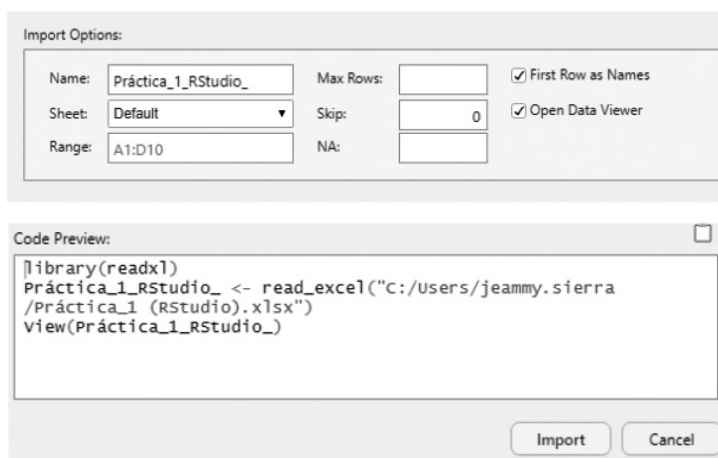


Figura 65. Importación de archivos en Rstudio

Fuente: elaboración propia.

Autoevaluación: gasto en salud per cápita

Elabora la tabla de frecuencias, gráfico y medidas de resumen del gasto en salud para los años 2017, 2018 y 2019.

- Descarga la base de datos 1¹⁹: <https://datos.bancomundial.org/indicador/SH.XPD.CHEX.PC.CD>
- Descarga la base de datos 2: <https://raw.githubusercontent.com/owid/covid-19-data/master/public/data/hospitalizations/covid-hospitalizations.csv>

19 Esta es la base de datos que brinda el Banco Mundial del gasto en salud per cápita de los países. Para esta práctica, solo se trabajará con los países europeos que aparecen, sin contar con los siguientes siete que no tienen información: Azerbaiyán, Albania, Bielorrusia, Ciudad del Vaticano, Letonia, Liechtenstein y Moldavia.

Recursos de aprendizaje

Autor, año y título del recurso	Ubicación	Observaciones
Carvajal, R. (2021). Indicios de calentamiento global en temperatura media y mínima para la ciudad de Bogotá, Colombia 1981-2015. <i>Geográfica digital</i> , 17(34), 19-30. https://revistas.unne.edu.ar/index.php/geo/article/view/4421/4379	https://revistas.unne.edu.ar/index.php/geo/article/view/4421/4379	Este artículo da cuenta del problema de cambio climático, presentando de manera detallada estadísticas descriptivas como temperatura media, mínima, desviaciones y demás calculadas desde estaciones metereológicas en Bogotá, D. C. El trabajo y sus datos en concreto son respecto al calentamiento con mayor concentración en algunas zonas de la capital. Así mismo, el artículo muestra cómo puede aumentar la tasa lineal de calentamiento en 0,07 °C para la temperatura media y en 0,273 °C para la temperatura mínima por década. Esto lleva al estudiante a preguntarse: ¿son las coberturas urbanas un agravante a largo plazo para las tasas de calentamiento global o estas se estabilizan al consolidarse la ciudad? (Carvajal, 2021).
Departamento Administrativo Nacional de Estadística [Dane]. (2017). Tercer Censo Nacional Agropecuario. http://microdatos.dane.gov.co/index.php/catalog/513/get_microdata	http://microdatos.dane.gov.co/index.php/catalog/513/get_microdata	Microdatos de la página del Dane del Tercer Censo Nacional Agropecuario realizado en el 2014 en Colombia. Los datos se encuentran divididos en archivos por cada departamento en una carpeta comprimida, dentro de la cual se encuentra el archivo tipo comas extensión CSV. De la misma manera, se encuentran archivos extensión SPSS. Igualmente, hay un archivo consolidado de todos los departamentos. Para usar estos datos se puede descargar la información de un departamento de Colombia sobre el cual trabajará y presentará un informe estadístico de su diagnóstico. • Proyecto: “Diagnóstico” en el Plan de Desarrollo Territorial. • Dimensión: económica (ver dimensiones Ley 152 y Ley 388) • Línea estratégica: por definir • Sector: agropecuario
Departamento Nacional de Planeación (DNP). (2019). Kit de Planeación Territorial- Unidad 2: Diagnóstico. https://portalterritorial.dnp.gov.co/kpt/	https://portalterritorial.dnp.gov.co/kpt/	Se deja a consulta el objeto virtual de aprendizaje “Kit de Planeación Territorial” que permite comprender el proceso de diagnóstico en la elaboración del Plan de Desarrollo Territorial con todas sus etapas, ubicado en el portal territorial del Departamento Nacional de Planeación (DNP, 2019)

Ministerio de Educación Nacional de Colombia [MEN] (2021). Información y Estadísticas Sectoriales. https://www.mineduccion.gov.co/1759/w3-propertyvalue-57277.html?_noredirect=1	https://www.mineduccion.gov.co/1759/w3-propertyvalue-57277.html?_noredirect=1	Se sugiere seguir las indicaciones en la guía para tener acceso a microdatos anonimizados del MEN, la cual se encuentra en: https://www.mineduccion.gov.co/1759/articles-369982_recurso_3.pdf 1. Realiza el registro del estudiante según indica la guía. 2. Navega por los datos abiertos del MEN, tales como: • Estadísticas sectoriales de educación preescolar, básica y media • Estadísticas sectoriales de educación superior • Estadísticas sectoriales de formación para el trabajo y desarrollo humano Allí encontrarás datos valiosos como: matrícula, establecimientos educativos, estadísticas por departamento y municipio, programas de educación superior, entre otros (MEN, 2021).
Instituto Nacional de Salud de Colombia. (2021). Herramientas de visualización de datos de salud pública. https://www.ins.gov.co/Direcciones/ONS/visor-de-indicadores-en-salud-p%C3%BAblica	https://www.ins.gov.co/Direcciones/ONS/visor-de-indicadores-en-salud-p%C3%BAblica	En este sitio del Instituto Nacional de Salud encontrará indicadores en materia de salud, de interés para el estudiante, que contribuyen en la fase diagnóstico de su ente territorial en el proceso de elaboración del Plan de Desarrollo Territorial (PDT). Algunas de las estadísticas e indicadores que se presentan son: casos de cáncer infantil, dengue, intento de suicidio, morbilidad materna extrema, cáncer de ovario y otros tipos de cáncer, así como muchos indicadores en el ámbito de salud pública. (Instituto Nacional de Salud de Colombia, 2021)
Instituto Nacional de Salud, Observatorio Nacional de Salud, covid-19 en Colombia. (2020). Covid-19 en Colombia, consecuencias de una pandemia en desarrollo. Bogotá, D. C. https://www.ins.gov.co/Direcciones/ONS/Informes/12COVID-19%20en%20Colombia,%20pandemia%20en%20desarrollo.pdf	https://www.ins.gov.co/Direcciones/ONS/Informes/12COVID-19%20en%20Colombia,%20pandemia%20en%20desarrollo.pdf	En este sitio del Instituto Nacional de Salud encontrará indicadores del comportamiento del covid-19 en Colombia.

Referencias

- Li, H., Squire, L. y Zou, H. (1998). Explaining International and Intertemporal Variations in Income Inequality. *The Economic Journal*, 108(446), 26-43. doi: <https://doi.org/10.1111/1468-0297.00271>
- Oddi, F. J., Miguez, F., Benedetti, G. y Garibaldi, L. A. (2020). *Cuando la variabilidad varía: Heterocedasticidad y funciones de varianza*. Asociación Argentina de Ecología.
- Reading, C. y Shaughnessy, J. (2004). Reasoning about variation. En D. Ben-Zvi y J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking* (pp. 169-200). Kluwer Academic Publishers.
- Shiller, R. J. (1989). *Market Volatility*. MIT Press.
- Watson, J. M. y Kelly, B. A. (2002). Can grade 3 students learn about variation? En B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics: Developing a statistically literate society*. International Statistical Institute.

Bibliografía comentada

Instituto Nacional de Salud, Observatorio Nacional de Salud, covid-19 en Colombia. (2020). Covid-19 en Colombia, consecuencias de una pandemia en desarrollo. Bogotá, D. C. <https://www.ins.gov.co/Direcciones/ONS/Informes/12COVID-19%20en%20Colombia,%20pandemia%20en%20desarrollo.pdf>

El Instituto Nacional de Salud presenta el décimo segundo informe técnico denominado covid-19 en Colombia: consecuencias de una pandemia en desarrollo. En el capítulo 1 se expone la naturaleza del SARS-cov-2 y las medidas sanitarias implementadas en el país para evitar el contagio; en el capítulo 2 se explica de manera detallada todo lo que en materia de investigación epidemiológica se había logrado a la fecha para contrarrestar el virus.

