

Proyecto Bicentenario Ciudad Segura

Manual del Profesor

Curso: Estadísticas.



Curso: Estadísticas.

Manual del Profesor

Subdirección de Desarrollo y Capacitación	TBD.
Gerencia de Nueva Tecnología	TBD
Código del Curso:	TBD
Desarrolladores:	Serge Ventoruzzo Juan Pasillas Salazar Arturo Arango Juan Pablo Arango Cristina Lara
Contacto	Parque Vía 198, Piso 8, Col. Cuauhtémoc, México, D. F. Tel. (0155) 52226430
Número de actualización:	02
Fecha de actualización	20 de Agosto de 2009.

Registro de actualizaciones

Nombre del Curso o Taller:	Estadística.
Nombre del 1er. desarrollador:	Serge Ventoruzzo Juan Pasillas Salazar
Nombre del último desarrollador:	Serge Ventoruzzo Juan Pasillas Salazar
Código del Curso o Taller:	TBD
Área responsable:	Gerencia de Nueva Tecnología TBD
Ubicación física del área:	Parque Vía 198, Piso 8, Col. Cuauhtémoc

No. actualización	Fecha	Nombre(s) del(los) desarrollador (es)	Modificaciones
		Serge Ventoruzzo. Juan Pasillas Salazar.	Documento Original
		Arturo Arango Juan Pablo Arango Cristina Lara	

Curso de Estadísticas.

Objetivo general:

Al término del curso, los participantes serán capaces de conocer las principales estadísticas disponibles y de interpretar los resultados de cada una.

Contenido

Registro de actualizaciones	3
Capítulo 1: Actualización de los principios de las estadísticas	7
Objetivos de las estadísticas	7
Que es estadística	7
El objetivo de la estadística.....	13
Ramas de la estadística	23
Estadística descriptiva.....	23
Estadística inferencial	23
¿Cuál es la importancia de las etapas del análisis estadístico?.....	27
Variables y constantes	27
Escalas de medición para variables cualitativas y cuantitativas.....	28
Razones	35
Proporciones	36
Porcentajes	37
Tasas	38
Diferentes fases.....	43
Organización de los datos.....	44
Como construir una distribución de Frecuencias	45
Gráficas de pastel	51
Gráficas de columna	57
Rango	66
Intervalo de la clase	66
Histogramas.....	71
Polígono de frecuencias.....	73
Ojiva o curva de frecuencia acumulada	73
Como interpretar	74
Utilización de los modelos.....	76
Principales métodos	85
Medidas de tendencia central	85
Medidas de variabilidad.....	97
Resumen	114
Capítulo 2: Estadísticas disponibles y sus interpretaciones	119
Panorama general	119

Introducción	119
Objetivo.....	119
Estadísticas centro de llamadas	120
Estadísticas incidentes	122
Estadísticas despacho	124
Estadísticas cruzadas con bases de datos de datos externos.....	126
Anexo I. Índice de imágenes, tablas y gráficas	128
Imágenes	128
Tablas	129
Gráficas	130
Bibliografía	131
Fórmulas usadas:	132
Referencias	134
Glosario	135

Capítulo 1:

Actualización de los principios de las estadísticas

Objetivo	Al término del capítulo, el participante reconocerá el concepto general de las estadísticas.
Tabla de contenidos	A continuación se enlistan los temas incluidos en este capítulo 1. Objetivos de las estadísticas. 2. Diferentes fases. 3. Principales métodos.

Objetivos de las estadísticas

Que es estadística

La estadística es una herramienta indispensable para todas las disciplinas del conocimiento universal. Este capítulo muestra la importancia que el estudio de esta materia tiene para la seguridad pública.

Según Rafael Ruiz Harrell, un conocido criminólogo mexicano y articulista del periódico “Reforma” “Cuando se haga el balance de los avances logrados por la criminología en el último cuarto del siglo XX se descubrirá, sin duda, que los más importantes ocurrieron en el terreno de los números. Tal vez haya quienes digan que el desenvolvimiento del concepto de la policía comunitaria, o entender a las tareas preventivas como las acciones necesarias para romper los patrones delictivos, son más importantes todavía, más lo cierto es que ninguna de estas dos tareas puede llevarse al cabo si no se hace de la estadística criminal un instrumento básico de lucha.”

La mayoría de los estudiantes y lectores de este material, probablemente estén tomando el primer curso de estadística en su vida, tanto profesional, como académica. Y lo están haciendo no por mera emoción intelectual, sino porque es la piedra que hay que salvar para obtener la licenciatura o algún grado, o bien porque, en su trabajo cotidiano, en alguna institución policíaca o de seguridad pública, se encuentran con información a la que es necesario clasificar, ordenar y analizar con objeto de combatir de forma eficiente a la delincuencia.

Con escasas excepciones, la mayoría de los lectores no fueron buenos, ni tuvieron notas sobresalientes, en el estudio de las matemáticas ni la estadística y probablemente estudiaron derecho, o eligieron alguna otra profesión, en donde pensaron que se ponían fuera del alcance del estudio de estas materias, o cualesquiera otras relacionadas con los números.

En un primer acercamiento, al estar impartiendo la materia de estadística a estudiantes de la maestría en criminalística, en el Instituto Nacional de Ciencias Penales, nos encontramos que la angustia y el sobresalto se apoderó de muchos ellos, ya que, señalaban, la mayoría venía de disciplinas en donde, tradicionalmente, las matemáticas y la estadística son consideradas como cuestiones esotéricas y “malvadas” inventadas para hacer sufrir a los hombres y mujeres, en su estudio; y más aún. Después de realizar una pequeña encuesta entre los asistentes a ese curso, nos encontramos que la mayoría (9 de cada 10) tenían escaso conocimiento de computación y estaban alejados de la operación de la computadora, y menos aún, sabían manejar algún programa estadístico. Para sorpresa nuestra, encontramos que, inclusive, había personas que, en esta era de la información, ¡jamás habían navegado en Internet!

También encontramos que su conocimiento de matemáticas, era elemental. En ese momento surgió la idea de realizar este material, y se pensó, que así, como esos estudiantes, existían muchas personas que, trabajando en instituciones de seguridad pública, se encuentran todos los días con problemas que pueden ser resueltos de una forma más fácil, utilizando las herramientas estadísticas, pero que, por desconocimiento o por “miedo” a la parte matemática, no tienen posibilidad de aprender dicha materia y, con ello, la forma en que intentan resolver los problemas, a veces no es la más adecuada.

Esperamos que este material ayude a los policías, abogados, analistas, tomadores de decisiones, criminólogos, criminalistas y en general a los profesionales de la seguridad pública en la operación diaria y, tengan, después de estudiar este texto, el conocimiento adecuado para el manejo y operación de grandes volúmenes de información, como suele suceder, en la seguridad pública.

Es importante señalar que este material es complemento del material “Estadística para analistas delictivos y criminólogos” que se presenta en Internet en la página www.seguridadpublicaenmexico.org.mx.

En dicha página se presenta material adicional que permite ejercitar las matemáticas mínimas necesarias, para obtener un mejor provecho de este instrumento. Además para aquellos que tengan el deseo de mejorar su nivel de manejo de las computadoras, se acompaña al material con material que permite al lector ejercitarse, en forma interactiva, no sólo en las matemáticas básicas, sino en el manejo del programa “Excel¹”.

¹ Excel es un programa de Microsoft y es propiedad intelectual de dicha compañía

En la misma página se tiene acceso a ejercicios y material en línea que le permitirán al lector ganar experiencia tanto en el manejo de las bases de datos de la información estadística oficial relativa a la seguridad pública como en la base de datos de las encuestas que ha realizado el Instituto Ciudadano de Estudios sobre la Inseguridad A. C. ICESI, y que hoy por hoy, representan al único material disponible sobre la “cifra negra” a nivel nacional.

El CIDE ha realizado encuestas de victimización sobre el área metropolitana de la ciudad de México y sus resultados pueden verse en la página www.biiacs.cide.edu.

La estadística en la vida diaria	Partamos de una noción ambigua y tentativa: es "delito" toda conducta que castiguen las leyes.
Cifra Negra Por Rafael Ruiz Harrell	El artículo 7o. del Código Penal dice que tiene ese carácter todo "acto u omisión que sancionen las leyes penales", pero hay otros estatutos -como el Código Fiscal, la Ley General de Salud y muchos más-, que también establecen castigos para los infractores, como sucede ante el causante remiso, frente al hospital o el médico incumplidos, o el laboratorio o la fábrica que elaboren productos dañinos.
Grupo Reforma Publicado en el periódico "Reforma" el 12 de febrero de 1996 en la columna "La ciudad el crimen".	Piénsese, así pues, en la totalidad de las leyes y los reglamentos vigentes y admítase que se está frente a un "delito" cuando alguno de ellos lo sanciona.
No se tomó la nota completa	La definición, aun a pesar de sus defectos, nos permite distinguir tres grandes grupos de delitos.
Nota de los autores: Esta nota se realizó en 1996, cuando no existía el ICESI.	Uno es el de aquellas conductas sancionadas que se llevan al cabo, pero no se detectan, como el entierro clandestino que viene a descubrirse setenta años después, o los fraudes de constructores que se ignoran hasta que los revela un terremoto. Otro es el de los delitos detectados, identificados propiamente como un crimen, pero que no se denuncian: la mujer golpeada por el marido o el amante; el robo de una cartera en el metro; las amenazas o los insultos que profirió el vecino; el billete falso que se intenta hacer pasar por bueno; la violación que prefiere ocultarse.
	Queda, por último, el grupo de delitos detectados y denunciados por la ciudadanía ante las autoridades o, en todo caso, descubiertos por ellas -como un cadáver encontrado en un parque, o el hallazgo policiaco de que un negocio tiene un escaparate roto.

Ahora bien: las estadísticas que maneja ... (eliminación parcial de los autores) toda procuraduría, resultan de contar, única y exclusivamente, los delitos del tercer grupo. Es decir: se trata, tan sólo, de los actos delictivos que llegaron a conocimiento de las autoridades respectivas por denuncia, por querrela o por parte policiaco. Los mal llamados "índices delictivos" nos informan de cuántos delitos llegaron a enterarse las autoridades, pero no nos dicen nada sobre cuántos detectó y sufrió la población ni, menos todavía, cuántos se cometieron.

La diferencia entre los delitos conocidos, registrados por ... "las procuradurías" ... (nota de los autores), y los que se cometieron, es lo que hoy se llama "cifra negra".

NEGRURAS DE LA CIFRA NEGRA

Quizá parezca absurdo señalar lo evidente, pero el hecho es que la característica fundamental de la "cifra negra" es que es negrísima. Vamos: se admite que es "oscura" precisamente porque es imposible precisarla. Es "negra" por desconocida.

Quizá deba calificar lo anterior y señalar que es posible llegar a tener, de manera en toda tentativa, una idea sobre la delincuencia detectada. Lograrlo requiere de investigaciones sociales tan complejas y de tan numerosos supuestos estadísticos, que los resultados siempre tienen el carácter de sospechas informadas, nunca el de afirmaciones con una probabilidad confiable.

Lo que es del todo imposible -y lo es por definición-, es el pretender calcular el monto de los delitos no detectados, de tal manera que el conjunto total de los delitos cometidos seguirá siendo indeterminado e indeterminable. Es imposible aclarar la "cifra negra". Hasta aquí la cita de Rafael Ruiz Harrell.

Se pensó en utilizar a la hoja de cálculo señalada en párrafos anteriores, debido a la facilidad de manejo que dicho programa presenta, así como a la comodidad para efectuar algunos programas interactivos, presentes también en la página, que permiten al lector realizar la mayoría de los ejercicios que acompañan al texto, en forma rápida y fácil.

Asimismo, en la página señalada se tendrán otras referencias, lecturas, problemas y ejercicios que no se incluyen en el material escrito; así como ligas a los principales proveedores y generadores de información estadística de seguridad pública, tanto de México como del mundo, de donde se obtuvo la información para la mayoría de los ejercicios que vienen en este texto.

Se buscó presentar la mayoría del material con sencillez y claridad, en las cuestiones teóricas, así como abundante material

práctico que les otorgue confianza para la realización de los ejercicios. Por ello, pensamos seriamente, que los lectores se darán cuenta que el estudio de la estadística no es aquella “cosa malvada” pensada para torturar a los pobrecitos estudiantes, sino que es una herramienta invaluable para la toma de decisiones y para el mejor desempeño de su trabajo cotidiano, en la seguridad pública.

Sólo nos queda señalar que, si bien, este material tiene por objetivo a lectores que se enfrentan, o pretenden hacerlo, a la delincuencia, todos los días, en cualquiera de sus manifestaciones; esto no implica que su uso este limitado, por lo que también puede ser utilizado en algún otro curso introductorio de estadística, ya que a pesar de su orientación, cubre rigurosamente todos los temas que deben estudiarse.

Así pues, en este capítulo se introducen los conceptos básicos de la estadística como son el conocimiento y predicción de poblaciones. Por ello, al finalizar éste, el alumno será capaz de entender cómo utilizar la estadística para caracterizar los fenómenos y problemas que se enfrentan en el combate a la delincuencia y encontrar parámetros que les permitan hacer comparaciones. Además se introducirá en forma muy breve, la importancia que han adquirido las computadoras como herramientas para el estudio de la estadística.

La estadística es una colección de información numérica, que se refiere a datos cuantitativos

La mayoría de las personas piensan en la estadística como una colección de información numérica sobre algún tema en particular; por ejemplo, el nivel de delincuencia en el Distrito Federal o la cantidad de vehículos que se roban todos los días en Morelos.

Así, estas personas, consideran a la estadística como una palabra de uso común que se refiere a datos numéricos o cuantitativos. Por ejemplo, las denominadas estadísticas vitales, representan datos numéricos sobre nacimientos, defunciones, casamientos, divorcios, etc. En los negocios y en la economía; las estadísticas son datos numéricos sobre empleo, producción, precios y costos, entre otros. Las estadísticas de seguridad pública contienen al número de delitos denunciados, la cantidad de averiguaciones previas iniciadas, a los presuntos delincuentes registrados, a los delincuentes sentenciados, al número de presos del país, al número de policías en activo, la cantidad de arrestos que se realizan, etc.

Es por ello que la estadística, desde este punto de vista, se refiere a información acerca de alguna actividad, expresada en números.

A la estadística también se le utiliza para denotar los métodos estadísticos. Siendo estos últimos los métodos que se aplican en

la recolección, organización, presentación, análisis e interpretación de datos cualitativos y cuantitativos. Dichos métodos tienen gran aplicabilidad en las distintas disciplinas: biología, economía, administración, sociología, contabilidad, etc.; y en particular en la seguridad pública.

Por lo que se ha visto la estadística se puede referir a la información cuantitativa numérica, es decir a los datos estadísticos, o puede referirse a los métodos que tratan con la información, es decir, a los métodos estadísticos aplicados en el análisis de los datos y se debe tener cuidado para diferenciar bien unos de otros.

La estadística en la vida diaria

La implementación de la Unidad de Protección Ciudadana, la instalación de cámaras, botones de emergencia y un centro de operaciones permitieron reducir en un 33 por ciento la incidencia delictiva en el primer cuadro de la ciudad.

Reducen delitos en el Centro

Informes de la Secretaría de Seguridad Pública indican que de marzo a la fecha el promedio de delitos diarios pasó de 7.5 a 5 en la zona que va del Correo Mayor a Balderas y de Izazaga a República de Venezuela

Por Arturo Sierra

Grupo Reforma

Sin embargo, en el resto de la Colonia Centro continúa la tendencia a la alza en el número de delitos.

La nota completa puede verse en el sitio de Internet del Periódico Reforma: <http://www.reforma.com/justicia/articulo/325669/> Distrito Federal (1 septiembre 2003).-

Estadísticas de la Procuraduría General de Justicia del DF señalan que en el primer semestre del 2002 se registraron en el Centro 3 mil 559 delitos y en el mismo periodo de este año fueron 4 mil 78, para un incremento del 14.5 por ciento

La aplicabilidad de la estadística presenta niveles de complejidad, identificándose desde los que requieren técnicas muy elaboradas hasta los que únicamente necesitan, por citar un ejemplo, la organización y presentación de un conjunto de datos numéricos en tablas, la elaboración de algunas gráficas o cálculos numéricos específicos.

Hay que considerar que, como la estadística involucra números, esto significa que se requiere un conocimiento básico de matemáticas. En este curso en particular, el nivel de matemáticas necesario, es muy elemental. Aquí se presentan las técnicas estadísticas y procedimientos más simples para auxiliar al investigador especializado en seguridad pública en la toma de decisiones en condiciones de incertidumbre.

El objetivo de la estadística

El objetivo de la estadística es proporcionar las técnicas adecuadas para realizar inferencias sobre una población a partir de una muestra

Es importante entender que la estadística no es sólo una materia más, sino que su objetivo es proporcionar las técnicas adecuadas que permitan efectuar inferencias que se puedan aplicar exitosamente en la toma de decisiones

Por ello, el propósito del material es dar respuesta a preguntas tales como: ¿Qué es estadística? ¿Cómo se puede utilizar la estadística en la seguridad pública? ¿Por qué la estadística es de gran importancia en el estudio del crimen y la delincuencia? ¿Cómo se puede utilizar en el combate a la delincuencia? Estas y otras cuestiones serán analizadas en los distintos capítulos.