Entre tanto, los capítulos 3 a 5 dan cuenta de las afectaciones que trajo el covid-19 –directas, indirectas– y sobre poblaciones especiales. De este intervalo documental se destaca la elaboración de un índice denominado “Índice de Afectación Económica”, a partir del cual se presentan estadísticas bastante importantes en la dinámica territorial.

Los capítulos 6 a 8 relacionan los impactos en poblaciones especiales, como grupos étnicos, personas privadas de la libertad (PPL), inmigrantes y comunidades inmersas en territorios de conflicto armado, entre otros. Las estadísticas presentadas en el informe resultan ser un insumo valioso para el análisis de problemáticas a partir de evidencias; aunque bien vale la pena confrontar los datos mostrados, a través de la sociedad civil organizada de tales grupos de interés, pues es muy frecuente subvalorar cifras, especialmente en contextos en los que las unidades de observación tienen previsiones, como es el caso de los inmigrantes venezolanos.

Números índice

ESTA SECCIÓN FINAL LE PERMITE AL LECTOR ADENTRARSE EN UN TEMA de gran utilidad cuando de análisis económicos se trata; no es raro encontrarse con noticias como “Costo de vida aumentó 10,21 % anual, algo que no se veía desde hace 22 años” (Quiceno Ramírez, 2022), pues el IPC se vio estimulado al alza según los analistas, principalmente por el valor de los alimentos. Pues bien, aunque el concepto es muy mencionado al referir este tipo de factores que inciden en la calidad de vida de los habitantes de un país, no es común que el ciudadano conozca qué representa este tipo de indicadores.

El índice de precios al consumidor es una medida que simboliza el cambio del precio de bienes y servicios representativos del consumo de los hogares del país, conocido como canasta, la cual se obtiene a partir de la encuesta nacional del presupuesto de hogares ENPH que el Dane realiza cada 10 años. Es bastante útil para comparar esos cambios de precios de los bienes y servicios que conforman la canasta, y también permite comparar la economía colombiana con la de otros países y así entender la evolución de la situación económica del país.

Sin embargo, para abordar este tipo de aplicaciones se empieza por explicar qué son los números índices; igualmente se plasman algunos conceptos necesarios como “año base” y “año actual” para la

comprensión y el cálculo de los diferentes tipos de índices que se presentan en esta sección final. Ordenar el gráfico de la unidad le permitirá tener una perspectiva holística de los diferentes tipos de índices, entre los que se cuentan los índices simples y los índices compuestos y cómo a su vez estos pueden clasificarse en otros subtipos.

Adicionalmente, como lectura de motivación, se presenta de forma resumida la metodología para calcular el índice de desarrollo humano, resaltando los resultados para Colombia en las mediciones realizadas para 2020, en donde se ubicó en el puesto 57 al ajustar el índice con los factores ambientales y al compararse con otros países con alta dependencia de los combustibles fósiles y su huella material.



Foto tomada de <https://pixabay.com/es/photos/manzanas-mercado-del-granjero-1841132/>
Autor: Pexels

Situación problema

Las medidas estadísticas de resumen, como el promedio, son herramientas valiosas en el análisis de datos. Sin embargo, su utilidad puede verse comprometida cuando se enfrenta a una alta dispersión de los datos o cuando la distribución de estos muestra un sesgo pronunciado. Un ejemplo claro de esta limitación del promedio se evidencia en la distribución de ingresos, como en el caso del salario promedio, que puede fallar en capturar la complejidad de la realidad económica de los colombianos.

En contraposición a esta limitación, el índice de Gini se erige como un indicador que logra capturar con mayor fidelidad la realidad. Este índice, que evalúa la desigualdad salarial en una escala

que oscila entre 0 (representando la igualdad máxima) y 1 (indicando la desigualdad máxima), proporciona una visión crítica de la situación económica real. A pesar de las fluctuaciones evidentes en el índice de Gini en Colombia, mostrando una reducción en la desigualdad salarial de 0,539 a 0,508 entre 2012 y 2017, el panorama cambia a partir de 2018, alcanzando un valor de 0,544 en 2020 y evidenciando una ligera disminución en 2021, con un Gini reportado de 0,523. Estos valores sitúan el índice en niveles comparables a los observados al inicio del período de referencia (Dane, 2022).

El peso de los datos es otro elemento para considerar, lo que es muy común en el ámbito social, político y económico. Allí los índices pueden dar señales de la existencia o no de instituciones adecuadas para el funcionamiento de la economía y, en consecuencia, medir el desempeño económico de un país con base en criterios institucionales. Una muestra de esto es la configuración del índice de desarrollo humano (IDH), que computa “logros medios obtenidos en las dimensiones fundamentales del desarrollo humano: tener una vida larga y saludable, adquirir conocimientos y disfrutar de un nivel de vida digno” (Eustat, 2022); ¿acaso una disminución del 1 % en la esperanza de vida al nacer tiene el mismo impacto en el IDH que una disminución del 1 % en la educación o los ingresos?

Lectura de motivación

Índice de desarrollo humano

En el marco económico, las propuestas que buscan medir el desarrollo de la humanidad, y las interrelaciones que se generan en este proceso, son numerosas, a la vez que contemplan diversos mecanismos y factores para ese fin. De ahí que la década de 1990 se haya creado como una alternativa al tradicional producto interno bruto (PIB) lo que se conoce como *índice de desarrollo humano* (IDH) clasificando a los países según la libertad y la oportunidad de su población de llevar una vida que valorara. De ese modo afloró un nuevo debate sobre lo que significa llevar una buena vida y las formas de conseguirlo (PNUD, 2020b). En consecuencia, el PNUD gestó el índice de desarrollo humano como una medida que se fundamenta en tres factores: 1. Vida larga y saludable; 2. Educación completa; y 3. Nivel de ingresos dignos.

Los tres factores son estimados a través del índice de esperanza de vida²⁰, índice de educación (este a su vez se calcula basado en 2 elementos: años promedio de escolaridad y años esperados de escolarización) e índice de ingresos (calculado a partir de los ingresos nacional bruto per cápita, INB).

Es preciso señalar que en la metodología se utilizan valores límites para el cálculo del IDH en cada una de las tres dimensiones como se aprecia en la tabla 29.

Tabla 29. Límites del IDH

Dimensión	Indicador	Mínimo	Máximo
Salud	Esperanza de vida (años)	20	85
Educación	Escolarización esperada (años)	0	18
	Promedio de escolaridad (años)	0	15
Estándar de vida	INB per cápita (PPA en US\$)	100	75000

Fuente: tomado de las notas técnicas del reporte de índice de desarrollo humano 2019 (PNUD, 2020a)

Sin embargo, algunos valores mínimos y máximos a nivel mundial que dan cuenta de la profunda desigualdad se pueden observar en la tabla 30.

Tabla 30. Valores dimensiones del IDH Colombia frente a mínimos y máximos en el mundo

Dimensión	Valor observado Colombia	Mínimo observado	Máximo observado
Esperanza de vida al nacer (años)	77,3	53,28 (República Centroafricana)	84,36 (Japón)
Escolarización esperada (años)	14,4	5 (Eritrea)	22 (Australia)
Promedio de escolaridad (años)	8,5	1,6 (Burkina Faso)	14,2 (Alemania)
INB per cápita (PPA en US\$)	14 257	993 (República Centroafricana)	131 032 (Liechtenstein)

Fuente: construcción a partir de cifras tomadas de Nota informativa para los países acerca del Informe sobre Desarrollo Humano presentado en el 2020, con información consolidada al 2019 (PNUD, 2020a); Banco Mundial Esperanza de vida al nacer, total (años) (Banco Mundial, 2020); cifras reportadas por el Instituto Vasco de Estadística (Eustat, 2022) y PNUD (2020b).

20 Es el tiempo estimado que una persona vivirá y depende de varios factores como: edad, género, estado de salud, etnia, ubicación geográfica e incluso nivel socioeconómico. Debido a que el futuro es incierto y sujeto a diversas contingencias, es imposible determinar la EV real de un individuo, sin embargo, se pueden hacer estimaciones utilizando datos empíricos provenientes de tablas de mortalidad (Silva, 2009).

A partir de las cifras anteriores, se calcula:

$$\text{Índice de la dimensión} = \frac{\text{Valor real} - \text{mínimo}}{\text{máximo} - \text{mínimo}}$$

Para la esperanza de vida se toman los valores originales real, mínimo y máximo, mientras que para las otras dos dimensiones se hace una transformación a esos valores, así:

Índice esperanza de vida	Índice escolaridad	Índice ingresos
$\frac{\text{Índice esperanza de vida}}{= \frac{77,3 - 20}{85 - 20} = 0,882}$	(debes hacer tres cosas): 1. Calcular índice de escolaridad esperada $\frac{14,4 - 0}{18 - 0} = 0,80$ 2. Calcular el índice de promedio de escolaridad $\frac{8,5 - 0}{15 - 0} = 0,567$ 3. Con los dos índices anteriores calcular la media aritmética. $\frac{0,80 + 0,567}{2} = 0,683$	A cada valor (real, máximo y mínimo) se le aplica el logaritmo: $\frac{\text{Índice ingresos}}{= \frac{\ln (14\,257) - \ln (100)}{\ln (75\,000) - \ln (100)}} = 0,749$

Fuente: elaboración propia.

Finalmente, se debe calcular un promedio geométrico con estos tres valores, así:

$$\sqrt[3]{0,882 * 0,683 * 0,749} = 0,767$$

En consecuencia, el IDH para Colombia presentado en el 2020 fue de 0,767 (83), lo que muestra un retroceso de cuatro puestos respecto a los resultados presentados el año anterior, cuando ocupaba el lugar 79. Así, el país está por debajo de países como Cuba, que logró un IDH de 0,783 (70); Uruguay, con 0,817 (55), y Argentina con 0,845 (46); posicionándose, con una diferencia mínima, por encima de Brasil, con un IDH de 0,765 (84) y Ecuador, con 0,759 (86).

Sin embargo, es de anotar que detractores de la metodología usada argumentan serios sesgos en los resultados a causa de fenómenos no contemplados que inciden drásticamente en el índice, razón por la cual el PNUD ha contemplado brechas en el acceso a la salud y la educación, incorporando efectos de la desigualdad de género.

Además, debido al interés mundial actual para lograr la adaptación al cambio climático, se han incorporado las presiones planetarias, donde se incorporan las emisiones de dióxido de carbono y la huella material de los países, creando un índice ajustado. Colombia escala al puesto 57 al compararse con otros países con alta dependencia de los combustibles fósiles y su huella material; ajustando el índice con los factores ambientales obtiene un IDH de 0,595 (PNUD, 2020a y b).

A la par de estos esfuerzos en el ajuste del índice, en la dimensión de ingresos estimada se adelantan estudios en la materia, como el trabajo de Cuenca (2020), que incorpora el índice de Gini como indicador de la distribución del ingreso estadísticamente significativo en correlación negativa con el índice de desarrollo humano.

Actividad de presaberes



¿Qué es un
indicador?

.....

.....

.....

.....

.....



¿Qué es un año base?

.....

.....

.....

.....

.....



¿Qué es
un índice?

.....

.....

.....

.....

.....

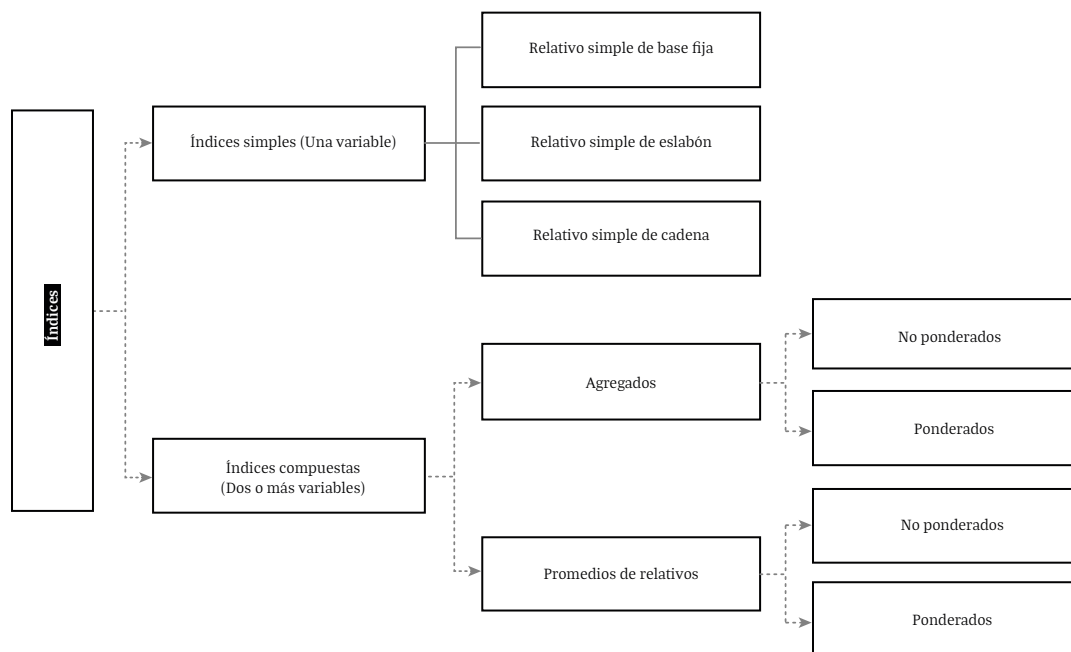


Figura 66. Ejes temáticos de la Unidad 5

Fuente: elaboración propia.

Números índices

Es una escala numérica utilizada para comparar variables entre sí o con algún número de referencia. También se define como una razón (un valor en una escala de medida) derivada de una serie de hechos observados; puede revelar cambios relativos en función del tiempo (Princeton University, 2022).

Es además aquella medida estadística que sirve para comparar una magnitud (o conjunto de magnitudes) en dos situaciones distintas, que pueden ser *temporales* o *espaciales*, y una de esas situaciones suele utilizarse como referencia. Es decir, los índices como medidas estadísticas sirven para comparar magnitudes de una o más variables en un periodo (o lugar) dado, con la magnitud de esa misma o mismas variables en otro periodo (o lugar), uno de esos puntos de espacio o tiempo se considera como referencia y se le denomina base (Aguirre *et al.*, 2005).

Algunos autores tratan como sinónimos los conceptos indicador e índice; no obstante, Delgado (2017) hace una adecuada aclaración al respecto, señalando que los indicadores sirven para indicar, mientras que los índices son los números que sirven para significar el valor de cada indicador.

Al italiano G. R. Carli se le acredita la creación de los números índice en 1764, con la presentación de un reporte de las fluctuaciones de precios en Europa de 1500 a 1750. El índice del costo de la vida se introdujo en 1913, actualmente denominado índice de precios al consumidor, y desde entonces se compila una larga lista de índices (Orion Fortress, 2021).

Periodo base o referencia

Será la situación inicial o periodo tomado como referencia y se denota con el subíndice cero (x_0).

Periodo actual

La situación que quiere compararse subíndice t (x_t)

Índices simples

Son aquellos que se refieren únicamente a una variable o fenómeno, es decir, se construyen para medir los cambios o variaciones (a través del tiempo o del espacio) de una sola variable (estadisticaparatodos.com, 2018, 16 de febrero).

Sea x_t variable a lo largo del tiempo, se obtendrán sus índices dividiendo cada valor de la variable (correspondiente a un momento del tiempo) entre el valor de dicha variable, en el instante que tomaremos como referencia y denominaremos periodo x_k ($k < t$)

$$I_{t/k}(x) = \frac{x_t}{x_k}$$

x_t : el valor de una variable a lo largo del tiempo en el punto t

x_k : el valor de una variable a lo largo del tiempo en el punto k

$l_{t/k}(x)$: Índice en el momento (t) usando como año base k

Es habitual definir el índice en términos porcentuales, para ello solo debemos multiplicar la expresión anterior por 100, es decir:

$$l_{t/k}(x) = \frac{x_t}{x_k} \cdot (100)$$

Al calcular un número índice estamos haciendo en realidad un cambio de variable, convirtiendo la magnitud original x_t en una magnitud $l(x)$. De este modo todos los estadísticos que hayamos obtenido para x_t podrán obtenerse para $l(x)$.

Sean los datos $x_1, x_2, x_3, \dots, x_k, \dots, x_t$ de una única variable x , esos valores pueden resultar de observaciones realizadas a una única unidad de análisis a través de diferentes momentos de tiempo (datos longitudinales) u observaciones realizadas transversalmente.

Ejemplo 5.1

Diferentes momentos del tiempo. Los precios mensuales de la gasolina durante los últimos doce meses en la ciudad de Bucaramanga.

Ejemplo 5.2

Observaciones realizadas transversalmente. Los precios de la gasolina en el último mes en las ciudades capitales provinciales de Colombia.

Variación del índice. Se corresponde con la diferencia relativa del dato en el periodo x_t y el dato en el periodo de referencia x_k , el cual puede ser el dato del periodo inmediatamente anterior (siendo $k < t$).

$$\begin{aligned} \Delta l_{t/k}(x) &= \frac{x_t - x_k}{x_k} \cdot (100) \\ &= \frac{x_t}{x_k} \cdot (100) - 100 \\ &= \frac{x_t}{x_k} - 100 \end{aligned}$$

Un $\Delta l_{t/k}(x) = 40$, sugiere que la magnitud x ha incrementado un 40 % desde el instante k hasta t .