¿Por qué debemos estudiar estadística?

Los datos estadísticos han sido utilizados durante muchos años, incluso se puede hablar hasta de siglos, por los gobiernos, como una ayuda en la administración.

En la antigüedad, la estadística era utilizada para reunir o compilar datos sobre el número de personas aptas para el servicio militar, también los datos fueron compilados para registrar nacimientos, casamientos y defunciones. En esa época muchos de los datos recolectados fueron tratados en secreto, dado que su utilización era muy reservada.

Los métodos estadísticos tienen una gran cantidad de aplicaciones; por ejemplo, para mejorar la producción agrícola, para determinar el diseño de los equipos telefónicos, para planificar el control de tráfico, para prevenir epidemias y para mejorar las actividades de las empresas y los gobiernos.

Respecto de la delincuencia en nuestro país, en un cuidadoso estudio sobre la criminalidad en la ciudad de México, a principios del siglo XX se sistematizan las primeras estadísticas sobre el crimen de la ciudad. En esa época, el robo (en la ciudad de México), encabezaba la lista con 39% del total de los motivos de detención; le seguían, en orden de importancia, las riñas y los delitos sexuales. Los homicidios, (...) representaban 6% del total de los delincuentes detenidos².

En 1922, Casimiro Cueto³, en el artículo “Consideraciones generales y aportes para la crítica, estadística de la criminalidad

² Banco Interamericano de Desarrollo. Análisis de la magnitud y costos de la violencia en México. .Documento de trabajo R 331 p. 7. Banco Interamericano de Desarrollo. 1998. Washington D. C.

³ Citado por Martín Gabriel Barrón Cruz. En “La mujer en las estadísticas judiciales (1985- 1997)” comentarios a la tesis que presenta Arturo Arango Durán para ingresar como académico en la Academia Nacional de la Mujer de la Sociedad Mexicana de Geografía y Estadística. Octubre 16, 2002. Mimeo

habida en el Distrito Federal durante el año de 1922”, señalaba lo siguiente:

Hoy, como pocas veces en México, “se ha discutido tanto [...] acerca del incremento que la criminalidad ha tomado [...]y hemos visto conclusiones basadas en números estadísticos formados fantásticamente, hemos apreciado conjeturas con tendencias de dogma [...por otra parte] la sociedad ha sido desastrosamente impresionada por las nuevas formas y por las públicas manifestaciones que ha presentado la criminalidad [...] la prensa [...] se ha encargado de hacer resaltar más estos hechos, y la sociedad juzga, por el cinismo de los delincuentes, que el mal ha llegado al grado más alto, deduciendo de esto, que la justicia se ha extinguido y hasta que se protege al criminal”⁴.

En la actualidad, los datos estadísticos son utilizados en distintas formas. Ello significa que la recolección de datos estadísticos ha presentado una gran expansión, la cual ha ejercido una profunda influencia en casi todos los campos de la actividad humana.

La búsqueda de relaciones Uno de los objetivos principales de la estadística aplicada en el campo de la seguridad pública es obtener relaciones entre diversas acciones, no en el sentido de causa- efecto, sino en el de probabilidad Uno de los objetivos principales de la estadística en el campo de la seguridad pública es obtener relaciones. Por ejemplo, en una nota periodística, de Arturo Sierra, del periódico “Reforma”, acerca de la reducción de delitos en el centro de la Ciudad de México, se señala que “La implementación de la Unidad de Protección Ciudadana, la instalación de cámaras, botones de emergencia y un centro de operaciones permitieron reducir en un 33 por ciento la incidencia delictiva en el primer cuadro de la ciudad” por lo que se está estableciendo una relación “negativa” entre la instalación de ese equipo y la reducción de delitos en el centro de la ciudad de México. Se dice que hay una relación negativa entre dos variables cuando a medida que aumenta una de éstas la otra disminuye

Otras relaciones interesantes pueden darse entre el número de policías y patrullas en determinada localidad y el número de robos existentes, así puede buscarse si cuando aumentan los primeros disminuyen los segundos o si la relación es que, a mayor número de policías, mayor delincuencia. O bien, si no existe alguna relación entre éstos componentes.

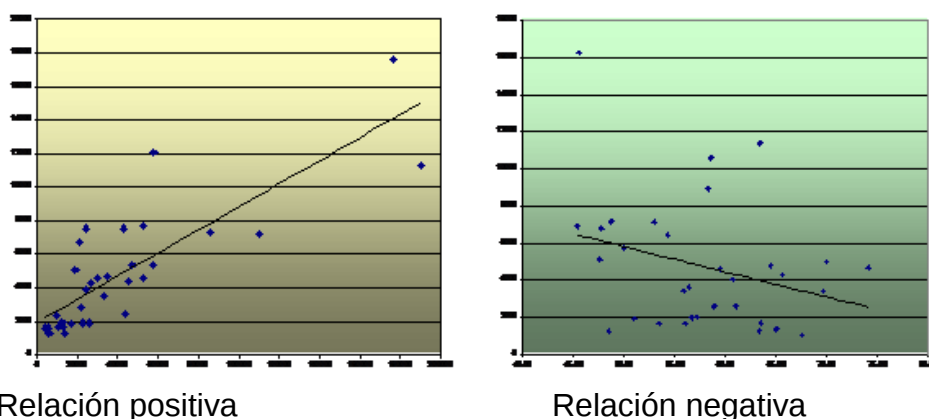
El número de relaciones que se pueden buscar y establecer en seguridad pública son enormes. Otro ejemplo puede ser la

⁴ Cueto, Casimiro.- “Consideraciones generales y aportes para la crítica, estadística de la criminalidad habida en el Distrito Federal durante el año de 1922” en Boletín de la Sociedad Mexicana de Geografía y Estadística, quinta época, T. XII, N° 1-6 (T. XXXVII de la edición completa), 1928, pp. 37-38

relación, que algunos criminólogos piensan que existe, entre los delitos patrimoniales y el nivel de ingreso o de empleo. A menor ingreso o empleo, mayor delincuencia. Otra forma de señalar esta relación es: A mayor desempleo, mayor delincuencia. A mayor pobreza, mayor delincuencia.

Las relaciones que se establecen pueden ser positivas o negativas; es decir, que si uno de los elementos de la relación aumenta, el otro aumenta, o si disminuye cuando el otro lo hace, se dice que dicha relación es positiva, ya que los cambios van en la misma dirección. La relación es negativa, si uno de los elementos aumenta y el otro disminuye, por lo que los cambios se dan en dirección contraria uno del otro.

Imagen 1. Relación positiva y negativa



Otra relación que los abogados y los legisladores, en México, piensan que existe, aunque no haya algún estudio estadístico al respecto, es la que se da entre las sanciones y la disminución de los delitos. Así se piensa que mayores penas en los códigos penales, harán que baje el número de delincuentes dedicados a las actividades cuyas penas son mayores. A mayor sanción menor delincuencia.

De hecho esta última relación, que genera gran controversia en nuestro país, es la que se encuentra presente entre el público y los medios cuando exigen, a los legisladores y a las autoridades, que los delincuentes permanezcan más tiempo en prisión.

La estadística en la vida diaria	Hay una pregunta fácil de plantear, pero difícil de responder: ¿las sanciones más severas disminuyen la delincuencia o la aumentan?
Delitos y Sanciones	La respuesta suele ser afirmativa, mas fuera de señalar que para llegar a esa conclusión no hace falta sino el sentido común, nunca se ofrecen datos y cifras que la apoyen. Esto se debe a varias razones. Una de ellas es que el monto de las penas -fuera del estudio de los códigos respectivos-, sólo puede conocerse analizando las estadísticas judiciales y, por desgracia, están elaboradas de tal manera que la comparación de un Estado a otro resulta imposible.
Por Rafael Ruiz Harrell	
Grupo Reforma	

Un ejemplo: al precisar las penas impuestas en las diversas entidades de la República por un delito dado, digamos el de homicidio, no distinguen en qué casos se refieren a homicidios accidentales y en cuáles a los dolosos. No separan tampoco los homicidios simples, o sea sin agravantes, de los calificados, y como la composición de los mismos es muy diferente de una entidad a otra, al pretender analizarlos se está comparando universos delictivos muy diferentes.

<http://busquedas.gruporeforma.com/utileriasr/imdservicios3W.DLL?JSearchformatS&file=MEX\REFORM01\00323\00323406.htm&palabra=harrell&sitereforma>

La nota completa puede verse en el sitio de Internet del Periódico Reforma

Así pues, La estadística juega un papel importante en la seguridad pública, ya que, “es a partir de los datos ‘crudos’, como podemos saber cuánta delincuencia existe y cuál es su comportamiento: si sube, baja, permanece igual, o si cambia la estructura del delito y del delincuente.”⁵

En las policías, la estadística juega un papel primordial, sobre todo en lo que toca a la prevención y planeación de operativos. Sería imposible saber ni distribuir al personal policiaco sin estar al tanto de en donde, como, y a que hora y día se suscitan los delitos.

En las procuradurías, también la estadística es auxiliar, ya que permite saber cuántas averiguaciones previas se inician y cuantas se concluyen, así como su eficiencia en la resolución de estas, ya sea mediante la consignación o facilitando la negociación entre las partes. Asimismo, es posible saber la actuación de los agentes del

⁵ Arango Durán, Arturo y Lara Medina Cristina. Indicadores de Seguridad Pública en México: La construcción de un sistema de estadísticas delictivas. P. 57. Instituto Nacional de Ciencias Penales

ministerio público respecto a los niveles de consignación vs. inicio así como del nivel de aciertos en la consignación.

En las cárceles la estadística también juega un papel invaluable, ya que es a partir de ésta que se pueden comprar alimentos, ropa, medicinas y planear las actividades educativas, de visita, de construcción de nuevos centros, de contratación de nuevo personal de custodia, etc.

La actividad del gobierno en particular, no podría desarrollarse sin la estadística, ya que, en general, la planeación de actividades de gasto y de desarrollo de obra e infraestructura, tienen como base a la estadística.

Sin lugar a dudas, la estadística ha logrado un gran desarrollo y ha permitido a las diferentes disciplinas hacerse de una herramienta de gran utilidad. La estadística se aplica en todos los campos de actividad, lo cual es la característica de los últimos tiempos.

Los estudiantes de ciencias naturales estudian estadística para convertirse en mejores investigadores; los estudiantes de economía estudian estadística para convertirse en mejores economistas; los estudiantes de administración y/o administración pública estudian estadística para lograr una administración más eficaz.

Los estudiosos de la seguridad pública no podrían realizar ningún análisis sin el auxilio de la estadística y por ello deben estudiarla ya que es a través de ésta como pueden tener un acercamiento a la comprensión de los fenómenos delictivos y sociales.

La estadística puede considerarse como una poderosa herramienta para la investigación en ciencias sociales. Así, en las distintas disciplinas que conforman las ciencias sociales; cada una de estas pone énfasis en diferentes aspectos de su campo de estudio, de ahí la diversidad de predicciones.

En todo proceso, El conocimiento y la predicción dependen del investigador, y la recolección de datos es una fase del desarrollo de cada investigación, con la cual se puede identificar el aspecto a ser considerado o en su caso realizar alguna hipótesis de trabajo que identifique los aspectos de interés. Identificándose la población para el estudio, así como los datos que se requieren utilizar.

Lo precedente permite establecer que, es a través de la estadística como se puede adquirir un conocimiento sobre la sociedad en su conjunto. Así se puede particularizar, en forma estadística, sobre las condiciones que presenta la delincuencia, considerando un espacio geográfico determinado, ya sea una

comunidad, una región o un país, y un tiempo en el que se cometen los delitos.

Por ello, en la investigación criminal es necesario identificar, tanto la fuente de los datos, así como los diferentes métodos del análisis estadístico, es decir, los diferentes métodos o técnicas, para organizar y presentar la información.

Al recopilar datos relativos a las características de individuos u objetos, ya sea por ejemplo edades y sexo de los menores infractores o nivel de ingreso y educación de los probables delincuentes sentenciados por violación, suele ser tedioso o nada práctico observar la totalidad del objeto de estudio, en especial si es muy grande. En vez de examinar al grupo entero, llamado población o universo, se selecciona una pequeña parte del objeto de estudio, al cual se le denomina muestra.

Cuando en estadística se habla de población, no se hace referencia a los seres humanos en particular, sino a los eventos o conjuntos de datos sobre los cuales se intenta realizar una investigación para describir o predecir comportamientos.

Población: Es el conjunto de todas las observaciones de interés, para el investigador social

Si una muestra es representativa de una población, entonces es posible inferir determinadas conclusiones derivadas de las características de la población a partir del análisis de dicha muestra.

Muestra: Es un subconjunto de las observaciones de interés, para el investigador social, seleccionado, en forma aleatoria

Para el estudio de la delincuencia, en el caso de la muestra, se tomará de la población en su conjunto a un número más pequeño de ésta, en donde se intentará averiguar cuál de las personas que están en la muestra ha padecido algún delito.

La parte de la estadística que estudia las condiciones, bajo las cuales, tales situaciones son válidas, se denomina estadística inductiva o inferencia estadística. Puesto que la inferencia no es del todo exacta, en este caso el concepto de probabilidad surge al establecer conclusiones por parte de quienes desarrollan dicho proceso.

Ejercicio 1

Dentro del estudio de la seguridad pública, proporciona cinco ejemplos de casos donde la estadística se aplique.

¿Por qué consideras importante el estudio de la estadística?

Expresa con tus palabras lo siguiente:

Estadística

Dato estadístico

Variable

Población y muestra

Ejemplo: Suponga que se analizan a 10 localidades en donde se encontró el tamaño de la población y la cantidad de policías en cada población, los cuales se presentan en la tabla siguiente:

Tabla 1. Datos hipotéticos de habitantes y policías para 10 localidades

Localidad	Habitantes	Policías
1	2731	30
2	3088	31
3	3235	21
4	4514	41
5	4811	28
6	5915	51
7	7890	53
8	8659	62
9	8934	71
10	9854	65

Supóngase que se desea estimar el promedio de población por localidad que tiene el municipio y se determinó que el tamaño de la muestra es 2, entonces el procedimiento a seguir sería: Seleccionar 2 localidades en forma aleatoria y calcular la media de la muestra (la cual se estudiará más adelante)

Supóngase que se selecciona la localidad 3 y 8, entonces tenemos que la media de la muestra (y por tanto de la población) sería $(3235+8659)/2 = 5947$

Debe notarse que el valor encontrado a partir de la muestra tomada de la tabla anterior, difiere del valor real de la media poblacional el cual es 5963.1, obtenido sumando a todos los habitantes y dividiendo entre la cantidad de localidades. Por otro lado, si las localidades seleccionadas en la muestra hubieran sido la 1 y la 4, entonces se tendría que la media muestral y por lo tanto la poblacional sería de $(2731 + 4514)/2 = 3622.5$ la cual se aleja demasiado del promedio y del valor hallado en la primera muestra.

Esto indica una cosa importante: Diferentes muestras dan lugar a diferentes estimaciones y, además, que algunos parámetros estimados pueden estar cerca del verdadero valor del parámetro poblacional, mientras que otros pueden estar muy lejos.

En el ejemplo que nos ocupa, no hay problema puesto que se conocen los verdaderos parámetros poblacionales, sin embargo

generalmente no se puede hacer un censo y por lo tanto no se conocen los verdaderos parámetros poblacionales, los cuales se deben estimar a partir de una muestra.

En este ejemplo se pueden obtener, de la tabla, 45 muestras de tamaño 2 las cuales se muestran en la tabla siguiente:

Tabla 2. Muestras de tamaño 2, obtenidas de la tabla 1.

Muestra No.	Elementos de la muestra		Media muestral
1	2731	3088	2909.5
2	2731	3235	2383.0
3	2731	4514	3622.5
4	2731	4811	7542.0
5	2731	5915	4323.0
6	2731	7890	5310.5
7	2731	8659	5693.0
8	2731	8934	5832.5
9	2731	9854	6292.5
10	3088	3235	3161.5
11	3088	4514	3801.0
12	3088	4811	3949.5
13	3088	5915	4501.5
14	3088	7890	5489.0
15	3088	8659	5873.5
16	3088	8934	6011.0
17	3088	9854	6471.0
18	3235	4514	3874.5
19	3235	4811	4023.5
20	3235	5915	4573.0
21	3235	7890	5562.0
22	3235	8659	5947.0
23	3235	8934	6084.5
24	3235	9854	6544.5
25	4514	4811	4662.5
26	4514	5915	5214.5
27	4514	7890	6202.0
28	4514	8659	6586.5
29	4514	8934	6724.0
30	4514	9854	7184.0
31	4811	5915	5363.0
32	4811	7890	6350.5
33	4811	8659	6733.0
34	4811	8934	6872.5
35	4811	9854	7332.5
36	5915	7890	6902.5
37	5915	8659	7287.0
38	5915	8934	7424.5
39	5915	9854	7884.5
40	7890	8659	8274.5
41	7890	8934	8412.0
42	7890	9854	8872.0
43	8659	8934	8794.0
44	8659	9854	9256.5
45	8934	9854	9394.0

En las 45 muestras se puede obtener la media muestral, la cual nos da desde un valor pequeño (muestra 1) hasta un valor muy grande (muestra 45).