Índice relativo simple de base fija

Este índice se construye para mostrar las variaciones relativas (porcentuales) en los valores de una sola variable, referidos todos estos cambios a un único valor de la serie, llamado valor del periodo base.

$$I_{t/o}(x) = \frac{x_t}{x_o} \cdot (100)$$

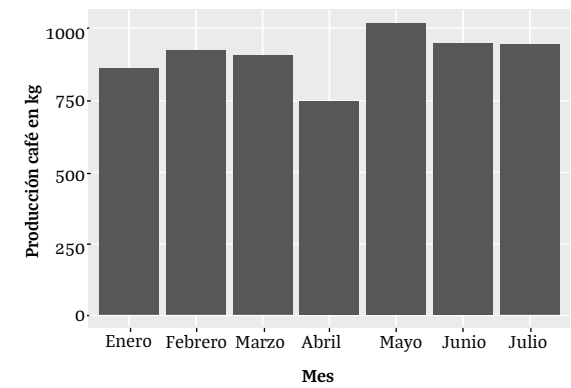
x_t : el valor de una variable a lo largo del tiempo en el punto t

x_o : el valor de una variable a lo largo del tiempo en el punto 0 (base fija)

$I_{t/o}(x)$: índice relativo simple en el momento (t) usando como año base 0

Ejemplo 5.3

La producción mensual de café en Colombia reportada por la Federación Nacional de Cafeteros del país se presenta en la figura 67. Calcular el índice relativo simple de base fija de cada mes hasta julio del 2022, tomando como base fija el mes de enero.



Nota: construcción a partir de estadísticas de la Federación Nacional de Cafeteros de Colombia. <https://federaciondecafeteros.org/wp/estadisticas-cafeteras/>

Figura 67. Producción mensual de café en Colombia 2018-2022

Fuente: elaboración propia.

Tabla 31. Índice relativo simple de base fija (producción de café)

Mes	t	Producción en (t)	$l_{t/0}(x)$	$l_{\text{Febrero/enero}}(x) = \frac{928}{868} \cdot (100) = 106.91\%$
Enero	0	868	100,00	$l_{\text{Marzo/enero}}(x) = \frac{914}{868} \cdot (100) = 105.30\%$
Febrero	1	928	106,91 %	$l_{\text{Abril/enero}}(x) = \frac{750}{868} \cdot (100) = 86.41\%$
Marzo	2	914	105,30 %	$l_{\text{Mayo/enero}}(x) = \frac{1017}{868} \cdot (100) = 117.17\%$
Abril	3	750	86,41 %	$l_{\text{Junio/enero}}(x) = \frac{951}{868} \cdot (100) = 109.56\%$
Mayo	4	1017	117,17 %	$l_{\text{Julio/enero}}(x) = \frac{944}{868} \cdot (100) = 108.76\%$
Junio	5	951	109,56 %	
Julio	6	944	108,76 %	

Fuente: elaboración propia

Los índices relativos con respecto al valor registrado en el mes base de enero muestran que en julio de 2022 la producción de café aumentó un 8,76 %, que es la diferencia entre 108,76 y 100. Analizando la serie completa durante el año, notamos que la producción de café muestra un comportamiento creciente a lo largo del periodo considerado; cada mes hay una producción superior a la registrada en el mes base, y solo decrece en abril, registrándose un 86,41 % de lo que fue en enero, lo que significa una caída del 13,59 %, que es el resultado de restarle a 100 el valor 86,41.

Índice relativo simple de eslabón

Este índice mide los cambios relativos de una sola variable entre dos periodos sucesivos (años, meses, semanas, días, etc.) de una misma serie; es decir, permite expresar en porcentajes la variación ocurrida en los datos entre un periodo “i” cualquiera y el periodo inmediatamente anterior (“i-1”).

$$l_{t/t+1}(x) = \frac{x_t}{x_{t-1}} \cdot (100)$$

x_t : el valor de una variable a lo largo del tiempo en el punto t

x_{t-1} : el valor de una variable a lo largo del tiempo en el punto t-1 (el anterior)

$l_{t/t+1}(x)$: índice relativo simple de eslabón en el momento (t) usando como año base t-1

Ejemplo 5.4

La tabla 32 incluye los valores de la TRM en Colombia durante julio del 2022. Obtén el índice simple de eslabón asociado a esta magnitud.

Tabla 32. TRM en Colombia durante julio de 2022

Fecha	TRM
30/07/2022	4300,30
29/07/2022	4375,5
28/07/2022	4420,75
27/07/2022	4445,01
26/07/2022	4461,63
25/07/2022	4423,86
24/07/2022	4423,86
23/07/2022	4423,86
22/07/2022	4410,14
21/07/2022	4303,34
20/07/2022	4303,34
19/07/2022	4315,41
18/07/2022	4395,63
17/07/2022	4395,63
16/07/2022	4395,63
15/07/2022	4519,65
14/07/2022	4558,05
13/07/2022	4627,46
12/07/2022	4513,28
11/07/2022	4388,27
10/07/2022	4388,27

Fuente: Banco de la República (s. f.).

Tabla 33. Índices simples de la TRM en Colombia durante julio de 2022

Fecha	TRM	
30/07/2022	4300,30	100,00 %
29/07/2022	4375,5	101,75 %
28/07/2022	4420,75	101,03 %
27/07/2022	4445,01	100,55 %
26/07/2022	4461,63	100,37 %
25/07/2022	4423,86	99,15 %
24/07/2022	4423,86	100,00 %
23/07/2022	4423,86	100,00 %
22/07/2022	4410,14	99,69 %
21/07/2022	4303,34	97,58 %
20/07/2022	4303,34	100,00 %
19/07/2022	4315,41	100,28 %
18/07/2022	4395,63	101,86 %
17/07/2022	4395,63	100,00 %
16/07/2022	4395,63	100,00 %
15/07/2022	4519,65	102,82 %
14/07/2022	4558,05	100,85 %
13/07/2022	4627,46	101,52 %
12/07/2022	4513,28	97,53 %
11/07/2022	4388,27	97,23 %
10/07/2022	4388,27	100,00 %

$$l_{t/t-1}(x)$$
$$= \frac{4375,5}{4300,30} \cdot (100)$$

$$l_{t/0}(x)$$
$$= 101,75\%$$

Práctica 14: Índices simples (tasa representativa del mercado, TRM)

Entre enero y junio del 2022 se impusieron 1 862 513 órdenes de comparendo, equivalentes a una variación de 39,3 % con relación al mismo periodo del 2021, cuando se impusieron 1 337 001. A partir de la información de la figura 68, calcular para los meses corridos del 2022:

- Índices simples con base fija, con mes base enero del 2022

• Índices simples de eslabón

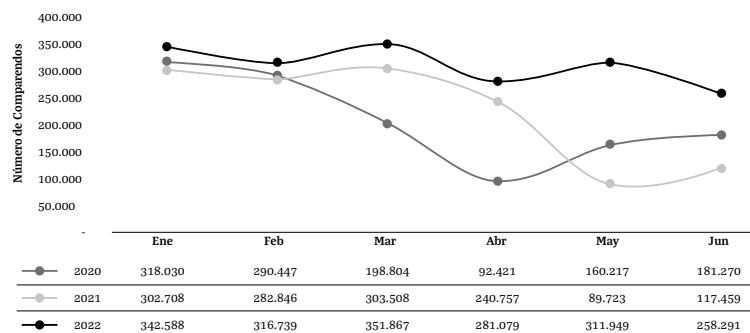


Figura 68. Evolución mensual de las órdenes de comparendos a nivel nacional en Colombia

Fuente: Federación Colombiana de Municipios (2022).

Índice relativo simple de cadena

En ocasiones, es necesario calcular los cambios relativos en una variable con respecto a un único periodo base a partir de los índices en eslabón. Esto implica la construcción de una secuencia de índices que se enlazan entre sí, formando lo que se conoce como un “índice de cadena”. Este tipo de análisis es de suma importancia cuando se busca determinar, por ejemplo, las fluctuaciones promedio en los precios y la producción en un país. La creación y comprensión de estos indicadores económicos es determinante en la toma de decisiones que pueden verse influenciadas por el comportamiento económico de una nación. Algunos ejemplos de estos índices incluyen el Índice de Precios del Productor (IPP) y el Índice de Precios al Consumidor (IPC), entre otros.

Un índice relativo simple de cadena se calcula multiplicando los índices de eslabón resultantes entre el periodo en estudio (“i”) y la base, excluyendo el eslabón del periodo base. Es decir:

$$I_{C\ t/0}(x) = I_{t/t-1}(x) \cdot I_{t/t-2}(x) \cdot \dots \cdot I_{1/0}(x) \cdot (100)$$

Ejemplo 5.5

Usando el ejemplo de la producción mensual de café en Colombia, reportada por la Federación Nacional de Cafeteros (tabla 34), calcula el índice relativo simple de cadena con referencia al periodo base marzo del 2022.