Esta tabla muestra la distribución de la variable aleatoria "media muestral", y puede ser usada para obtener la definición de variable aleatoria.

La probabilidad de escoger una muestra particular de dos localidades y a partir de dicha muestra obtener un valor particular de la variable aleatoria "media muestral" es $1/45$. Sin embargo, no hay forma posible de saber cual muestra es la que va a ser escogida dado que todas tienen la misma posibilidad de serlo.

Este aspecto de aleatoriedad tiene serias implicaciones y nos enfrenta al problema de la estimación.

Es decir, generalmente se conoce sólo una muestra y es a partir de ésta que se tienen que hacer estimaciones acerca de la población. La tabla anterior nos da una imagen de cuán cerca o cuán lejos puede estar esta estimación del verdadero parámetro poblacional.

Aleatoriedad: ¿La representatividad de una muestra de que depende?

El concepto de aleatoriedad, se vuelve muy importante cuando se quiere describir a una población, a partir de una muestra.

Implica que todos los miembros o elementos de una población bajo estudio tienen la misma probabilidad de ser elegidos y que la elección de cada elemento, se realiza por medio del azar, es decir no está predeterminada

Es decir, ¿Cuál es el criterio para establecer si una muestra es representativa o no de la población? En otras palabras ¿cuál debe ser el tamaño de la muestra n ?

Existen teorías sobre cómo determinar la amplitud de la muestra poblacional. Sin embargo, en este momento nos interesa destacar que, fundamentalmente, la representatividad de la muestra depende del investigador y del objeto de estudio.

Por ejemplo, supongamos el caso del estudio clínico de un paciente, quien requiere un estudio de la sangre, para esto sólo se extraen del paciente cinco centímetros de sangre, misma que es enviada al laboratorio. Esos cinco centímetros corresponden a la muestra de donde se obtienen inferencias del estado que guarda toda la sangre del paciente.

O el caso de las encuestas de preferencias electorales, tan de moda en nuestro país en los últimos tiempos, donde las encuestas de preferencias del voto se realizan con muestras representativas de entre 1500 a 4000 electores, y en donde, la capacidad predictiva ha sido de gran exactitud.

En el caso de las encuestas realizadas por el Instituto Ciudadano de Estudios sobre la Inseguridad⁶ A. C. (ICESI) “... son estudios victimológicos que buscan obtener de informantes victimizados una radiografía de lo que sucede, es decir, un acercamiento al número y tipo de delitos cometidos en un período específico, así como recabar información detallada de determinados tipos de incidentes, las actitudes en torno a la delincuencia y la percepción acerca de las instituciones encargadas de prevenir y procurar justicia y de los niveles de seguridad o inseguridad.”⁷

Esto implica que, a partir de una muestra se pueden obtener estimadores sobre la población que nos permiten realizar inferencias acertadas.

Ejercicio 2

En la siguiente tabla se presentan los datos hipotéticos de 5 delegaciones donde se muestra a los delitos de robo de vehículo y robo a casa habitación

Tabla 3. Datos hipotéticos del delito de robo de vehículo y robo a casa habitación para 5 localidades

Localidad	Vehículo	Casa Habitación
1	27	31
2	38	25
3	25	32
4	41	29
5	81	65

Suponga que a partir de una muestra de tamaño 2, se va a estimar el número de robos de vehículo y el de robos a casa habitación.

- Obtenga la cantidad total de muestras de tamaño 2
- ¿Cuál sería el menor valor estimado?
- ¿Cuál sería el mayor valor?
- ¿Cuál es el verdadero valor?

Tip. Estudie la siguiente tabla

Tabla 4. Obtención de muestras a partir de una población.

Muestra	Localidad	Vehículo	Casa Habitación
1	1, 2	27, 38	31, 25
2	1, 3	27, 25	31, 32

⁶ www.icesi.org.mx véase la sección referente a encuestas, en particular la síntesis metodológica correspondiente a la ENSI- 3

⁷ Tercera Encuesta Nacional sobre Inseguridad 2005 (ENSI - 3). Síntesis Metodológica. P.3. en www.icesi.org.mx

Ramas de la estadística

Por otro lado, tenemos la parte de la estadística que se ocupa de la descripción y análisis de un grupo de información específico, sin obtener conclusiones particulares sobre un grupo mayor, y que se llama estadística descriptiva. Las dos partes de la estadística sobre las que hemos hablado son la estadística descriptiva y la estadística inferencial, mismas que también se les denomina las ramas de la estadística, y que son complementarias, una con otra.

Estadística descriptiva

Partiendo de los datos obtenidos, estos deben ser utilizados en forma práctica. Generalmente los datos de una muestra pueden describirse mediante dos métodos: numérico y gráfico (Tablas y gráficas).

El método gráfico, y con mucho el más simple, incluye las descripciones tabulares, e implica la construcción y elaboración de tablas; además, permite la elaboración de esquemas gráficos. Así, por ejemplo, se elaboran diagramas circulares, gráficos de barras, histogramas, diagramas de dispersión etc. Estos describen de una manera visual la naturaleza de los datos.

El método numérico, implica cálculos numéricos muy simples y permite la descripción aritmética, cuya interpretación proporciona aspectos de la naturaleza de la población bajo estudio a partir de una muestra. Así, se determinan diversos cálculos como las medidas de tendencia central: media aritmética, moda, mediana, etc.; las medidas de dispersión: rango, desviación media, varianza, desviación estándar; los coeficientes: de correlación, los de regresión, y muchos otros más. Dichos números permiten conocer aritméticamente la naturaleza de un conjunto de datos, así como realizar comparaciones de dos o más conjuntos de datos, lo cual permite tomar algunas decisiones.

Estadística inferencial

La estadística inferencial también es conocida como estadística inductiva; sirve para realizar inferencias estadísticas. Para poder desarrollarla, es necesario en un primer momento describir al objeto de estudio (conjunto de datos).

Una población es un conjunto de todos los individuos u objetos de los cuales se quiere saber algo. De antemano sabemos que existen algunos impedimentos para el estudio de la totalidad de la población.

Quizá en algunas cuestiones, muy escasas, si sea posible realizar algún análisis sobre el total de la población bajo estudio. Por ejemplo, en nuestro país se llevan a cabo estudios censales cada

diez años, que pretenden involucrar a toda la población. Se realizan también censos agrícolas, ganaderos y de servicios. Pero los costos y tiempo de proceso de la información, que esta actividad implica, los vuelven muy difíciles de realizar con frecuencia.

Si se desea saber cuál es la edad promedio de los alumnos de un salón de clases, es muy fácil pasar una lista a cada miembro del grupo, y a partir de ésta, obtener la edad promedio. Si se desea saber la edad promedio de los alumnos de una escuela, implica más tiempo y se puede hacer con un poco más de dificultad, pero si se desea saber la edad promedio de los alumnos universitarios de todas las escuelas del país, los costos y tiempo que esto implica, hacen que sea más recomendable, trabajar con una muestra representativa, a partir de la cual se puedan realizar inferencias hacia toda la población en estudio.

Es decir, en los casos en que la población es muy grande, es más fácil obtener una muestra, y a partir de esta, derivar una descripción de la población bajo estudio y se toman decisiones, para toda la población, basados en esa muestra.

Los principales aspectos que se desean conocer de una población se conocen como medidas de tendencia central, y como medidas de dispersión. Estas medidas se conocen como **parámetros poblacionales**. Estos últimos se estiman, o se calculan en forma aproximada mediante valores, frecuentemente denominados **valores estadísticos**, obtenidos de la muestra.

Las estimaciones de los parámetros poblacionales y las pruebas de hipótesis, que se realizan sobre esos parámetros son parte importante de la estadística inferencial.

De esta forma las técnicas de estimación permiten, obtener parámetros poblacionales por medio de valores estadísticos. A partir de estos parámetros se puede describir a toda la población.

Por ejemplo, puede estimarse la educación promedio de menores infractores, de todo el país. En este ejemplo, la población en estudio son todos los menores que cometen una infracción, en particular, y la muestra son sólo ciertos menores infractores, seleccionados de manera aleatoria, en los cuales se busca saber el número de años promedio de educación. Así, puede inferirse el número de años promedio que tienen todos los menores infractores. Es decir, a partir de la muestra se toman decisiones sobre toda la población.

Otro ejemplo puede ser cuando se obtiene mediante encuesta aleatoria, el porcentaje de electores que votaran por un candidato en las elecciones del Distrito Federal. Aquí, la población en estudio consta de todos los habitantes del Distrito Federal, con 18

años o más, que cuentan con credencial de elector y que no están privados de sus derechos electorales, como los delincuentes en prisión, etc. Si la encuesta se aplica a menores de edad, esta encuesta no será válida. Si la encuesta se aplicará por teléfono, ésta tampoco sería válida, ya que no tomaría en cuenta a toda la población bajo estudio, ya que se estaría eliminando a aquellos posibles electores que no cuentan con teléfono en su domicilio.

Un aspecto a considerar, es que al estudiar a la estadística inferencial también se desarrolla lo que hemos denominado, estadística descriptiva, es decir, la estadística inferencial incluye a la estadística descriptiva.

No hay que olvidar que la información cuantitativa puede tener distintas fuentes, mismas que hay que diferenciar, dado que no toda información cuantitativa se representa con datos estadísticos. Así, la información estadística cuantitativa, con características para el análisis estadístico debe ser un conjunto de valores numéricos que permitan establecer relaciones, desde el punto de vista del investigador.

Ello significa que los datos estadísticos estén representados por números que permitan su comparación, su análisis y su interpretación. Por ejemplo si identificamos como fuente de información numérica la edad de una persona, ésta debe ser representada por un número, en este caso de años; sin embargo la sola cuantificación de la edad no es un dato estadístico, si ésta no es comparada con otra. La situación cambia cuando se cuenta con la información de la edad de 100 personas; representando este conjunto de edades, datos estadísticos, ya que permite el estudio de las edades de las personas, puesto se pueden comparar, analizar y sus resultados ser interpretados.

Partiendo del objetivo fundamental que tiene la estadística, que es realizar inferencias o predicciones acerca de las características de una población, basadas en información contenida en una muestra. El estudio de cualquier problema sobre estadística involucra cinco etapas:

La primera, y tal vez la más importante etapa de la investigación estadística, es la formulación o definición del problema.

La segunda etapa de la investigación estadística es la decisión de cómo seleccionar la muestra.

La tercera etapa de la investigación estadística involucra, la recopilación y análisis de los datos estadísticos o de la información estadística;

La cuarta etapa de la investigación estadística, está constituida por la descripción y la organización de los datos;

Finalmente, la quinta etapa de la investigación estadística, es el uso de los datos de la muestra para realizar inferencias o predicciones acerca de la población. Cabe destacar que los problemas de inferencia pueden ser de dos tipos: de estimación o de comprobación de hipótesis.

Una hipótesis estadística es la creencia a priori, que se tiene acerca de una población.

Como ejemplo de una hipótesis acerca de una población puede señalarse que los hijos de familias con desintegración familiar tienen mayor posibilidad de convertirse en delincuentes que los hijos de las familias que tienen mayor integración familiar.

Una vez que se realiza el estudio para probar nuestra hipótesis, es decir, se decide cuál es la población (familias, familias con desintegración familiar, hijos de familia, etc.), se obtiene una muestra representativa y se obtienen los parámetros poblacionales, entonces a través de las técnicas de prueba de hipótesis permiten determinar si se debe aceptar o rechazar la hipótesis supuesta.

Ejercicio 3

Distinga entre población y muestra.

Consulte la Síntesis Metodológica de la ENSI-3 (www.icesi.org.mx) y distinga entre unidades de análisis, unidades de observación, población y muestra.

Describa la relación entre una muestra y la población en referencia a la ENSI- 3.

Busque en algún periódico, e identifique tres aplicaciones de la estadística descriptiva en cuestión de seguridad pública.

Explique los siguientes términos:

Inferencia estadística

Estadística descriptiva

Realice algunos comentarios sobre la relación entre conocimiento y predicción.

¿Cuál considera que es la importancia de una hipótesis?

Con base en la Síntesis Metodológica de la ENSI- 3, ejemplifique la selección muestral e identifique la población específica.

¿Los métodos estadísticos se limitan al estudio de las ciencias sociales? Explique su respuesta y Ejemplifique

¿Cuál es la importancia de las etapas del análisis estadístico?

Mediciones

La investigación en los distintos campos del saber está interesada en el desarrollo de nuevos conocimientos. En este proceso, las hipótesis de trabajo específicas representan guías de investigación. En el capítulo anterior identificamos que a la estadística le conciernen dos tipos de problemas de investigación: a) la descripción de los datos (incluye explicación y análisis de datos) y b) realizar inferencias a cerca de la población o universo; basadas en la información contenida en una muestra.

Al revisar algún artículo de investigación social, están implícitos hechos que nosotros podemos medir, es decir, fundamentalmente se identifican variables de interés tanto políticas, económicas y sociales.

Por ejemplo, es posible clasificar a las personas de acuerdo a la característica racial o realizar mediciones sobre número de nacimientos (índice de natalidad), ingreso familiar, cuantificar la producción total de bienes y servicios de un país, ésta es más difícil de medir. También, se puede medir el desarrollo tecnológico de una sociedad, o la religión de un determinado grupo social. Consecuentemente, antes de realizar discusiones sobre la descripción de datos e inferencia estadística, pasaremos a considerar los tipos de mediciones que se requieren en un estudio de los fenómenos sociales, económicos y políticos.

La primera cuestión a responder es la siguiente: ¿qué es medir? Medir implica el establecimiento de criterios para diferenciar fenómenos, sean estos sociales o físicos. En el estudio de la estadística el medir implica cuantificar un conjunto de observaciones o características.

Variables y constantes

Para el estudio de la estadística es indispensable identificar aspectos de las observaciones, comúnmente denominadas constantes y variables.

Cuando se realizan observaciones de un mismo fenómeno y este no cambia en sucesivos ensayos, se dice que el fenómeno es una constante.

De la misma forma, cuando se realizan observaciones de un mismo fenómeno y este presenta cambios o variaciones en diferentes ensayos, se dice que el fenómeno es una variable. La variable, representa una característica de interés que se puede observar en todas y cada una de las observaciones.

Una variable es cualquier elemento al que se le pueden atribuir diversos valores, Así sexo es una variable que puede asumir dos

valores: masculino y femenino; el peso es una variable que puede asumir una diversa escala de valores que van desde los 0 gramos hasta una gran cantidad de kilos.

Al conjunto de valores que puede asumir una variable se le conoce como distribución de frecuencias.

Si partimos del hecho que la población está integrada por un número específico de elementos y estén estos representados por $\{X_1, X_2, X_3, \dots, X_N\}$, entonces X_1 puede identificar un objeto o persona, según sea el caso. Por otro lado, generalmente a las variables se denotan por las últimas letras del alfabeto (X, Y), mismas que miden las características de interés. Así, partiendo del ejemplo anterior, si la población es un grupo de personas, la variable a considerar puede ser: nivel de ingreso, estatura, sexo, etc. Si la población se identifica con un bosque, la variable puede ser: tipos de árbol, tamaño, clasificación de plantas, etc.

Las variables tienen distintas clasificaciones, es decir, existen las aleatorias, dependientes, independientes, discretas, continuas, cualitativas y cuantitativas. Donde cada una implica particularidades:

Una variable aleatoria, es la que toma al azar los valores de los posibles resultados de un ensayo.

Una variable dependiente, es la que toma los valores correspondientes a un modelo matemático. Ej.: $Y = 3x + 2$. Es decir que el valor de ésta depende de los valores que toma la otra, así en este caso el valor de y depende del valor que tome x, multiplicado por 3 y sumándole 2.

Una variable continua, se deriva de una medición (medida fraccionaria) y puede tomar cualquier valor decimal de una recta. Ej.: 5.22, 8.07, 10.06, 16.1, 54.586, etc.

Una variable discreta, se deriva de un conteo y es la que puede tomar cualquier valor entero. Ej.: 5, 28, 146, 528, etc.

Las variables cuantitativas están representadas en cantidades numéricas, lo cual significa que pueden ser discretas o continuas. Por ejemplo: 8; 9.3; 15; 32.36.

Las variables cualitativas son las que se expresan mediante características. Por ejemplo: Peor, malo, regular, bueno o excelente; defectuoso, aceptable; corto, estándar o largo.

Escalas de medición para variables cualitativas y cuantitativas

La estadística implica el trabajar con observaciones, mismas que poseen un determinado o indeterminado número de elementos (personas u objetos). Los datos, a su vez, tienen su caracterización, es decir, si el estudio involucra a estudiantes;

entre sus características pueden estar el sexo, la edad, el peso, origen, las materias aprobadas, etc.

Al construir estadísticas respecto de sus características, mismas que como se menciona anteriormente se les denomina variables, es posible, contarlas, jerarquizarlas y mediarlas. Así, siguiendo el ejemplo anterior, es posible clasificar y agrupar a las características. Por ejemplo, con relación al sexo de los estudiantes; se puede determinar cuántos son hombres y cuántos son mujeres, de esta forma se clasifica al conjunto de estudiantes en dos tipos: mujeres y hombres.

Dependiendo del objeto de estudio, a las características se les puede jerarquizar, lo cual significa, clasificar en un orden, ya sea creciente o decreciente. Otras características y/o variables que se pueden medir son: tiempo de transporte de los estudiantes, estatura, etc.

Lo precedente pone de manifiesto de que los datos numéricos pueden diferir en cuanto al tipo de medición, según sea una u otra medición con que se pueden trabajar los datos y por consiguiente las aplicaciones matemáticas serán diferentes.

Como fue señalado con anterioridad se identifican las variables cuantitativas y las cualitativas, mismas que contienen una sub clasificación, es decir, a partir de éstas se derivan los niveles de medición.

Los niveles de medición que existen son:

- a) Nominal o clasificatorio
- b) Ordinal
- c) De intervalo, y
- d) De razón

Los dos primeros están asociados a las variables cuantitativas y el tercero y el cuarto a las variables cualitativas.

Nivel de medición nominal.

Representa un nivel de medición donde únicamente la clasificación se realiza según la característica o características del objeto de estudio. Para denotar la medición nominal, se utilizan símbolos, letras o números. Cuando se utilizan números, para la diferenciación de la clasificación, éstos tienen un valor simbólico y no numérico. Ejemplo: clase I, clase II; la diferenciación de los sexos por letras: M para masculino y F para femenino o M para mujer y H para hombre etc.

Nivel de medición ordinal.

Son variables que presentan un orden jerárquico establecido. Ejemplo primero, segundo, último, etc.

Nivel de medición de intervalo.

Tienen el cero arbitrariamente situado y éste no representa la ausencia de la característica que se está midiendo. Ejemplo: 5-10-15; 2-4-6.

Nivel de medición de razón.

Representa la ausencia de la característica que se quiere medir y tienen al cero como absoluto. Además de tener las propiedades de las escala nominal, ordinal y de intervalo.

Ejercicio 4

- 1 Explique la importancia de las mediciones.
- 2 Describa las diferentes formas de realizar mediciones.
- 3 Explique la diferencia entre una variable y una constante.
- 4 ¿Cuál es la importancia de las variables cuantitativas y cualitativas?
- 5 Señale la diferencia que existe entre los niveles de medición.
- 6 Para cada una de las siguientes variables, defina el nivel de medición como cualitativo o cuantitativo. Igualmente señale si son de carácter nominal, ordinal, de intervalo o de razón.
 - a. Sentencia de un delincuente en años
 - b. Delitos cometidos por un delincuente
 - c. Edad de los policías
 - d. Delitos en el código penal
 - e. Castigo por infracciones al código penal, en años.
 - f. Sexo de las víctimas
 - g. Nivel de educación de los delincuentes
 - h. Tipo de drogas utilizadas
 - i. Frecuencia de utilización de drogas, por los delincuentes detenidos por ese motivo
 - j. Monto de lo robado en casas habitación
 - k. Tipo de arma utilizada para cometer un delito
 - l. Marca de los vehículos robados

Comparación de variables utilizando razones, proporciones, porcentajes y tasas

Al igual que hay distintos niveles de medición también existen diversas formas de presentar los datos, para ello se pueden utilizar razones, proporciones, porcentajes y tasas.

Ayuda para el maestro

Actividades

Ejercicios

Capítulo 1.

Tabla 1.xls

Niveles de medición

El primer paso en la estadística es contar. La tabla 1, que se muestra más abajo, presenta la cuenta de víctimas hombres, mujeres y total de víctimas.

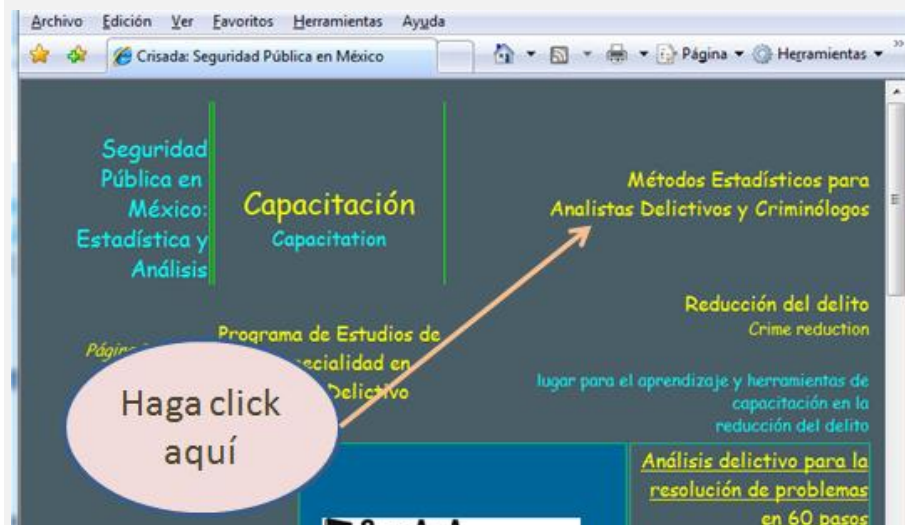
Si observamos la tabla con atención encontraremos que el mayor número de víctimas se da en el grupo de edad de 20 a 24 años, seguido por el de 25 a 29 años, etc.

La tabla de ejercicios para este apartado se encuentra en Internet en:

<http://www.seguridadpublicaenmexico.org.mx/cursos.htm>

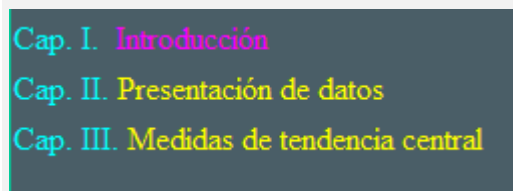
En el apartado de capacitación en la sección de Métodos Estadísticos para Analistas Delictivos y Criminólogos,

Imagen 2. Pantalla para selección



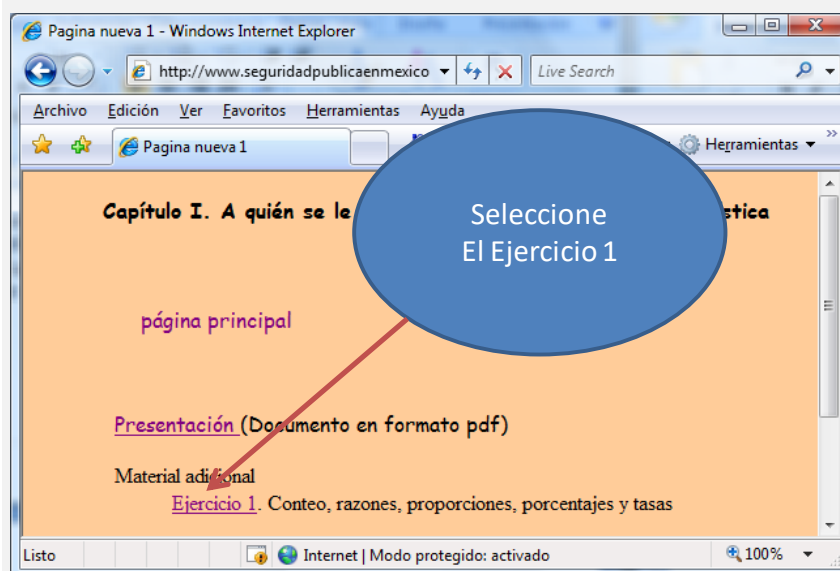
En la hoja resultante seleccione Cap. I. Introducción

Imagen 3. Selección del tema que contiene al capítulo 1



Y finalmente, seleccione el Ejercicio 1.

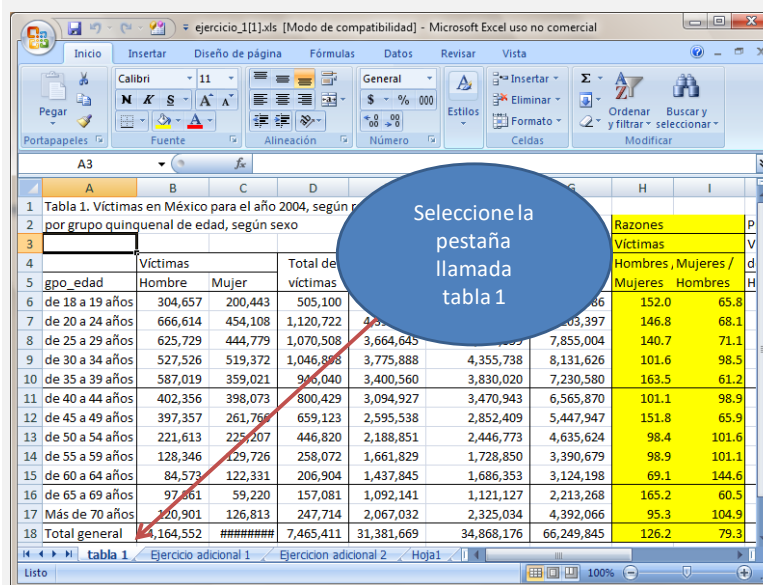
Imagen 4. Selección del ejercicio 1



Se abre un archivo en Excel, como se muestra en la siguiente imagen

En caso que exista un password, solicítelo vía correo electrónico a aarango54@hotmail.com

Imagen 5. Selección del libro que contiene a la tabla 1



En este archivo se encuentra replicada la tabla 1, Víctimas en México para el año 2004, según resultados de la ENSI- 3 por sexo y grupo quinquenal de edad.

La forma más elemental para presentar información es como se muestra en las columnas B, C y D, que muestran el conteo o la frecuencia de las víctimas en México. Cuando la información está en esta forma simple se dice que ésta está en datos brutos. Un conteo o frecuencia sólo nos indica el número de veces que un evento ocurre en sus datos. La tabla 1, presenta el número de víctimas hombres, mujeres o total de víctimas por grupo quinquenal de edad, según lo reportado por la ENSI-3

Estos números nos señalan exactamente cuántas víctimas de cada sexo se encontraron mediante la mencionada encuesta.

Sin embargo, estos datos brutos o conteo simple pueden orientarnos de manera inadecuada.

Así, por ejemplo, las víctimas en el grupo de edad de 20 a 24 años, superan en un poco más del doble a las víctimas en el grupo de edad de entre 18 a 19 años.

Imagen 6. Víctimas por grupo de edad

gpo_edad	Víctimas	
	Hombre	Mujer
de 18 a 19 años	304,657	200,443
de 20 a 24 años	666,614	454,108

Las limitaciones fundamentales que presenta el conteo simple o la presentación en datos brutos, es que no indica cuantas víctimas

ocurrieron en relación con el tamaño de la población en esos grupos de edad.

Si se divide a cada dato entre la población por sexo, en esos grupos de edad, encontramos que la población que se encuentra entre los 18 a 19 años es más vulnerable que la que se encuentra entre los 20 a los 24 años.

Imagen 7. Algunos cálculos

gpo_edad	Hombre	Mujer	Vh / Ph *100	Vm / Pm *100
de 18 a 19 años	304,657	200,443	15,197	9,754
de 20 a 24 años	666,614	454,108	15,158	9,449

Donde Vh = Víctimas hombres

Ph = Población hombres

Vm = Víctimas mujeres

Pm = Población mujeres

Un grupo de edad puede tener más víctimas debido a que en dicho grupo hay más población.

El conteo simple puede estar bien si no se esta interesado en hacer comparaciones entre los grupos o el tiempo, o incluso regiones. La forma correcta de hacer comparaciones entre unidades con diferentes valores poblacionales, es a través de las tasas.

Una tasa se encuentra dividiendo el número observado entre la población que teóricamente puedes ser observada al interior de la población de interés.

Así, para el grupo de 18 a 19 años, tenemos que

$$\frac{\text{Víctimas hombres}}{\text{Población de hombres}} = \frac{306457}{2004680} * 100000 = 15197$$

Mientras que para el grupo de 20 a 24 años

$$\frac{\text{Víctimas hombres}}{\text{Población de hombres}} = \frac{666614}{439733} * 100000 = 15158$$

De donde resulta que el grupo de hombres entre los 18 a 19 años es más vulnerable que el grupo de hombres entre los 20 a 24 años, ya que tienen un mayor número de víctimas

15197 por 100 mil habitantes en el primer grupo, vs. 15158 por 100 mil en el segundo grupo.

En general siempre debe intentar compararse contra la población total teórica de interés. Así, si se habla de robo a casa habitación debe compararse (obtenerse una tasa) contra el número total de casas habitación. En robo de vehículo contra el total de vehículos

en circulación (incluso, por marca y modelo, si ello es posible). En robo a negocio, contra el total de negocios, separando farmacias, zapaterías, bancos, etc.

Realice los cálculos en las hojas “Ejercicio adicional 1” y “Ejercicio adicional 2” Puede elaborar preguntas del estilo

En datos brutos

¿Cuál es la entidad donde hubo más víctimas hombres?

¿Cuál es la entidad donde hubo menos víctimas mujeres?

Considerando el total de víctimas, ¿en que entidad hubo más víctimas?

En tasas

¿Cuál es la entidad que tuvo más víctimas hombres por cada 100 mil habitantes?

Estas preguntas sólo sirven como ejemplo, pudiendo trabajar con todas aquellas de su interés.

Razones

Una razón es la relación que existe entre dos elementos o conjuntos. Por ejemplo en la Tercera Encuesta Nacional sobre Inseguridad ENSI- 3 realizada por el ICESI, se encontró que el número de víctimas en México, en el año 2004, según el sexo es el mostrado en la siguiente tabla:

Tabla 5. Víctimas en México para el año 2004, según resultados de la ENSI- 3 por sexo y grupo quinquenal de edad

Grupo de edad	Víctimas		Total Víctimas	Población en los grupos de edad		Total Población en los grupos de edad
	Hombre	Mujer		Hombre	Mujer	
de 18 a 19 años	304,657	200,443	505,100	2,004,680	2,054,906	4,059,586
de 20 a 24 años	666,614	454,108	1,120,722	4,397,733	4,805,664	9,203,397
de 25 a 29 años	625,729	444,779	1,070,508	3,664,645	4,190,359	7,855,004
de 30 a 34 años	527,526	519,372	1,046,898	3,775,888	4,355,738	8,131,626
de 35 a 39 años	587,019	359,021	946,040	3,400,560	3,830,020	7,230,580
de 40 a 44 años	402,356	398,073	800,429	3,094,927	3,470,943	6,565,870
de 45 a 49 años	397,357	261,766	659,123	2,595,538	2,852,409	5,447,947
de 50 a 54 años	221,613	225,207	446,820	2,188,851	2,446,773	4,635,624
de 55 a 59 años	128,346	129,726	258,072	1,661,829	1,728,850	3,390,679
de 60 a 64 años	84,573	122,331	206,904	1,437,845	1,686,353	3,124,198
de 65 a 69 años	97,861	59,220	157,081	1,092,141	1,121,127	2,213,268
Más de 70 años	120,901	126,813	247,714	2,067,032	2,325,034	4,392,066
Total general	4,164,552	3,300,859	7,465,411	31,381,669	34,868,176	66,249,845

Fuente: elaboración propia con base en el análisis de resultados de la ENSI- 3 por grupo de edad y sexo.

De donde:

n1 = Víctimas mujeres 3 300 859

n2 = Víctimas hombres 4 164 552

N = Total de víctimas 7 465 411

Realizando la operación de razón

$$\frac{Mujeres}{Hombres} = \frac{3,300,859}{4,164,552} = 0.79$$

Es costumbre en estadística delictiva multiplicar el resultado por 100, por lo que:

$$0.79 * 100 = 79.26$$

Lo cual significa que existen 79.26 víctimas mujeres por cada 100 víctimas del sexo masculino. Si procedemos de forma inversa. La relación de hombres a mujeres será:

$$\frac{Hombres}{Mujeres} = \frac{4,164,552}{3,300,859} = 1.26$$

Multiplicando el resultado por 100

$$1.26 * 100 = 126.17$$

El resultado final indica que existen 126.17 víctimas hombres por cada 100 víctimas del sexo femenino.

Proporciones La proporción es la relación que se establece entre una parte con respecto al todo. En las proporciones se omite multiplicar el cociente por cien, dando por resultado que la relación se establece con respecto a la unidad.

Partiendo del ejemplo inicial:

n1 = Víctimas mujeres 3 300 859

n2 = Víctimas hombres 4 164 552

N = Total de víctimas 7 465 411

La proporción de hombres es

$$\frac{4,164,552}{7,465,411} = 0.56$$

La proporción de mujeres es:

$$\frac{3,300,859}{7,465,411} = 0.44$$

Por formula:

$$\frac{n_1}{N} = P_1, \quad \frac{n_2}{N} = P_2$$

$$P_1 + P_2 = 1$$

Donde n1 = valor 1,

n2 = valor 2,

P1 = porcentaje 1, y

P2 = porcentaje 2

Las proporciones son utilizadas frecuentemente en los cálculos estadísticos, pero en la presentación final de los datos o cifras son más comunes los porcentajes. Los porcentajes y las proporciones indican la misma información, medida en dos formas diferentes. Los primeros son de mayor utilización y por lo mismo son de más fácil comprensión.