Primero calcular los índices de eslabón:

Tabla 34. Índices de eslabón (producción de café)

Mes	t	Producción	$I_{t/t-1}(x)$
Enero	0	868	100,00
Febrero	1	928	106,91 %
Marzo	2	914	98,49 %
Abril	3	750	82,06 %
Mayo	4	1017	135,60 %
Junio	5	951	93,51 %
Julio	6	944	99,26 %

Nota: producción en miles de sacos de café de 60 kg.

Fuente: elaboración propia.

Ahora realizar los productos:

$$I_{c\text{ julio/marzo}}(x) = (0,9926) \cdot (0,9351) \cdot (1,356) \cdot (0,8206) = 1,03281 \cong 103,281\%$$

Índices compuestos

Camacho y Horta (2020) refieren que si bien no existe una “definición oficial” de lo que es un indicador compuesto, entre las diferentes definiciones propuestas puede decirse que, esencialmente, un indicador compuesto refleja un “sistema complejo” que consiste en numerosos componentes, haciendo más fácil de entender el sistema en su conjunto, en lugar de reducirlo a sus partes separadas

Sin embargo, una definición adecuada de los números índices complejos es que hacen referencia a varias magnitudes a la vez y a su evolución en el espacio o tiempo. Se puede decir que utilizan magnitudes complejas o variables n-dimensionales. Un ejemplo de número índice complejo sería el IPC que aporta información sobre la variación de precios de los productos incluidos en la lista de la compra de las familias de una región.

Greco *et al.* (2018, citado por Camacho y Horta, 2020) presentan una lista de ventajas y desventajas en el uso de los índices compuestos:

Ventajas

- Pueden resumir conceptos complejos o multidimensionales para apoyar a los tomadores de decisiones.

- Son más sencillos de interpretar que tratar de encontrar tendencias entre muchos indicadores separados.
- Facilitan la tarea de construir *rankings* de unidades de análisis (países, regiones, ciudades, etc.) en asuntos complejos, en ejercicios de comparación.
- Permiten evaluar el progreso de las unidades de análisis (países, regiones, ciudades, etc.) en el tiempo en asuntos complejos.
- Reducen el tamaño de un conjunto de indicadores sin perder información relevante, permitiendo incluir más información y manteniendo un límite de tamaño adecuado.
- Centran los asuntos del desempeño y el progreso de la unidad de análisis (país, región, ciudad, etc.) en la arena del diseño de políticas.
- Facilitan la comunicación con el público general (i.e. ciudadanos, medios de comunicación, políticos, etc.) y promueven la responsabilidad.
- Ayudan al desarrollo de un lenguaje común entre los conocedores y los no conocedores del tema.

Desventajas

- Pueden enviar mensajes erróneos o engañosos si están mal contruidos o son mal interpretados.
- Pueden invitar a conclusiones de política demasiado simplistas.
- Pueden ser mal utilizados, por ejemplo, para apoyar una política deseada, si la construcción del índice no es transparente y si carece de principios sólidos tanto estadísticos como conceptuales.
- La selección de los indicadores y los pesos puede ser un desafío político.
- Pueden disimular serios problemas en algunas dimensiones e incrementar la dificultad de identificar las acciones adecuadas para remediarlos.
- Pueden llevar a políticas inadecuadas si las dimensiones de desempeño que son difíciles de medir son ignoradas

Índice compuesto agregado no ponderado

Sea una magnitud compleja H conformada por las magnitudes simples $x_1, x_2, x_3, \dots, x_k, \dots, x_l$ entonces el índice $H(\frac{l}{o})$ será:

$$H(\frac{j}{o}) = \sum_{i=1}^k \frac{X_{ij}}{X_{io}} \cdot (100) \quad i = 1,2, \dots k$$

$H(\frac{j}{o})$: índice compuesto agregado no ponderado

$I(H_i)$: índice simple de la magnitud i

k: cantidad de magnitudes simples que conforman a la magnitud compleja

Ejemplo 5.6

El costo de un programa social dirigido al adulto mayor es C, el cual depende de 3 componentes con la misma importancia relativa C_1, C_2 y C_3 . Calcula para C el índice de la media aritmética simple a partir de la información de la tabla 35.

Tabla 35. Costos de programa social adulto mayor

t	C ₁	C ₂	C ₃
2016	3	1	3
2017	3,5	3	2
2018	3	3	1
2019	2,5	2	1
2020	3	4	4
2021	4	5	5
2022	4,5	7	2

Nota: costos en miles de millones de pesos

Fuente: elaboración propia.

Tabla 36. Índice compuesto relativo no ponderado (costos programa social adulto mayor)

t	C ₁	C ₂	C ₃	$\sum_{i=1}^3 C_{ij}$	I C _{ij}
2016	3	1	3	7	100,00%
2017	3,5	3	2	8,5	121,43%
2018	3	3	1	7	100,00%
2019	2,5	2	1	5,5	78,57%
2020	3	4	4	11	157,14%
2021	4	5	5	14	200,00%
2022	4,5	7	2	13,5	192,86%

$$\begin{aligned} &= 3,5 + 3 + 2 \\ &= 8,5 \end{aligned}$$

$$\begin{aligned} &= 8,5 / 7 = 100 \\ &= 121,43 \% \end{aligned}$$

Fuente: elaboración propia.

En consecuencia, los costos C1, C2, C3 del programa social en el 2017 crecieron en conjunto un 21,43 % (121,43-100) con relación a los costos que se registraron en los tres en el año base 2016.

Índice compuesto relativo no ponderado

Sea una magnitud compleja H, conformada por las magnitudes simples $x_1, x_2, x_3, \dots, x_i$ Entonces el índice $I(\hat{H})$ será el promedio²¹ de todos los índices de cada una de las magnitudes.

$$I(\hat{H}) = \frac{1}{k} \sum_{i=1}^k I(H_i) = \frac{H_1 + H_2 + \dots + H_k}{k}$$

$I(\hat{H})$: índice compuesto no ponderado

$I(H_i)$: índice simple de la magnitud i

k: cantidad de magnitudes simples que conforman la magnitud compleja

Ejemplo 5.7

Usa la información del ejemplo de costo de un programa social dirigido al adulto mayor. Calcula para C el índice compuesto relativo no ponderado a partir de la información de la tabla 35.

1. El primer paso será calcular los índices simples para los precios de cada uno de los componentes.				2. Calcular en cada momento la media aritmética simple de los índices simples.			
t	C ₁	C ₂	C ₃	I(C ₁)	I(C ₂)	I(C ₃)	I(C)
0	3	1	3	100,00	100,00	100,00	100,00 %
1	3,5	3	2	116,67 %	300,00 %	66,67 %	161,11 %
2	3	3	1	100,00 %	300,00 %	33,33 %	144,44 %
3	2,5	2	1	83,33 %	200,00 %	33,33 %	105,56 %
4	3	4	4	100,00 %	400,00 %	133,33 %	211,11 %
5	4	5	5	133,33 %	500,00 %	166,67 %	266,67 %
6	4,5	7	2	150,00 %	700,00 %	66,67 %	305,56 %
= 3,5/3 · 100 = 116,67%				= 4/1 · 100 = 400%			
				= 100+400+133/3 = 211,11%			

21 El promedio puede ser aritmético, armónico o geométrico.

Podemos decir que el costo del producto es un 61,11 % mayor en el momento 1 que en el momento 0. En el momento 2 se ha producido un incremento en el costo del producto del 44,44 % del coste inicial y así sucesivamente.

Se deja al lector usar la misma información para calcular los índices complejos no ponderados como:

- Índice promedio geométrico
- Índice promedio armónico
- Índice promedio agregativo

Tabla 37. Tipos de índices compuestos no ponderados

Índice compuesto no ponderado	Fórmula
Índice promedio aritmético	$I(H) = \frac{1}{k} \sum_{i=1}^k I(H_i) = \frac{H_1 + H_2 + \dots + H_k}{k}$
Índice promedio geométrico	$I(\bar{H}_G) = \sqrt[k]{\prod_{i=1}^k H_i} = \sqrt[k]{H_1 + H_2 + \dots + H_k}$
Índice promedio armónico	$I(\bar{H}_h) = \left(\frac{1}{k} \sum_{i=1}^k l(H_i)^{-1} \right)^{-1} = \frac{k}{\frac{1}{H_1} + \frac{1}{H_2} + \dots + \frac{1}{H_k}}$
Índice promedio agregativo	$I(\bar{H}_A) = \frac{\sum_{i=1}^k X_{it}}{\sum_{i=1}^k X_{i0}} = \frac{X_{1t} + X_{2t} + \dots + X_{kt}}{X_{1-0} + X_{2-0} + \dots + X_{k0}}$

Fuente: adaptado de Trueba, *et al.* (s. f.).

Índice compuesto agregado ponderado

Al construir un índice de precios (o de cantidades) podemos tomar la decisión de ponderar por las cantidades (o precios) del año base, del año que se está analizando o por un valor que promedia ambas magnitudes. Según sea la ponderación que adjudiquemos a cada variable al construir el índice, vamos a estar en presencia de un tipo particular de índice de agregados ponderados.

Los números índices complejos ponderados tienen en cuenta la importancia relativa de cada una de las magnitudes simples que componen el índice complejo.

Para ponderar estas importancias relativas debemos asignar a cada una de las magnitudes un peso que denominaremos w_i , de forma que la suma de los pesos asociados a todas las magnitudes simples debe ser igual a la unidad, es decir:

$$\sum_{i=1}^k w_i = 1$$

Estos pesos pueden variar también a lo largo del tiempo, por lo que definiremos como w_{it} al peso referido a la importancia relativa de la magnitud i en el instante t .

Índice de Laspeyres

Considerando de nuevo la magnitud compleja H , formada por k magnitudes simples H_i , el índice de Laspeyres se define como la media ponderada de los índices simples de las magnitudes, es decir:

$$IP^L_{j/0} = \frac{\sum_{i=1}^k P_{ik} Q_{i0}}{\sum_{i=1}^k P_{i0} Q_{i0}} \cdot (100)$$

Donde:

P_{ik} : es el precio correspondiente al “i-ésimo” artículo en el periodo j .

P_{i0} : es el precio correspondiente al “i-ésimo” artículo en el periodo base.

q_{i0} : es la cantidad correspondiente al “i-ésimo” artículo en el periodo base.

Ejemplo 5.8

Los siguientes son los datos de producción y precio interno de café colombiano y de tomate (enero a julio de 2022). Toma enero como mes base.

Tabla 38. Índices (café versus tomate)

Mes		Producción de café	Precio del café	Producción de tomate*	Precio del tomate
Enero	0	868	\$ 2 148 333	520	\$ 2769
Febrero	1	928	\$ 2 213 333	876	\$ 2841
Marzo	2	914	\$ 1 988 774	867	\$ 2704
Abril	3	750	\$ 2 027 448	967	\$ 3090
Mayo	4	1017	\$ 2 096 733	836	\$ 3550
Junio	5	951	\$ 2 172 233	773	\$ 3024
Julio	6	944	\$ 2 250 290	625	\$ 1970

Nota: *Estimaciones realizadas a partir de evaluaciones agropecuarias municipales de agronet.gov.co (aproximación a sacos de tomates de 60 kg)

Fuente: elaboración propia.

1. Usa una nueva hoja de cálculo en Excel y coloca la información desde la casilla A1

	A	B	C	D	E
1	Mes	Producción Café	Precio café	Producción Tomaté	Precio Tomaté
2	Enero	868	\$ 2.148.333	520	\$ 2.769
3	Febrero	928	\$ 2.213.333	876	\$ 2.841
4	Marzo	914	\$ 1.988.774	867	\$ 2.704
5	Abril	750	\$ 2.027.448	967	\$ 3.090
6	Mayo	1017	\$ 2.096.733	836	\$ 3.550
7	Junio	951	\$ 2.172.233	773	\$ 3.024
8	Julio	944	\$ 2.250.290	625	\$ 1.970

2. Los cálculos para realizar en la columna F, Casilla F2 son:
$$= 868(2148333)+520(2769)$$

Es decir: $=B2*C2+D2*E2$

	A	B	C	D	E	F
1	Mes	Producción Café	Precio café	Producción Tomaté	Precio Tomaté	
2	Enero	868	\$ 2.148.333	520	\$ 2.769	$=B2*C2+D2*E2$
3	Febrero	928	\$ 2.213.333	876	\$ 2.841	\$ 2.056.462.049
4	Marzo	914	\$ 1.988.774	867	\$ 2.704	\$ 1.820.083.981
5	Abril	750	\$ 2.027.448	967	\$ 3.090	\$ 1.523.574.237
6	Mayo	1017	\$ 2.096.733	836	\$ 3.550	\$ 2.135.345.600
7	Junio	951	\$ 2.172.233	773	\$ 3.024	\$ 2.068.131.452
8	Julio	944	\$ 2.250.290	625	\$ 1.970	\$ 2.125.505.315

Arrastra la fórmula para que se completen las casillas F3 hasta F8

3. Los cálculos para realizar en la columna G son:
 $G2= 100$ (pues el mes base)

$$G3 = \frac{868 (2213333) + 520 (2841)}{868 (214333) + 520 (2769)} \cdot (100) = 103,03\%$$

Realiza el mismo proceso en las casillas G4 hasta G8²²

	A	B	C	D	E	F	G
1	Mes	Producción Café	Precio café	Producción Tomaté	Precio Tomaté	$\sum_{i=1}^k P_{ik} q_{i0}$	IP^L_{enero}
2	Enero	868	\$ 2.148.333	520	\$ 2.769	\$ 1.866.193.213	100,00%
3	Febrero	928	\$ 2.213.333	876	\$ 2.841	\$ 2.056.462.049	$=(C3*B2+E3$
4	Marzo	914	\$ 1.988.774	867	\$ 2.704	\$ 1.820.083.981	$*D2)/(C2*$
5	Abril	750	\$ 2.027.448	967	\$ 3.090	\$ 1.523.574.237	$B2+E2*D2)$
6	Mayo	1017	\$ 2.096.733	836	\$ 3.550	\$ 2.135.345.600	103,44%
7	Junio	951	\$ 2.172.233	773	\$ 3.024	\$ 2.068.131.452	103,58%
8	Julio	944	\$ 2.250.290	625	\$ 1.970	\$ 2.125.505.315	103,55%

22 Note que lo que cambia del denominador frente al numerador es el precio, pues se usa la misma producción (cantidades).

Conclusión: el precio de estos cultivos expone –en promedio– un comportamiento creciente. A partir de febrero los valores crecen sostenidamente a lo largo del periodo analizado, y solo en el mes de marzo se presenta el menor valor de la serie, presentando una caída conjunta del 10,47 % (89,86-100) respecto a los precios de enero de 2022. Se resalta además que en julio el café y el tomate crecieron –en promedio– el 3,58 % (103,58-100) con relación a los precios registrados en el mes base.

Tabla 39. Índice compuesto de café y tomate

Mes	Producción Café	Precio café	Producción Tomate	Precio Tomate	$\sum_{i=1}^k = P_{ik} Q_{io}$	IP^L_{enero}
Enero	868	\$ 2 148 333	520	\$ 2769	\$ 1 866 193 213	100,00 %
Febrero	928	\$ 2 213 333	876	\$ 2841	\$ 2 056 462 049	103,03 %
Marzo	914	\$ 1 988 774	867	\$ 2704	\$ 1 820 083 981	89,86 %
Abril	750	\$ 2 027 448	967	\$ 3090	\$ 1 523 574 237	101,96 %
Mayo	1017	\$ 2 096 733	836	\$ 3550	\$ 2 135 345 600	103,44 %
Junio	951	\$ 2 172 233	773	\$ 3024	\$ 2 068 131 452	103,58 %
Julio	944	\$ 2 250 290	625	\$ 1970	\$ 2 125 505 315	103,55 %

Fuente: elaboración propia.

Otros índices

Así como el índice de Laspeyres existen otros índices de precio como el índice de Paasche y el índice de Marshall y Edgeworth. Donde:

P_o : precio del periodo base

P_n : precio del periodo actual

Q_o : cantidad del año base

Q_n : cantidad del periodo actual

Índice de precios de Paasche. La autoría del índice de Paasche se atribuye a Hermann Paasche, un estadístico de origen alemán. Este índice se calcula como un cociente, donde el numerador corresponde a la suma de los precios del periodo real multiplicados por las cantidades respectivas en el mismo periodo. De manera similar, el denominador es

una suma ponderada de los precios en el periodo base multiplicados por las mismas cantidades (Dodge, 2008, p. 302). El cálculo se realiza de la siguiente manera:

$$IP_{(Paasche)n/o} = \frac{\sum P_n Q_n}{\sum P_o Q_o}$$

Índice de precios de Marshal-Edge worth. Este índice se define como un promedio agregado de los índices de precios simples para cada artículo. Este promedio está ponderado por la suma de las cantidades consumidas, producidas o vendidas tanto en el año base como en el año actual (Dubey y Shukla, 2014). El cálculo se realiza de la siguiente manera:

$$IP_{(Marshal-Edgeworth)n/o} = \frac{\sum P_n (Q_o + Q_n)}{\sum P_o (Q_o + Q_n)}$$

Índice de precios de Fisher. propuso la utilización de la media geométrica, dando origen al índice de precios de Fisher, que, como se analizará posteriormente, es considerado uno de los índices superlativos.

$$IP_{(F)n/o} = \sqrt{\frac{\sum P_n Q_n}{P_o Q_o} \cdot \frac{\sum P_n Q_n}{P_o Q_n}}$$

Autoevaluación: ¡Practica tus conocimientos!