Porcentajes: Representan una relación que se establece de una de las partes con respecto al todo o total multiplicado por cien.

Siguiendo el ejemplo inicial:

n1 = Víctimas mujeres 3 300 859

n2 = Víctimas hombres 4 164 552

N = Total de víctimas 7 465 411

La proporción de víctimas hombres es

$$\frac{4,164,552}{7,465,411} = 0.56$$

Multiplicando por 100

$$0.56 * 100 = 56\%$$

La proporción de víctimas mujeres es:

$$\frac{3,300,859}{7,465,411} = 0.44$$

Multiplicando por 100

$$0.44 * 100 = 44\%$$

Por fórmula:

$$Porcentaje = \left(\frac{n_1}{N} + \frac{n_2}{N} \right) 100 = 100$$

$$Porcentaje = \left(\frac{n_1}{N} \right) 100 + \left(\frac{n_2}{N} \right) 100 = 100$$

El principal papel de los porcentajes es el de obtener comparaciones, es decir, con las cifras absolutas no es muy recomendable, en virtud de que, al ser menores a 1, no permiten identificar las relaciones. La comparación en este caso es posible porque los números absolutos se reducen a una escala que es fácil de multiplicar y dividir (escala de 100), lo cual permite la determinación de su magnitud relativa.

Es importante señalar que la suma de todos los porcentajes debe ser igual 100

Finalmente, los porcentajes, son especialmente útiles cuando se realizan comparaciones entre dos o más conjuntos de variables.

Tasas:

También son conocidas como coeficientes o números índices. Una tasa o índice, es similar al porcentaje, el numerador indica el número de veces que un suceso específico ocurre durante un particular periodo de tiempo y el denominador es el número de veces que es posible que ocurra el suceso.

$$tasa = \frac{\text{número de ocurrencias actuales de un evento}}{\text{número de ocurrencias posibles}}$$

A este respecto los coeficientes más utilizados en la estadística delictiva son las tasas de victimización. Existen otras tasas como son: mortalidad general, natalidad general o específica, delincuencia, fertilidad general o específica, divorcios, etc. Por lo regular las tasas se multiplican por algún número, usualmente se multiplican por mil, diez mil y cien mil, a fin de expresarlas como el número de veces que el suceso del numerador acontece en cada mil, diez mil u otras cifra según sea el caso.

Ejemplo:

Partiendo del ejemplo inicial:

n1 = Total p1 de víctimas mujeres 3 300 859,

p1 = población total de Mujeres 34 868 176

n2 = Total de víctimas hombres 4 164 552,

p2 = población total de Hombres 31 381 669

Por lo que

$$\frac{\text{Víctimas hombres}}{\text{Población total de hombres}} = \frac{4,164,552}{31,381,669} = 0.1327$$

Multiplicando por 100 mil

$$0.1327 * 100\,000 = 13270.7$$

$$\frac{\text{Víctimas mujeres}}{\text{Población total de mujeres}} = \frac{3,300,859}{34,868,176} = 0.0947$$

Multiplicando por 100 mil

$$0.0947 * 100\,000 = 9466.7$$

Es decir que la tasa de hombres es de 13 270.7 víctimas hombres por cada 100 mil hombres, mientras que la tasa de mujeres es de 9 466.7 víctimas mujeres por cada 100 mil mujeres.

Determine para cada grupo de edad

- a) La razón de víctimas hombres y mujeres y viceversa.
- b) La proporción de víctimas hombres, así como de mujeres
- c) El porcentaje de víctimas mujeres y hombres y construya una tabla de porcentajes.
- d) Las tasas de víctimas hombres y mujeres

Tasas de incremento (crecimiento)

Otro tipo de operaciones que se usa con regularidad en la descripción de datos son las tasas de Incremento porcentual y Promedio de incremento.

Incremento (crecimiento) porcentual

El incremento porcentual se usa para resolver problemas como el siguiente: En una localidad hubo 1302 personas registradas como delincuentes en 2007 y en el 2008 ya eran 1894. ¿Cuál fue el incremento porcentual? La fórmula correspondiente es la siguiente:

$$IP = \frac{P_1 - P_0}{P_0} (100)$$

Donde:

IP = incremento porcentual

P_1 = población última (en el año 2008)

P_0 = población anterior (en el año 2007)

Sustituyendo la fórmula:

$$IP = \frac{1894 - 1302}{1302} (100) = 45.47\%$$

Entonces, se puede afirmar que la tasa de incremento porcentual de una localidad equis en el período que va de 2007 a 2008, fue de 45.47%.

Promedio de incremento (crecimiento)

En caso de que lo que interesara saber no fuera el período completo, sino el promedio de incremento por año, entonces se usaría la siguiente fórmula:

$$PI = \frac{P_1 - P_0}{P_1 + P_0} \frac{2}{N} (100)$$

Donde:

PI = promedio de incremento

P₁ = población última

P₀ = población anterior

N = número de períodos (pueden ser años, semestres, semanas, etc.)

Sustituyendo la fórmula:

$$PI = \frac{1894 - 1302}{1894 + 1302} \frac{2}{1} (100) = \frac{592}{3196} (2)(100) = (0.1852)(200) = 37.05\%$$

Así, el promedio de incremento anual de la población descrita fue de 37.05%.

Actividades

En Internet, en el archivo ejercicios capitulo 1.xls vea la hoja ejercicios adicional 1 y ejercicios adicional 2

Ejercicio 6

Considerando la información de la tabla siguiente

Tabla 6. Delito de robo a transeúnte registrado por la PGJ DF para las delegaciones de Iztapalapa y Tlalpan en los años 1997^a 2006.

AÑO	Iztapalapa	Tlalpan
1997	3436	1419
1998	3721	1178
1999	4458	1969
2000	2761	846
2001	2220	714
2002	2250	677
2003	2759	742
2004	2781	756
2005	2786	612
2006	2715	549

El incremento porcentual del año 2003, con respecto del año 2002, para Iztapalapa es el que se muestra

$$\text{Iztapalapa} \quad IP = \frac{2759 - 2250}{2250} (100) = 22.62\%$$

El promedio de los años 1997 a 2001 para Tlalpan es

$$PI = \frac{714 - 1419}{714 + 1419} \frac{2}{4} (100) = \frac{-705}{2133} (0.5)(100) = (-0.331)(50) = -16.53\%$$

El robo a transeúnte registrado en Tlalpan ha disminuido en 5 años en promedio en 13.22%

Calcule

1. Cuantos robos se cometían en Iztapalapa por cada 100 robos a transeúnte de Tlalpan, para los años 2000 y 2006.
2. Cuantos en Tlalpan por cada 100 de Iztapalapa, para los años 2004 y 2006.
3. La tasa de crecimiento porcentual, para cada año para las dos delegaciones
4. La tasa de crecimiento promedio, para las dos delegaciones.

[illegible]

Diferentes fases

Descripción General

Antes de intentar realizar cualquier tipo de análisis estadístico propiamente dicho, es muy importante entender la distribución y las características que posee la información sobre la delincuencia, que tendencia presentan, que forma es la que describe mejor este tipo de información.

En este capítulo nos limitaremos a conocer los distintos tipos de gráficos que pueden realizarse y aprenderemos a utilizar el programa de Microsoft Excel®, el cual es una hoja electrónica muy útil para el manejo de grandes volúmenes de información, para la realización de los diferentes tipos de gráficos que se utilizan con mayor frecuencia en la estadística, y aunque los métodos gráficos aquí revisados pueden utilizarse en una gran cantidad de situaciones y aplicarse a cualquier problema nosotros haremos énfasis en las aplicaciones delincuenciales y de criminalidad.

La gran mayoría de las personas tienen dificultades para interpretar la información cuando esta corresponde a un gran conjunto de datos. Una solución sencilla que nos ayuda con este problema es presentar a dicho conjunto de datos en forma gráfica. Ya lo dice el conocido refrán “una imagen dice más que mil palabras”.

Para el caso de la representación de información sobre el delito a través de gráficas e imágenes estas deben realizarse en forma simple, sin demasiado adorno, ya que el objetivo no es hacer gráficos “bonitos” sino comunicar algo; es decir que las tablas, figuras y gráficas son herramientas efectivas para mostrar información, en la medida en que estas mantienen un diseño sencillo y no se “sobrecargan” con adornos superfluos.

Cuando se realiza cualquier estudio estadístico, el investigador o analista debe recolectar información de la variable o variables que se encuentran bajo estudio. Así si se desea estudiar al número de personas que han sido víctimas del delito de robo a transeúnte en un área geográfica en particular, por ejemplo, el centro histórico de la Ciudad de México, durante un periodo de tiempo determinado, por ejemplo, el mes de enero; se debe buscar información de diversos tipos que va desde aquella que se encuentra en las encuestas sobre victimización hasta la que se encuentra en los centros de atención y servicio de emergencia, como el 066, la que se encuentra en los partes de novedades de la policía preventiva y la que se encuentra en poder de los agentes del Ministerio Público en las averiguaciones previas iniciadas.

Así, para describir la situación que guarda este problema, realizar conclusiones u obtener inferencias acerca de este delito y sus víctimas el analista debe organizar la información de tal forma que sea entendible y arroje sentido. La forma más conveniente para organizar la información es mediante la construcción de distribuciones de frecuencia.

Después de organizar los datos, el analista debe presentarlos de tal forma que sean fácilmente entendibles por aquellos que se verán beneficiados de la lectura del estudio. La forma más sencilla para hacer esto es a través de la construcción de gráficos estadísticos. Existe una gran cantidad de éstos y cada uno tiene un propósito específico.

En este apartado se explicará cómo organizar los datos mediante las distribuciones de frecuencia y como presentar los datos a través de la construcción de gráficos. En particular se revisará la forma para construir histogramas, polígonos de frecuencia, ojivas, gráficas de pastel, gráficos de Pareto y gráficas de series de tiempo.

Organización de los datos

Suponga que el analista después de investigar sobre el delito de robo a transeúnte en el Centro Histórico en los diferentes puntos de concentración de información obtiene la siguiente información de las víctimas, por edad, que padecieron dicho delito en un día en particular.

Tabla 7. Edades de las víctimas del delito de robo a transeúnte en el centro histórico, en un día en particular

56	26	57	49	37	34	62
61	34	37	49	43	15	61
40	31	26	28	48	43	43
22	22	23	41	50	44	33
22	33	35	51	46	20	28
50	43	49	34	55	62	15
30	47	17	29	63	31	32
38	64	24	44	22	62	40
58	21	28	31	63	44	55
62	45	28	15	56	25	65

Cuando la información se encuentra en su forma original se conoce como datos brutos. En virtud que sobre la información en datos brutos no se puede obtener ningún tipo de análisis, el analista organiza ésta construyendo una distribución de frecuencias.

La frecuencia es el número de valores en una clase específica de la distribución. Para la información de las edades de las víctimas, la distribución de frecuencias se muestra a continuación:

Tabla 8. Distribución de frecuencias para las edades de las víctimas en el centro histórico mostradas en la tabla 2.

Nº. de Clases:		10	punto	frecuencia		
Lim. inferior	Lim. superior		medio	frecuencia	acumulada	% acumulado
15.0	20.0		17.5	5	5	7.14%
21.0	25.0		23.0	8	13	18.57%
26.0	30.0		28.0	8	21	30.00%
31.0	35.0		33.0	10	31	44.29%
36.0	40.0		38.0	5	36	51.43%
41.0	45.0		43.0	9	45	64.29%
46.0	50.0		48.0	8	53	75.71%
51.0	55.0		53.0	3	56	80.00%
56.0	60.0		58.0	4	60	85.71%
61.0	65.0		63.0	10	70	100.00%
66.0						
# de datos				70		

Una vez que la información se encuentra ordenada se pueden realizar algunas observaciones generales sobre las víctimas de robo a transeúnte. Esta tabla nos indica que hubo diez personas entre 31 años y hasta 35 que sufrieron el delito de robo, 8 tenían entre 26 y 30 años. Asimismo hubo 4 víctimas que se encontraban en el rango de edad de entre 56 a 60 años, etc.

Las clases en el ejemplo anterior son 15- 20, 21- 25, 26- 30, etc. Estos valores se conocen como límites de clase. Así las edades de cada víctima se van enumerando y colocando en cada clase según corresponda.

Como construir una distribución de Frecuencias

Para construir una distribución de frecuencias,

Paso 1.

Encuentre el valor máximo y el valor mínimo de los datos

Valor máximo = 65 años

Valor mínimo = 15 años

Para ello, puede recorrer por filas primero buscando el valor máximo de cada fila y coloque el resultado a la derecha de la fila. Haga el mismo recorrido por filas, pero ahora buscando el valor mínimo. De las columnas resultantes, seleccione el valor máximo y el valor mínimo

Tabla 9. Edades de las víctimas del delito de robo a transeúnte. Como selecciona valores máximos y mínimos.

						Valor máximo	Valor mínimo
56	26	57	49	37	34	62	26
61	34	37	49	43	15	61	15
40	31	26	28	48	43	43	48
22	22	23	41	50	44	33	50
22	33	35	51	46	20	28	51
50	43	49	34	55	62	15	62
30	47	17	29	63	31	32	63
38	64	24	44	22	62	40	64
58	21	28	31	63	44	55	63
62	45	28	15	56	25	65	65

Paso 2.

Encuentre el rango. Este se obtiene restando el valor mínimo del valor máximo.

$$\text{Rango} = \text{Valor máximo} - \text{Valor mínimo} = 65 - 15 = 50$$

Paso 3.

Decida cuantas clases va a generar. En este paso el analista debe seleccionar el número de clases en las cuáles debe mostrar la información. En nuestro ejemplo se seleccionó mostrar la información en 10 clases.

Paso 4.

El rango permite obtener la anchura de clase. Para ello divida el rango entre el número de clases deseado.

$$\text{Anchura de clase} = \frac{\text{Rango}}{\text{Número de clases deseado}} = \frac{50}{10} = 5$$

Paso 5.

Ahora se obtienen los límites de clase,

Para obtener los límites de clase de la clase 1.

El límite inferior de la clase 1, se obtiene colocando el valor mínimo como límite inferior de la clase 1, y se le suma la anchura de clase, con lo que se obtiene el límite superior de la clase 1.

Clase	Límite inferior	Anchura de clase	Límite superior
1	15	5	20

Para obtener los límites de clase de la clase 2 y subsiguientes.

El límite inferior de la clase 2, se obtiene colocando el límite superior de la clase 1, como límite inferior de la clase 2, y se le suma un valor adicional. Esta operación se hace para garantizar que no haya duda respecto de a que clase debe pertenecer una

observación, ya que si se coloca el valor 20 como límite inferior de la clase 2, sin sumar un valor adicional, una observación que tuviere ese valor podría quedar tanto en la clase 1 como en la clase 2.

Es convención de que cuando se trata de números enteros, se le suma un valor 1 al valor que va a formar el límite inferior de la clase 2 y subsiguientes, con objeto de evitar duplicidad o duda de en que clase debe caer una observación. Si contiene decimales, se le suma al valor mínimo la cantidad de 0.1, si contiene centésimas se le suma un valor 0.01, si contiene milésimas se le suma al valor mínimo la cantidad de 0.001 y así sucesivamente.

El límite superior de la clase 2 se obtiene sumando la anchura de clase al límite superior de la clase 1.

Así los límites de clase quedan como sigue:

Tabla 10. Límites de clase. Como calcularlos.

Clase	Límite inferior	Anchura de clase	Límite superior	Valor a sumar al límite inferior
1	15	5	20	
2	$20 + 1 = 21$	5	$20 + 5 = 25$	1
3	$25 + 1 = 26$	5	$25 + 5 = 30$	1
4	$30 + 1 = 31$	5	$30 + 5 = 35$	1
5	$35 + 1 = 36$	5	$35 + 5 = 40$	1
6	$40 + 1 = 41$	5	$40 + 5 = 45$	1
7	$45 + 1 = 46$	5	$45 + 5 = 50$	1
8	$50 + 1 = 51$	5	$50 + 5 = 55$	1
9	$55 + 1 = 56$	5	$55 + 5 = 60$	1
10	$60 + 1 = 61$	5	$60 + 5 = 65$	1

Paso 6.

En este paso se hace un conteo de cada observación y se coloca en la clase correspondiente. Por ejemplo para las observaciones en la clase 1, se buscan las edades observadas en el rango de 15 a 20 años.

Tabla 11. Ejemplo de conteo de observaciones para saber la frecuencia de cada clase.

56	26	57	49	37	34	62
61	34	37	49	43	15	61
40	31	26	28	48	43	43
22	22	23	41	50	44	33
22	33	35	51	46	20	28
50	43	49	34	55	62	15
30	47	17	29	63	31	32
38	64	24	44	22	62	40
58	21	28	31	63	44	55
62	45	28	15	56	25	65

Tabla 12. Conteo para obtener la frecuencia de cada clase.