En la sección Estadísticas por tema del Dane, consulta precios y costos, y practica calculando algunos índices.



Recursos de aprendizaje

Autor, año y título del recurso	Ubicación	Observaciones
Dane. (2019). Índice de precios al consumidor – IPC. [video de YouTube]. https://youtu.be/rSwdsNCs8xM	https://youtu.be/rSwdsNCs8xM	Video del Dane que explica en qué consiste el IPC.

United Nations Development Programme [UNDP]. (2020). Human Development Reports	http://hdr.undp.org/en/data	Este es el centro de datos de desarrollo humano (United Nations Development Programme, 2020) que le permite ver datos por dimensión, indicador, año o país de forma interactiva y dinámica. Proporciona descargas de más de 150 indicadores globales e índices compuestos para más de 190 países en CSV y PDF. Esta página también proporciona documentación y metodologías, preguntas frecuentes.
Organización Mundial del Turismo [Unwto] (2021). Unwto tourism data dashboard. Unwto. https://www.unwto.org/es/unwto-tourism-dashboard	https://www.unwto.org/es/unwto-tourism-dashboard	El tablero de datos de turismo de la OMT proporciona estadísticas y conocimientos sobre indicadores clave para el turismo receptor y emisor a nivel mundial, regional y nacional. Los datos cubren las llegadas de turistas, la participación del turismo en las exportaciones y la contribución al PIB, los mercados de origen, la estacionalidad y el alojamiento (datos sobre el número de habitaciones, huéspedes y noches). (Organización Mundial del Turismo–Unwto, 2021)
Departamento Nacional de Planeación [DNP]. (2021). Metadatos y Microdatos. https://anda.dnp.gov.co/index.php/catalog	https://anda.dnp.gov.co/index.php/catalog	El Catálogo del Archivo Nacional de Datos (Anda) es un sistema web de catalogación de encuestas que sirve como portal para que los investigadores puedan navegar, buscar y descargar datos y metadatos relevantes de censos, encuestas o registros administrativos Este recurso es una fuente importante para el estudiante para consultar, entre otros: • Encuesta de Percepción ciudadana sobre el Plan Nacional de Desarrollo. • Sisbén III: es el Sistema de Identificación de Potenciales Beneficiarios de Programas Sociales que. A través de un puntaje, clasifica a la población de acuerdo con sus condiciones socioeconómicas. Se utiliza para identificar de manera rápida y objetiva a la población en situación de pobreza y vulnerabilidad para focalizar la inversión social y garantizar que esta sea asignada a quienes más lo necesitan. • Las unidades de análisis son: viviendas, hogares y personas. (DNP, 2021)

Referencias

- Aguirre, C., Niño, M. F., Simonetti, E. F. (2005). *Estadística aplicada en las ciencias sociales y humanas*. Editorial Universitaria de Misiones.
- Banco de la República (s. f.). Tasa Representativa del Mercado (TRM–Peso por dólar). <https://www.banrep.gov.co/es/estadisticas/trm>
- Banco Mundial (2020). *Esperanza de vida al nacer, total (años)*. DataBank. <https://datos.bancomundial.org/indicador/SP.DYN.LE00.IN?lang=de>

- Camacho, M. y Horta, R. (2020). Metodologías para la Construcción de Índices Compuestos. UCU Departamento de Administración y Finanzas. https://www.researchgate.net/profile/Micaela-Camacho/publication/340607621_Metodologias_para_la_Construccion_de_Indices_Compuestos/links/5e9484c8a6fdcca789152a5d/Metodologias-para-la-Construccion-de-Indices-Compuestos.pdf
- Delgado, E. (2017, 13 de junio). Hablemos correctamente: indicador, índice y parámetro. <http://www.acn.cu/hablemos-del-idio-ma/27003-hablemos-correctamente-indicador-indice-y-parametro>
- Departamento Administrativo Nacional de Estadística [Dane] (2022, 26 de abril). Comunicado de prensa. Pobreza monetaria 2021. https://www.dane.gov.co/files/investigaciones/condiciones_vida/pobreza/2021/Comunicado-pobreza-monetaria_2021.pdf
- Dodge, Y. (2008). Paasche Index. En *The Concise Encyclopedia of Statistics*. Springer. https://doi.org/10.1007/978-0-387-32833-1_302
- Dubey, J. y Shukla, D. (2014). Sampling Distribution of Some Special Price Index Numbers. *International Journal of Management and Social Sciences Research*, 3(1), 28.
- Estadística para todos.com (2018, 16 de febrero). Tema 5a. Números índices. <https://estadisticaparatodos.com/numeros-indices/>
- Federación Colombiana de Municipios (2022). Boletín Estadístico Nacional. https://www.fcm.org.co/wp-content/uploads/2022/08/00_Nacional_BE_2022p_I-Sem.pdf
- Greco, S., Ishizaka, A., Tasiou, M. y Torrisi, G. (2018). On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness. *Social Indicators Research*, 1-34.
- Instituto Vasco de Estadística [Eustat] (2022, 10 de noviembre). Índice de Desarrollo Humano. https://www.eustat.eus/estadisticas/tema_852/opt_1/tipo_6/ti_indice-de-desarrollo-humano/temas.html
- Orion fortress. (2021, 22 de junio). Números índice. *Orion Fortress [Blog]*. <https://oriong.weebly.com/sociedad/numeros-indice>

- Princeton University. (2022). Index. A Lexical Database for English.
<https://wordnet.princeton.edu/search?keys=index>
- Programa de las Naciones Unidas para el Desarrollo (PNUD).
 (2020a). *Informe sobre Desarrollo Humano 2019* [Informe].
<https://www.undp.org/es/colombia/publications/informe-sobre-desarrollo-humano-2019>
- Programa de las Naciones Unidas para el Desarrollo [PNUD]. (2020b).
Informe sobre Desarrollo Humano 2020. La próxima frontera. El desarrollo humano y el Antropoceno. <https://hdr.undp.org/system/files/documents/global-report-document/hdr2020spinformesobredesarrollohumano2020pdf.pdf>
- Quiceno Ramírez, J. C. (2022, 5 de agosto). Costo de vida aumentó 10,21 % anual, algo que no se veía desde hace 22 años. *El Colombiano*. <https://www.elcolombiano.com/negocios/inflacion-en-julio-llego-a-1021-niveles-de-hace-22-anos-FJ18330317>
- Silva, R. (2009). Reflexiones acerca de la Esperanza de Vida y una implementación en una hoja de cálculo. *Revista Actuari@*, 1-8. <https://www.colegioactuarial.org/Revista%20Actuario/Publicaciones/023.pdf>
- Trueba, C., Remuzgo, L., Jordá, V., Sarabia, J. M. (s. F.). Estadística I. Tema 7. Números índices y tasas de variación. Open Course Ware, Universidad de Cantabria. https://ocw.unican.es/pluginfile.php/858/course/section/934/cap_7.pdf

Bibliografía comentada

- Programa de Naciones Unidas para el Desarrollo [PNUD]. (2020).
Informe sobre Desarrollo Humano 2020: la próxima frontera. El desarrollo humano y el Antropoceno. PNUD. http://hdr.undp.org/sites/default/files/hdr_2020_overview_spanish.pdf

Este valioso material corresponde a la 30ª edición del informe sobre el índice de Desarrollo Humano [IDH], el cual ha sido denominado “La próxima frontera: El desarrollo humano y el Antropoceno” desarrollado por el Programa de Naciones Unidas para el Desarrollo [PNUD] a

diciembre del 2020 con la información consolidada a 2019. Es un documento bien estructurado de 456 páginas. En esta versión se busca promover el desarrollo, eliminando las presiones planetarias, de ahí que en el apartado n.º 1 se define senda del Desarrollo Humano a partir del respeto por la naturaleza, impulsando la transformación más allá de la necesidades o supervivencia. Dadas las capacidades de la sociedad de actuar basándose en estos datos como nunca antes, es imperativo realizar elecciones que nos alejen de trayectorias potencialmente catastróficas “El Antropoceno no se ha establecido aún formalmente como una nueva época geológica, pero diversos geólogos y científicos que estudian el sistema de la Tierra proponen datar su inicio a mediados del siglo XX”; la sociedad humana está vinculada a través de numerosas redes no solo comerciales y de información, sino también políticas y de infraestructura. El comportamiento humano –condicionado por normas y valores– provoca alteraciones en el funcionamiento del sistema de la Tierra, lo que a su vez tiene efectos de realimentación en las normas, los valores y los comportamientos humanos.