Nº. de Clases:		10	punto	
Lim. inferior	Lim. superior		medio	frecuencia
15.0	20.0	17.5		5
21.0	25.0	23.0	III	8
26.0	30.0	28.0	III	8
31.0	35.0	33.0		10
36.0	40.0	38.0		5
41.0	45.0	43.0		9
46.0	50.0	48.0	III	8
51.0	55.0	53.0		3
56.0	60.0	58.0		4
61.0	65.0	63.0		10

Para decidir cuantas clases utilizar y la anchura de clase, durante la construcción de una distribución de frecuencias se deben seguir las siguientes reglas:

Deben existir entre 5 a 10 clases. No es un error tener menos de 5 o más de 10 clases, sin embargo la mayoría de los que realizan análisis delictivo están de acuerdo con esta convención.

Si no se utiliza computadora, debe buscarse que el punto medio sea un número non. El punto medio, también conocido como marca de clase se obtiene sumando los límites inferior y superior de cada clase y dividiendo entre 2

$$\text{punto medio} = \frac{\text{límite inferior} + \text{límite superior}}{2}$$

$$\text{punto medio}_{\text{clase 1}} = \frac{15 + 20}{2} = 17.5$$

Las clases deben ser mutuamente exclusivas; esto implica que no exista duda; es decir que no haya traslape entre los límites de clase, de tal forma que una observación no pueda ser colocada en dos clases a la vez.

Tabla 13. Ejemplo de clases no mutuamente exclusivas

	Limite inferior	límite superior
1	15	20
2	20	25
3	25	30

Con estos límites, en que clase se colocará una observación cuando tenga el valor de 20, en la clase 1 o en la clase 2. Es decir que no debe existir duda de en qué clase debe colocarse cada observación

Las clases deben ser continuas. Aun cuando no existan valores en una clase, esta debe aparecer. La única excepción es si el valor cero esta antes de la primera clase o en la última.

En nuestro ejemplo, aunque existen personas que tienen 14 años o menos, en virtud que hubo cero observaciones (no existieron personas en esas edades que hayan sido víctimas del delito de robo a transeúnte) no existe una clase para señalar ese hecho, se sobre entiende que el valor de las observaciones empezó en la edad de 15 años. Lo mismo sucede para las personas de 66 años o más, ya que la edad máxima observada fue de 65 años.

Las clases deben ser mutuamente exclusivas

Las clases deben ser iguales en anchura.

Paso 7.

Se acostumbra obtener la frecuencia y la frecuencia acumulada, así como los porcentajes sobre estos valores.

Tabla 14. Obtención de frecuencia y frecuencia acumulada

Nº. de Clases:		10	punto	frecuencia		
Lim. inferior	Lim. superior	medio	frecuencia	acumulada	% frecuencia	% acumulada
15	20	17.5	5	5	7.1%	7.1%
21	25	23	8	13	11.4%	18.6%
26	30	28	8	21	11.4%	30.0%
31	35	33	10	31	14.3%	44.3%
36	40	38	5	36	7.1%	51.4%
41	45	43	9	45	12.9%	64.3%
46	50	48	8	53	11.4%	75.7%
51	55	53	3	56	4.3%	80.0%
56	60	58	4	60	5.7%	85.7%
61	65	63	10	70	14.3%	100.0%
# de datos			70			

La frecuencia es la cantidad de veces que una observación cae entre las clases. La frecuencia para la clase 1 es 5, en virtud que hubo 5 observaciones (5 personas entre las edades de 15 a 20

años fueron víctimas del delito de robo a transeúnte), para la clase 3, la frecuencia fue de 8.

La frecuencia acumulada para la clase 1 es de 5, es decir que hasta esa clase se han considerado a 5 observaciones, para la clase dos es de 13 (5 de la clase 1 más las 8 de la clase 2).

El porcentaje de frecuencia para la clase 1 resulta de dividir el número de observaciones de dicha clase entre el total de observaciones

$$\% \text{ frecuencia}_{\text{clase 1}} = \frac{\text{Numero de observaciones en la clase 1}}{\text{total de observaciones}} * 100$$

$$\% \text{ frecuencia acumulada}_{\text{clase 1}} = \frac{5}{70} * 100 = 7.1$$

Para la clase 2

$$\% \text{ frecuencia acumulada}_{\text{clase 2}} = \frac{8}{70} * 100 = 11.4$$

Y así sucesivamente

Finalmente se obtiene la frecuencia acumulada que resulta de dividir cada valor acumulado obtenido para cada clase entre el total de observaciones

$$\% \text{ frecuencia acumulada}_{\text{clase 1}} = \frac{\text{Numero de observaciones en la clase 1}}{\text{total de observaciones}} * 100$$

$$\% \text{ frecuencia acumulada}_{\text{clase 1}} = \frac{5}{70} * 100 = 7.1$$

Para la clase 2

$$\% \text{ frecuencia acumulada}_{\text{clase 2}} = \frac{13}{70} * 100 = 18.6$$

Para la clase 3

$$\% \text{ frecuencia acumulada}_{\text{clase 3}} = \frac{21}{70} * 100 = 30.0$$

Ejercicio 7

Obtenga la distribución de frecuencias para las edades de las víctimas del delito de robo de vehículo, que se muestran a continuación:

53	32	61	23	31
54	31	20	50	56
37	60	19	26	61
58	36	32	49	23
56	61	29	33	24
53	28	59	28	29
55	28	56	37	29
46	44	37	30	49

Se presentan algunos cálculos para ayudar en su ejercicio

Tabla 15. Cálculos para obtener la distribución de frecuencias.

Nº. de Clases:	10	punto			frecuencia		
Lim. inferior	Lim. superior	medio	Cuenta	frecuencia	acumulada	% frecuencia	% acumulado
15.0	19.5	17.3	II	2	2	5.7%	5.71%
19.6	24.0	21.8			6		17.14%
24.1	28.5				10		28.57%
28.6	33.0			7			
33.1	37.5						57.14%
37.6	42.0					0.0%	
42.1	46.5	44.3					57.14%
46.6	51.0				22	5.7%	
51.1	55.5	53.3	IIII	4			
55.6	60.0	57.8		9		25.7%	
60.1							
# de datos				35			

Gráficas de pastel:

La forma más simple de representar a nuestra información de forma gráfica es mediante las gráficas de pastel o “pie”. La tabla 2.1 muestra a las denuncias ante las agencias del ministerio público del fuero común para los delitos patrimoniales del año 2006, mientras que la tabla 2. Muestra las denuncias ante las agencias del ministerio público del fuero común para los delitos de robo del año 2006

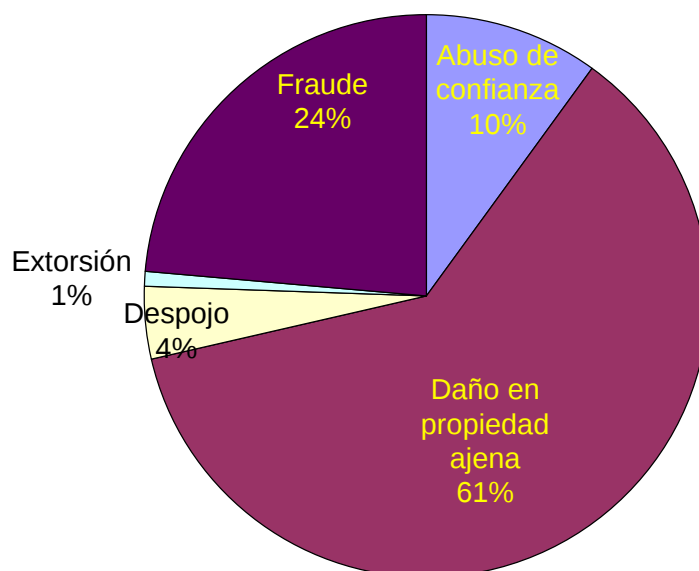
Tabla 16. México: denuncias ante las agencias del ministerio público del fuero común por delitos patrimoniales, según modalidad, para el año 2006.

Delito	Modalidad	denuncias	%
Patrimonial	Abuso de confianza	226,544	9.9%
	Daño en propiedad ajena	1,403,078	61.5%
	Despojo	95,142	4.2%
	Extorsión	19,835	0.9%
	Fraude	537,389	23.5%
Total Patrimonial		2,281,988	100.0%

Fuente: Arturo Arango y Cristina Lara, "Denuncias ante agencias del ministerio público del fuero común: 1997 a 2006" En "Estadísticas de Seguridad Pública en México". Elaborado con base en información del Sistema Nacional de Seguridad Pública. (Septiembre 2007). CrisAdA Editorial: Seguridad Pública en México. Tabla Arango y Lara. www.seguridadpublicaenmexico.org.mx

Para la mayoría de las audiencias y propósitos el sólo mostrar la tabla 1, puede ser suficiente, sin embargo esta información puede expresarse en forma alternativa utilizando una gráfica de pastel. Normalmente los datos en la gráfica de pastel se representan en porcentaje.

Gráfica 1. Denuncias ante las agencias del ministerio público del fuero común por delitos patrimoniales, según modalidad, para el año 2006



Aunque el número de categorías que puede representarse en una gráfica de este tipo es muy amplio, la gráfica se vuelve difícil de leer, cuando éstas son muchas; por ello se recomienda que en este tipo de gráficas se represente a un máximo de 5 clases o apartados.

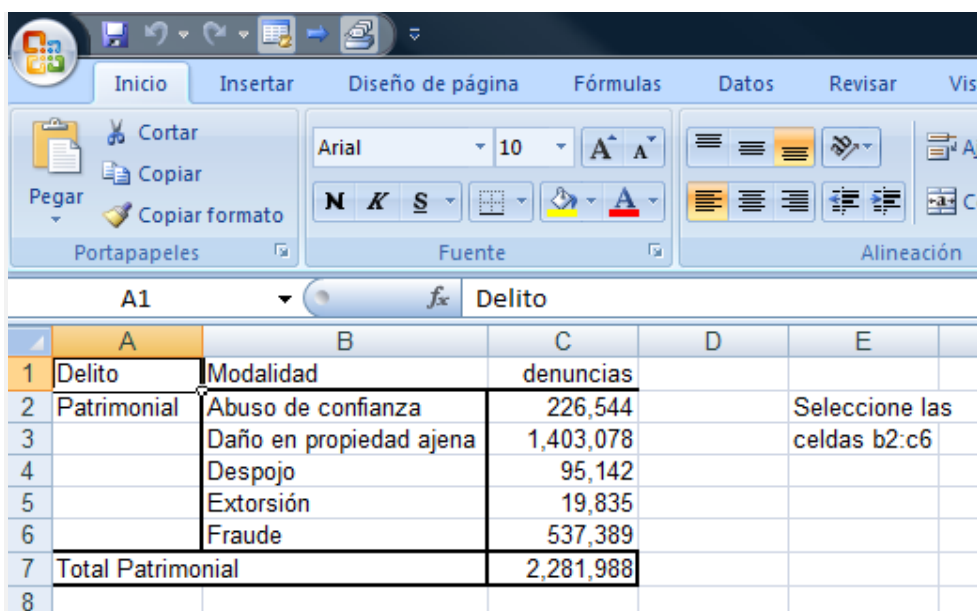
Paso a Paso con Excel: gráficas de pastel

Ayuda para el
Maestro

Capitulo 2.
Tabla 1.xls

Para generar las gráficas de pastel en Excel, sólo teclee los datos en una hoja nueva y seleccione la información que va a graficar.

Imagen 8. Datos para gráfico de pastel

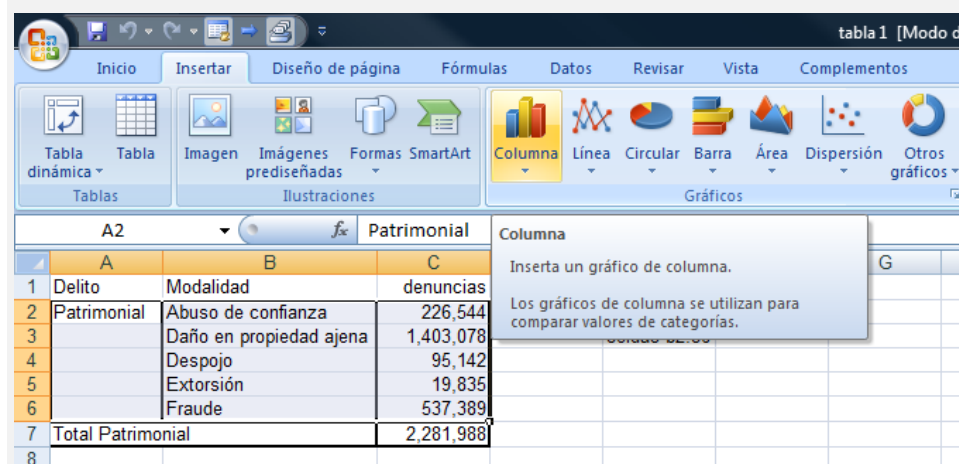


The screenshot shows the Excel interface with the 'Inicio' (Home) tab selected. The data table is as follows:

	A	B	C	D	E
1	Delito	Modalidad	denuncias		
2	Patrimonial	Abuso de confianza	226,544		Seleccione las
3		Daño en propiedad ajena	1,403,078		celdas b2:c6
4		Despojo	95,142		
5		Extorsión	19,835		
6		Fraude	537,389		
7	Total Patrimonial		2,281,988		
8					

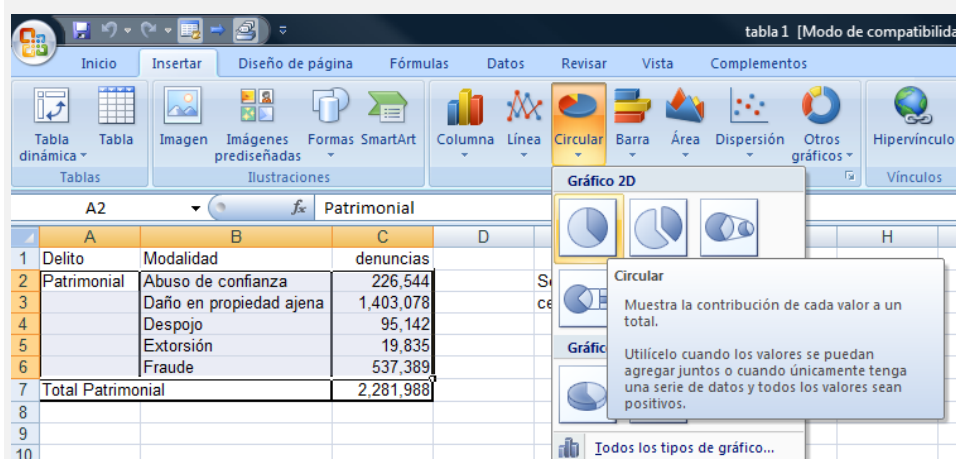
Seleccione el asistente para gráficos como se muestra en la imagen.

Imagen 9. Selección de gráficos



En el menú emergente seleccione graficas circular.

Imagen 10. Comando para selección de gráfico circular

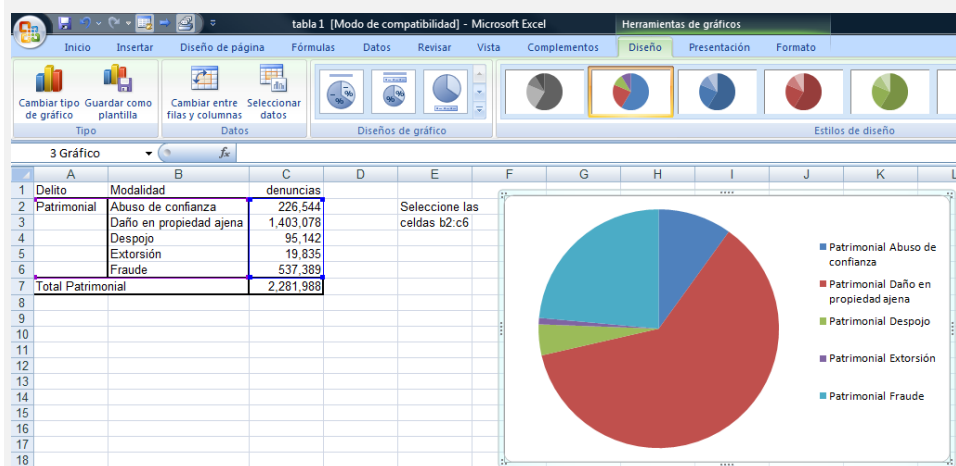


Seleccione finalizar.

Se crea un gráfico en la misma hoja donde encuentran los datos, como un gráfico incrustado. Este gráfico puede modificarse colocando el cursor sobre él.

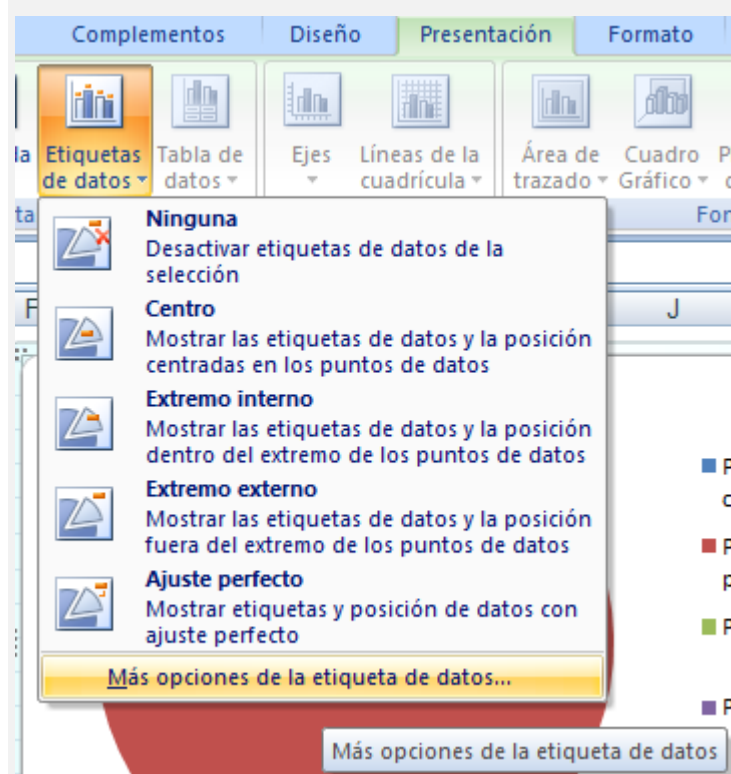
Cuando coloca el apuntador del ratón sobre el gráfico y hace clic para seleccionarlo, aparecen en la parte superior las herramientas de gráficos Diseño, Presentación y Formato. Puede utilizar los comandos que aparecen en estas herramientas para modificar el gráfico hasta que se muestre como desea.

Imagen 11. Gráfico circular



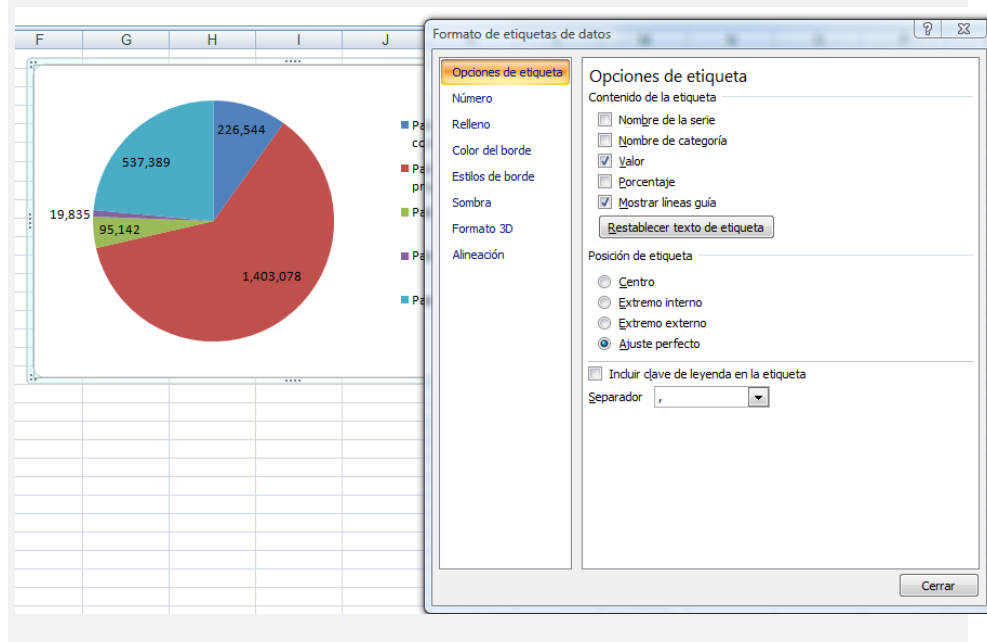
Ahora, en la herramienta Presentación, seleccione Etiquetas de datos, más opciones.

Imagen 12. Comando para agregar etiquetas a un gráfico



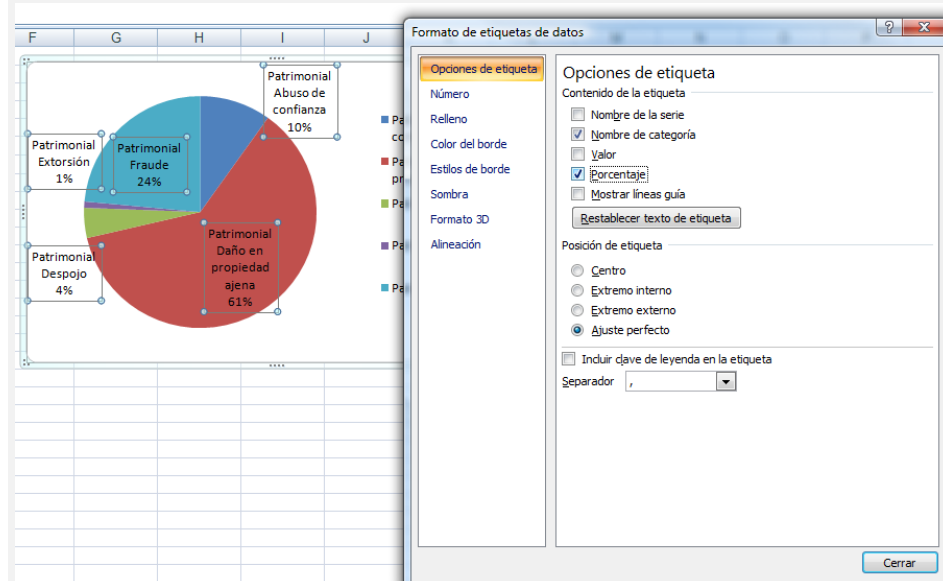
En forma automática se colocan los valores de los datos en donde corresponde, asimismo aparece el menú emergente de Formato de etiquetas de datos.

Imagen 13. Como dar formato a las etiquetas de datos



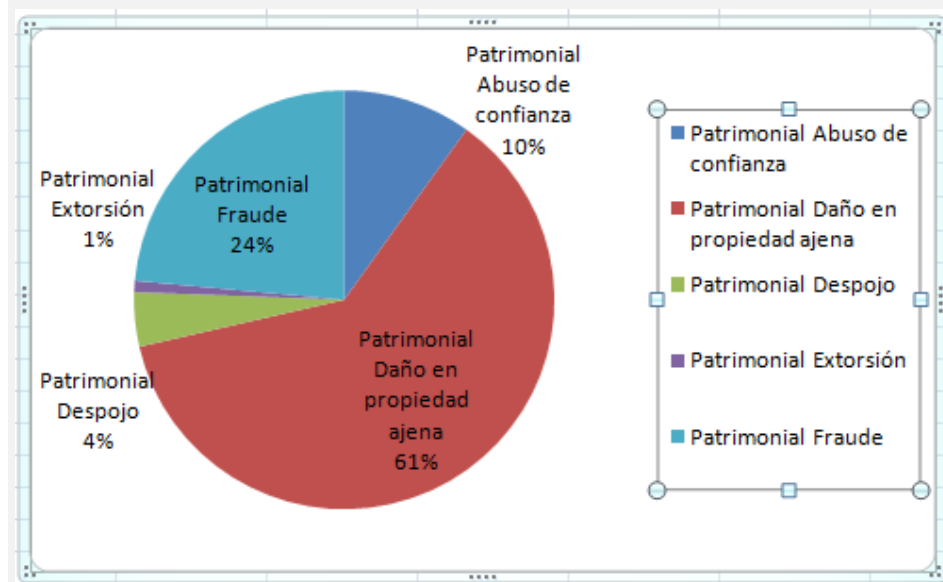
En dicho menú seleccione Nombre de categoría y porcentaje. Luego haga clic en Cerrar.

Imagen 14. Formato de etiqueta de datos. Como mostrar nombre y porcentaje



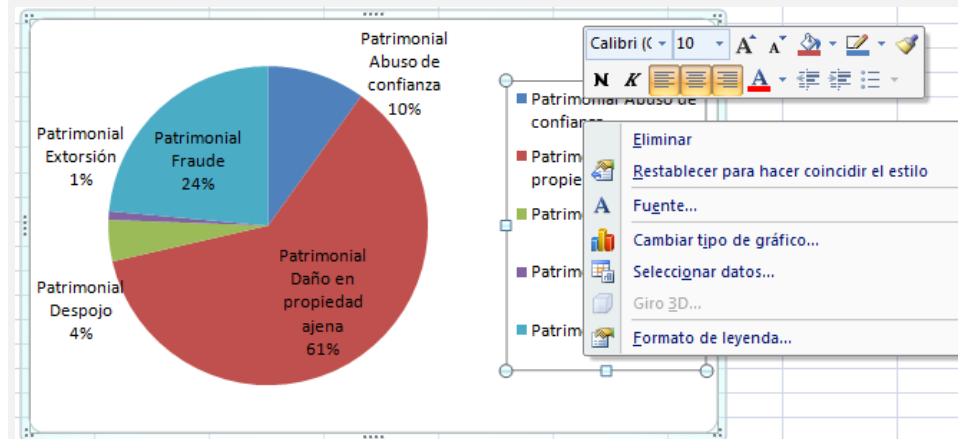
Coloque el apuntador del ratón sobre la leyenda que se encuentra al lado derecho, y haga clic con el botón izquierdo del ratón, para seleccionarla.

Imagen 15. Paso 1 para no mostrar la leyenda en un gráfico



Seleccione eliminar en el menú emergente

Imagen 16. Paso 2 para no mostrar la leyenda en un gráfico



Pruebe a cambiar el tamaño del gráfico, el tipo de letra y la ubicación de las leyendas. Para ello puede utilizar las herramientas de gráficos o bien colocando el apuntador del ratón sobre cualquier parte de la gráfica y haciendo doble clic con el botón derecho del ratón

Su gráfica debe verse como se muestra en la hoja gráfico 1 del archivo Capítulo 2. Tabla 1.xls.

Gráficas de columna

Las gráficas de barra o de columna son útiles cuando se necesita representar a más de 5 categorías.

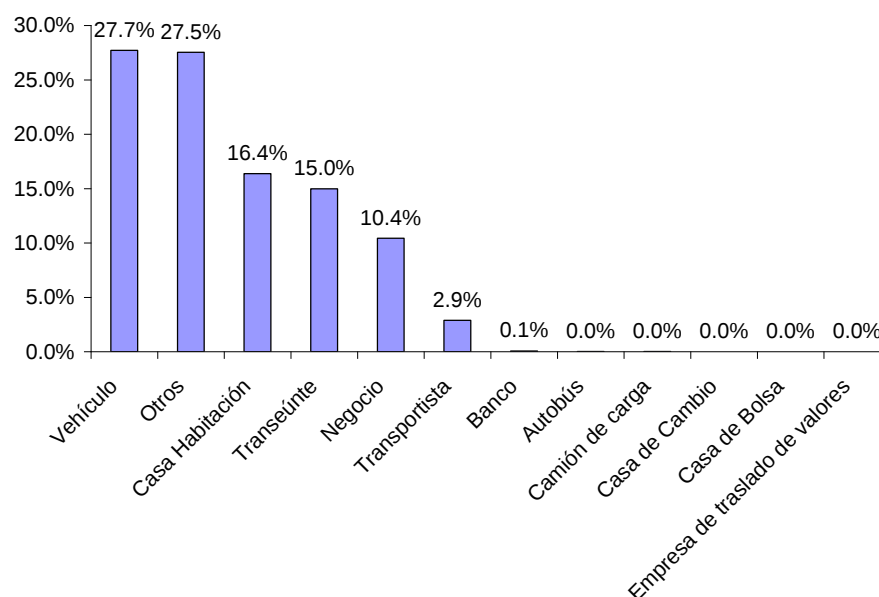
En la tabla 3 se muestra al delito de robo y sus modalidades y debido a que son más de 5 categorías las que se desea representar en forma gráfica, se presenta la información mediante una gráfica de columnas, ya que en conjunto son 12.

Tabla 17. México: denuncias ante las agencias del ministerio público del fuero común por delitos de robo, según modalidad, para el año 2006.

Delito	Modalidad	denuncias	%
Robo	Autobús	945	0.0%
	Banco	3,294	0.1%
	Camión de carga	935	0.0%
	Casa de Bolsa	15	0.0%
	Casa de Cambio	150	0.0%
	Casa Habitación	880,720	16.4%
	Empresa de traslado de valores	10	0.0%
	Negocio	561,039	10.4%
	Otros	1,481,207	27.5%
	Transeúnte	805,518	15.0%
	Transportista	155,306	2.9%
	Vehículo	1,490,722	27.7%
Total Robo		5,379,861	100.0%

Fuente: Arturo Arango y Cristina Lara, "Denuncias ante agencias del ministerio público del fuero común: 1997 a 2006" En "Estadísticas de Seguridad Pública en México". Elaborado con base en información del Sistema Nacional de Seguridad Pública. (Septiembre 2007). CrisAdA Editorial: Seguridad Pública en México. Tabla Arango y Lara. www.seguridadpublicaenmexico.org.mx

Gráfica 2. México: denuncias ante las agencias del ministerio público del fuero común por delitos de robo, según modalidad, para el año 2006



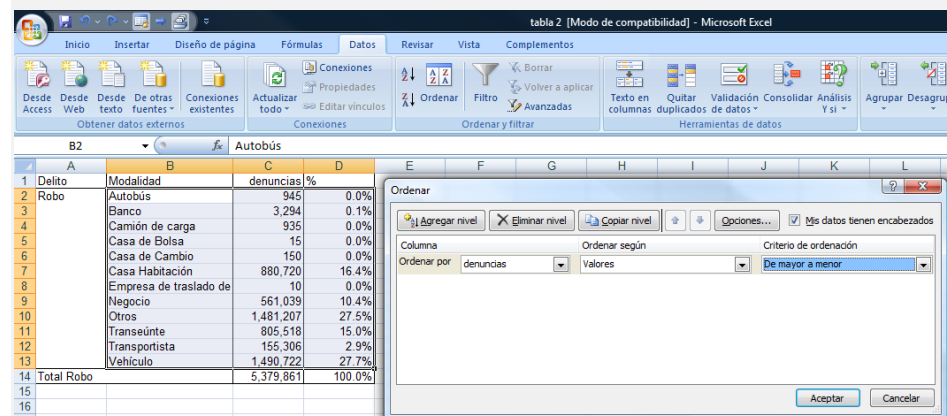
Paso a Paso con Excel: gráficas de columnas

Ayuda para el
Maestro

Capitulo 2.
Tabla 1.xls

Para generar las gráficas de columnas en Excel, sólo teclee los datos en una hoja nueva y seleccione la información que va a graficar.

Imagen 17. Como ordenar la información en Excel.



Tip: Una vez tecleados los datos éstos se ordenan en forma descendente, sobre la columna que contiene a los porcentajes.

Para ello coloque el apuntador en la celda c1, haga clic con el botón izquierdo del ratón, y en Datos.Ordenar, seleccione Ordenar. En el menú emergente seleccione Ordenar por denuncias (denuncias es el nombre que se puso a la columna), según Valores, escogiendo el criterio de ordenación de mayor a menor. Para ello, puede cambiar los valores de cada opción haciendo clic sobre el triángulo invertido que aparece a la derecha de cada ventana de opciones.

Esto garantiza que se muevan solo los valores y que la cifra de total permanezca.

Las fórmulas utilizadas para calcular la suma del total de robo y los porcentajes se muestran a continuación:

Imagen 18. Fórmulas para obtener el porcentaje

	A	B	C	D
1	Delito	Modalidad	denuncias	%
2	Robo	Vehículo	1490722	=C2/\$C\$14
3		Otros	1481207	=C3/\$C\$14
4		Casa Habitación	880720	=C4/\$C\$14
5		Transeúnte	805518	=C5/\$C\$14
6		Negocio	561039	=C6/\$C\$14
7		Transportista	155306	=C7/\$C\$14
8		Banco	3294	=C8/\$C\$14
9		Autobús	945	=C9/\$C\$14
10		Camión de carga	935	=C10/\$C\$14
11		Casa de Cambio	150	=C11/\$C\$14
12		Casa de Bolsa	15	=C12/\$C\$14
13		Empresa de traslado de valores	10	=C13/\$C\$14
14	Total Robo		=SUMA(C2:C13)	1

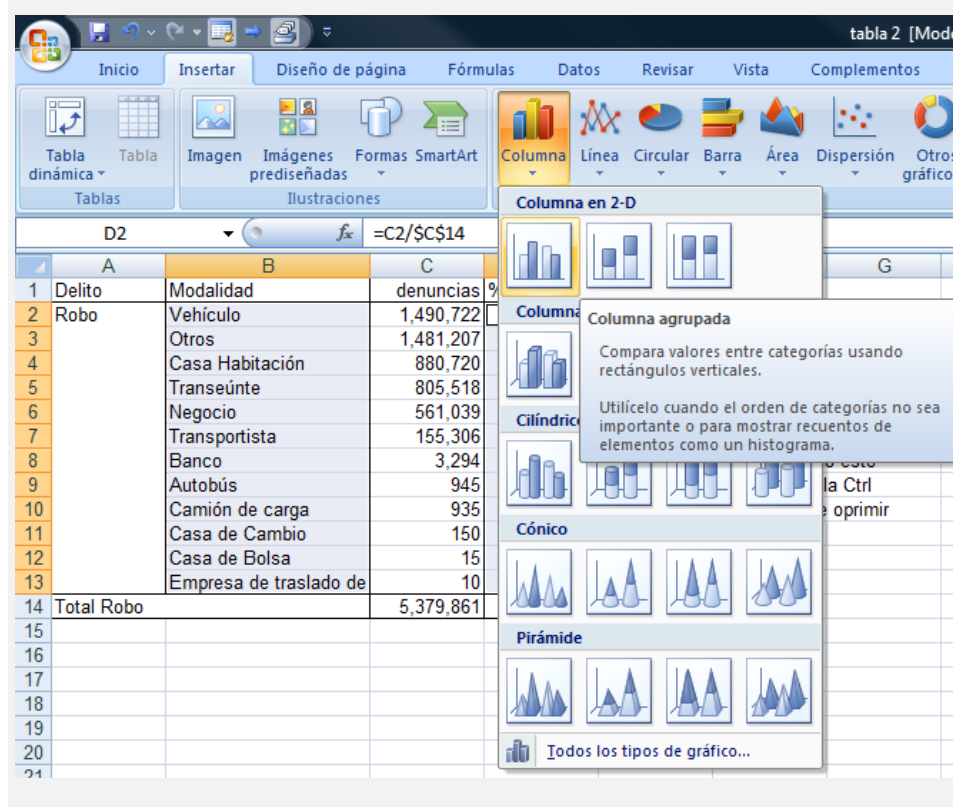
Seleccione los datos en las columnas B y D. Para ello seleccione el rango B2:B13 y una vez hecho esto oprima la tecla Ctrl y sin dejar de oprimirla con el ratón seleccione el rango D2:D13.