El informe presenta los resultados del índice. Colombia ocupa la posición 83, con un IDH de 0,767, Colombia escala 26 puestos, es decir que llega al lugar 57 de los 189 países al ajustar los valores, considerando estas alteraciones climáticas que son muy alarmantes en otros países dada su dependencia a los combustibles fósiles y otros que son auspiciadores de emisiones de CO₂.

Glosario

Armonización estadística: es un componente de la gestión de las normas estadísticas que realiza el Dane, que tiene como fin la producción y definición continua de conceptos unificados para garantizar integración de la información que hace parte del SEN con la producida en otros países. También hacen parte de la gestión de las normas estadísticas: la adopción de estándares para la documentación metodológica, las buenas prácticas estadísticas y la adopción y adaptación de nomenclaturas y clasificaciones.

Desviación estándar: valor de la variable que indica qué tan alejados están en promedio los datos a la media, es decir, que tan variable es el conjunto.

Estadísticas oficiales: estadísticas producidas y difundidas por las entidades integrantes del SEN que permiten conocer la situación económica, demográfica, ambiental y social a nivel nacional y territorial para la toma de decisiones. Estas cumplen las condiciones y características establecidas en el artículo 1.3.2.1 del Decreto

1743 del 2016: “1. Que la operación estadística que la genera esté incorporada en el Plan Estadístico Nacional. 2. Que la operación estadística que la genera haya aprobado la evaluación de la calidad estadística establecido para el SEN. Parágrafo. Las estadísticas oficiales son de uso obligatorio por parte de las entidades del Estado en documentos y comunicaciones oficiales y para la transmisión de información del país a organismos internacionales.”

Estándares estadísticos: son mecanismos usados por las entidades pertenecientes al SEN que permiten la comparabilidad de las estadísticas oficiales, no solo a nivel territorial, sino en el marco nacional e internacional.

Marco estadístico: conjunto de información (ficheros, listados, directorios, etc.) que permite identificar a todos los individuos de la población. Es la base informativa que se emplea para seleccionar la muestra. En el marco estadístico no siempre está contenido todo el universo (por las omisiones, duplicaciones, unidades mal clasificadas, etc.).

Metadato: información necesaria para el uso e interpretación de las estadísticas. Los metadatos describen la conceptualización, calidad, generación, cálculo y características de un conjunto de datos estadísticos. Atributos o características que identifican y describen la información. En síntesis, el metadato es la información acerca de la información.

Medidas de tendencia central: es el conjunto de medidas estadísticas, entre las cuales se tiene: media aritmética, mediana, moda y cuantiles. Se destaca que estas son las más usadas, pero existen otras medidas de tendencia central.

Media aritmética: valor de la variable que corresponde a una especie de punto de equilibrio, permitiendo que el peso de los datos a lado y lado se balancee, es decir, haciendo que la suma de los valores a izquierda sea igual a derecha, sin importar que puedan ubicarse más datos a un lado que al otro.

Mediana: valor de la variable que corresponde al centro del conjunto de datos ordenados, permitiendo que haya el mismo número de datos a lado y lado del valor mediano.

Moda: valor o valores que ocurren con mayor frecuencia en un conjunto de datos. En el contexto de la administración pública, la moda puede aplicarse para identificar el valor más común o recurrente en un conjunto de datos específico relacionado con aspectos administrativos, por ejemplo, la moda de los salarios.

Microdatos: cada uno de los datos sobre las características de las unidades de estudio de una población (individuos, hogares, establecimientos, entre otras) que se encuentran consolidados en una base de datos.

Problema público: un problema público se reconoce como un escenario negativo de ciudad, país o municipio que abarca a la mayoría de su población; pasa de ser individual para verse en colectivo (Matamoros, 2013, p. 98).

Reserva legal: es la protección de la información cuyo acceso podrá ser rechazado o denegado siempre que dicho acceso estuviere expresamente prohibido por una norma legal o constitucional. Los sujetos obligados deberán mantener un índice actualizado de los actos, documentos e informaciones calificado como clasificado o reservado, de conformidad con la ley de transparencia (Ley 1712 del 2014). El índice incluirá sus denominaciones, la motivación y la individualización del acto en que conste tal calificación. Además, la Ley de Hábeas Data (Ley 1581 del 2012) protege los datos personales, y la Ley 79 de 1993 establece la reserva estadística, según la cual los datos suministrados en el desarrollo de los censos y encuestas no podrán darse a conocer, sino únicamente en resúmenes numéricos que no hagan posible deducir de ellos información alguna de carácter individual.

Sistema Estadístico Nacional (SEN): conjunto articulado de componentes que, de manera organizada y sistemática, garantiza la producción y difusión de las estadísticas oficiales a nivel

nacional y territorial que requiere el país. Sus componentes son las entidades u organizaciones que lo integran, usuarios, procesos e instrumentos técnicos para la coordinación, políticas, principios, fuentes de información, infraestructura tecnológica y talento humano. El SEN se divide en cuatro subsistemas (PEN, p. 22):

1. Sistema de información agropecuario y rural: administrado por el Ministerio de Agricultura y Desarrollo Rural
2. Sistema Integrado de Información sobre Violencias de Género (Sivige)
3. Sistema de Estadísticas en Justicia
4. Sistema de Información Ambiental de Colombia (SIAC)

El SEN dicta que debe haber, a nivel nacional, un plan estadístico nacional (PEN) y, a nivel territorial, un plan estadístico territorial (PET), que se alimentan de los procesos de producción de información (PPI). A modo de ejemplo, véase el PET Cali: <https://www.dane.gov.co/files/sen/plani-ficacion/planes/plan-estadistico-cali.pdf>

Derechos sobre las imágenes

Todas las imágenes diferentes a iconos de uso libre de Office e imágenes con licencias *copyright* tipo “CC: Reconocimiento” utilizadas en la presente producción son de autoría de Jeammy Julieth Sierra Hernández (profesor ocasional tiempo completo de la Territorial Tolima-Escuela Superior de Administración Pública- ESAP) y su uso está bajo la licencia CC-BY-ND:



Sobre la autora

Jeammy Julieth Sierra Hernández

Doctorante en Modelado en Política y Gestión Pública, con estudios de Doctorado en Educational Technology. Es magíster en Administración Financiera y especialista en Actuaría; de profesión, matemática y estadística. Experta en el diagnóstico de problemáticas públicas soportadas en análisis de datos. Ha realizado estudios para la consolidación de líneas base que permitan medir el índice de progreso social (IPS) de las regiones, así como la satisfacción de los beneficiarios de subsidios por parte del Ministerio de Minas y Energía de Colombia. También fue consultora estadística en estudios de Representaciones Sociales de las Mujeres Desplazadas en Ibagué. Se desempeña como docente investigadora vinculada de tiempo completo a la ESAP. Ha sido profesora universitaria por más de 10 años e investigadora en temas como administración pública, administración financiera, administración de empresas e ingenierías, en instituciones de educación superior como la Universidad del Tolima, la ESAP, la Universidad Nacional Abierta y a Distancia (UNAD), a la Corporación Universitaria Minuto de Dios, la Universidad Cooperativa de Colombia y la Universidad Antonio Nariño. Es líder del grupo de investigación Territorium de la Escuela Superior de Administración Pública, colectivo que aborda, entre otras líneas de investigación, “Capacidad institucional” y “Tendencias de cambio en la administración pública”.

ORCID: <https://orcid.org/0000-0003-3267-5283>

Correo electrónico: jeammy.sierra@esap.edu.co

**Estadística I. Análisis de datos para
la toma de decisiones informadas en
la administración** hace parte de la
Colección Didáctica. Para su composición
se usaron fuentes Bagatela y un formato
de 21 cm × 24 cm. Su cuidado estuvo a
cargo de Editorial ESAP, sello editorial de
la Escuela Superior de Administración
Pública, y se imprimió en la Imprenta
Nacional de Colombia.

Los retos y realidades glociales actuales demandan líderes y profesionales capacitados para tomar decisiones informadas en un entorno complejo y cambiante. Desde sus conceptos básicos hasta su aplicación práctica en el sector público, esta obra ofrece un recorrido detallado y accesible sobre la relevancia de la estadística en la toma de decisiones. Así, responde a preguntas como: ¿de qué manera influye la estadística en la planificación del desarrollo territorial? ¿Qué papel desempeñan los datos abiertos en la administración moderna?

Con un enfoque didáctico, cada capítulo explora temas fundamentales como medidas de tendencia central, variación y forma, números índice, entre otros. Además, incluye actividades de autoevaluación, recursos prácticos y bibliografía comentada para fortalecer la comprensión y aplicación de los conceptos estudiados.

Este libro es un recurso valioso tanto para estudiantes como para profesionales en el ámbito administrativo y de la gestión pública, y la administración en general. Su estructura clara y su contenido relevante lo convierten en un compañero indispensable para aquellos que buscan mejorar su capacidad analítica y tomar decisiones fundamentadas.