Imagen 19. Selección de columnas no contiguas

	A	B	C	D
1	Delito	Modalidad	denuncias	%
2	Robo	Vehículo	1,490,722	27.7%
3		Otros	1,481,207	27.5%
4		Casa Habitación	880,720	16.4%
5		Transeúnte	805,518	15.0%
6		Negocio	561,039	10.4%
7		Transportista	155,306	2.9%
8		Banco	3,294	0.1%
9		Autobús	945	0.0%
10		Camión de carga	935	0.0%
11		Casa de Cambio	150	0.0%
12		Casa de Bolsa	15	0.0%
13		Empresa de traslado de	10	0.0%
14	Total Robo		5,379,861	100.0%

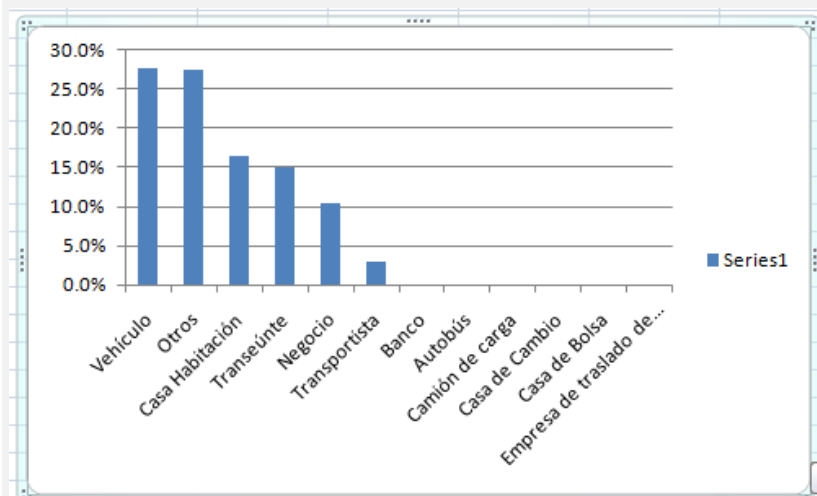
.En el menú insertar, seleccione Columna agrupada

Imagen 20. Comando para gráficos de columna



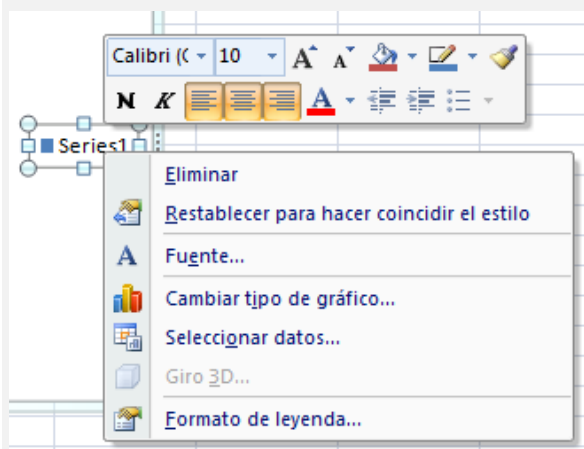
Su gráfico debe verse así.

Imagen 21. Gráfico de columnas



Ahora coloque el cursor sobre la leyenda , que se muestra a la derecha del gráfico y haga doble clic con el botón derecho del ratón.

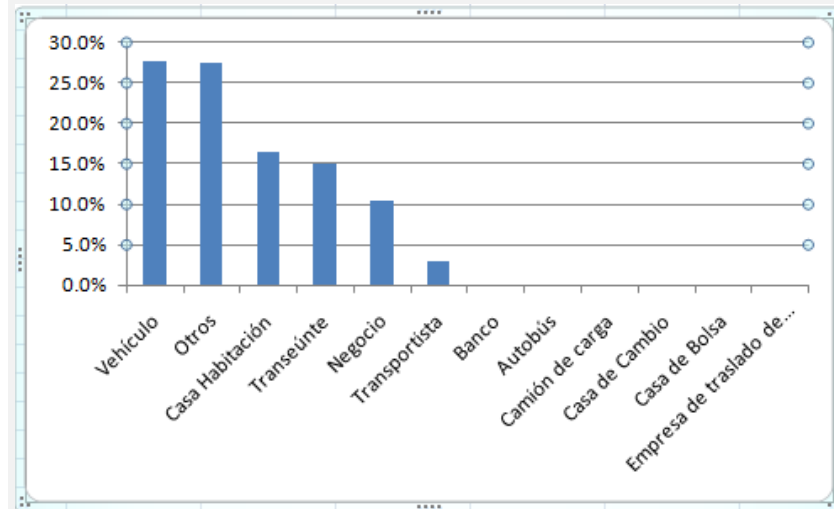
Imagen 22. Como eliminar la leyenda en un gráfico



En el menú emergente haga clic en eliminar.

Coloque el cursor sobre alguna de las líneas de división principales (las paralelas al eje de las x's) y haga clic con el botón izquierdo.

Imagen 23. Gráfico de columnas donde no se muestra la leyenda



Haga clic con el botón derecho del ratón y seleccione borrar en el menú emergente

Su gráfica debe verse como se muestra en la hoja gráfica 1, del archivo tabla 2, que acompaña a este material impreso.

Ejercicio

Actividades



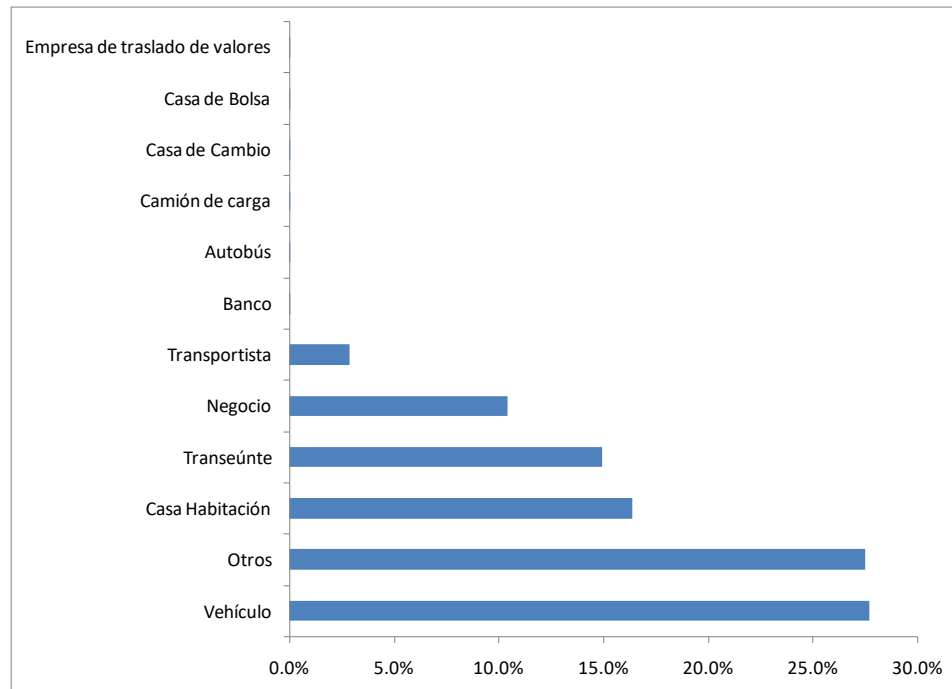
Estudie los archivos tabla 1.xls y tabla 2.xls. Una vez que está seguro que los comprende elimine las hojas gráfico 1 y gráfico 2 y siguiendo las instrucciones, genere los gráficos por usted mismo hasta que éstos sean iguales a los que se muestran en la gráfica 1 y gráfica 2 de este apartado. Pruebe cambiando el gráfico 2 a gráfica de barras en vez de columnas

Coloque etiquetas en cada barra.

Elimine el eje de las Y

Una vez terminado el ejercicio, su gráfica debe verse así

Gráfica 3. México: denuncias ante las agencias del ministerio público del fuero común por delitos de robo, según modalidad, para el año 2006. En barras.



Para generar una tabla de distribución de frecuencias con Excel.

La tabla 4, que se presenta a continuación muestra las tasas de víctimas hombres por cada 100 mil habitantes hombres en México para el año 2004, según los resultados obtenidos por usted, al realizar el ejercicio adicional 1, en el archivo de ejercicios capítulo 1.xls. Revise su archivo con los ejercicios realizados con objeto de recordar el procedimiento de obtención de esta tabla. Observe en particular la columna N, de la hoja ejercicio adicional

Tabla 18. Tasas de víctimas hombres por 100 mil habitantes hombres en México para el año 2004, según resultados de la ENSI-3, por entidad de ocurrencia del delito. (Sólo considera a las personas de 18 años y más)

Tasas por				Tasas por			
entidad_levantamiento	100 mil hom.	frecuencia	Porcentaje	entidad_levantamiento	100 mil hom.	frecuencia	Porcentaje
Aguascalientes	14304	1	0.03125	Morelos	14001	1	0.03125
Baja California	21564	1	0.03125	Nayarit	8095	1	0.03125
Baja California Sur	12070	1	0.03125	Nuevo León	8489	1	0.03125
Campeche	11116	1	0.03125	Oaxaca	10726	1	0.03125
Chiapas	2599	1	0.03125	Puebla	13041	1	0.03125
Chihuahua	12414	1	0.03125	Querétaro de Arteaga	8905	1	0.03125
Coahuila de Zaragoza	9516	1	0.03125	Quintana Roo	17020	1	0.03125
Colima	8410	1	0.03125	San Luis Potosí	6846	1	0.03125
Distrito Federal	22236	1	0.03125	Sinaloa	15634	1	0.03125
Durango	9307	1	0.03125	Sonora	16677	1	0.03125
Guanajuato	8976	1	0.03125	Tabasco	7525	1	0.03125
Guerrero	8302	1	0.03125	Tamaulipas	13125	1	0.03125
Hidalgo	9511	1	0.03125	Tlaxcala	9775	1	0.03125
Jalisco	16585	1	0.03125	Veracruz-Llave	4888	1	0.03125
México	19706	1	0.03125	Yucatán	15190	1	0.03125
Michoacán de Ocampo	9043	1	0.03125	Zacatecas	7541	1	0.03125

Fuente: Resultados de la ENSI-3 en www.icesi.org.mx

Las filas de frecuencia y porcentaje se obtuvieron asignando una frecuencia de 1 a cada entidad de la república y a dividir ese valor de 1, entre la suma de 32 que es el número de entidades federativas con que cuenta México, respectivamente.

Técnicamente a este tipo de tabla se le conoce como **distribución de frecuencia desagrupada**. Esto es así, ya que se muestran las tasas de víctimas hombres que cada entidad federativa tiene, en conjunto con la frecuencia (**f**) que cada estado tiene en el conjunto general. En esta distribución ninguna entidad muestra la misma tasa que alguna otra, es decir que las tasas para cada estado son únicas, sin embargo cada entidad y su tasa ocurren una sola vez por lo que tienen una frecuencia (**f**) de 1. En virtud de que existen muchos datos (32) esta información se vería mucho mejor utilizando una **distribución de frecuencias agrupada**.

El primer paso para construir una **distribución de frecuencias agrupada** es decidir el número de clases o el intervalo de clase necesario para agrupar a su distribución. Aunque esta decisión es de criterio, es crucial para la forma en que se verá su distribución, y existen algunas reglas al respecto.

La primera de ellas es que su número de clases no sea tan grande de tal forma que, al ser tantas, la nueva tabla no sea tan confusa como la tabla original. Asimismo el número de clases debe ser el necesario de tal forma que, al agruparse la información, no se oculte información relevante en alguna clase.

Otra regla a seguir en la construcción de la tabla de distribución de frecuencias agrupada es que cada dato debe pertenecer a una y sólo una clase. Es decir no debe haber ambigüedad de si una

observación pertenece a una u otra clase por lo que no se deben traslapar los datos.

Ejercicio

Actividades
Paso a Paso
con Excel:
distribución de
frecuencias

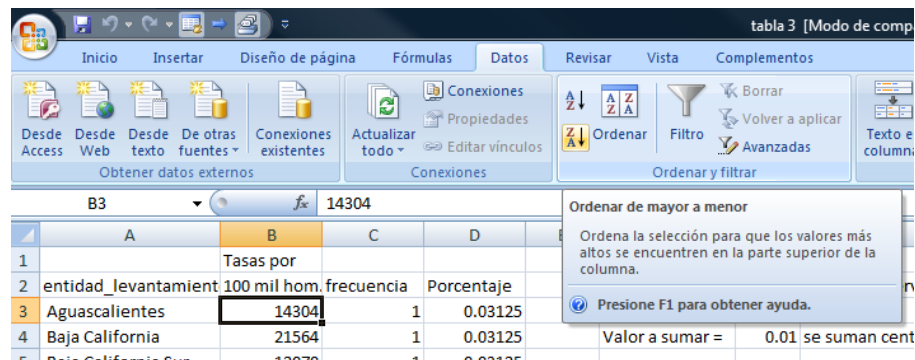


Vaya al archivo tabla3.xls en Internet Abra el archivo tabla 3.xls disponible en el material en Internet que acompaña a este material.

El primer paso es ordenar la información de forma ascendente, donde la observación con el menor valor se pone al inicio y la observación con el mayor valor se pone al final. Abra el archivo tabla 3.xls, disponible en Internet, que acompaña a este material.

Seleccione la celda b3 y en datos seleccione Ordenar de mayor a menor.

Imagen 24. Selección de datos para ordenar en orden ascendente.



Su información debe verse como se indica a continuación

Imagen 25. Datos ordenados en orden ascendente

	A	B	C	D
1		Tasas por		
2	entidad_levantamiento	100 mil hom. frecuencia		Porcentaje
3	Chiapas	2599	1	0.03125
4	Veracruz-Llave	4889	1	0.03125
5	San Luis Potosí	6846	1	0.03125
6	Tabasco	7526	1	0.03125
7	Zacatecas	7542	1	0.03125
8	Nayarit	8096	1	0.03125

Intervalo de clase

Una vez que los datos están puestos en orden ascendente se calcula el rango de los datos para decidir el número de intervalos o clases necesarias para describir a nuestra información agrupada. (La regla es que haya entre 5 a 10 clases.)

En el caso del modelo que se presenta se utilizan 10 clases

Rango es la diferencia entre el valor máximo y el valor mínimo del conjunto de datos.

Una de las medidas de variación más simple es el rango.

Ejemplo Halle el rango de los siguientes datos: 3, 3, 4, 4, 4, 4, 5, 5

Solución:

Como los datos varían de 3 a 5 entonces $\text{rango} = 5 - 3 = 2$

Ejemplo Halle el rango de los siguientes datos: 1, 2, 3, 4, 5, 6, 7

Solución:

Los datos varían de 1 a 7 entonces $\text{rango} = 7 - 1 = 6$

De los datos ordenados, sabemos que el valor máximo corresponde al mostrado por el Distrito Federal con una tasa de 22 236 víctimas hombres por cada 100 mil hombres; mientras que el valor mínimo corresponde al estado de Chiapas con 2 599 víctimas hombres por cada 100 mil hombres; de donde el rango es:

$$\text{Rango} = 22\,236 - 2\,599 = 19\,637$$

El siguiente paso consiste en dividir este rango entre un número suficiente de intervalos que nos garantice una imagen detallada y precisa de la información. Conviene mencionar nuevamente que la regla es que el número de intervalos o clases es de entre 5 a 10 clases. Con esto en mente, digamos que, para el caso que nos ocupa 10 clases son suficientes, por lo que dividimos el valor del rango, encontrado en el paso anterior, entre el número de clases deseadas.

Intervalo de la clase

El intervalo de la clase se obtiene dividiendo al rango entre el número de clases deseadas.

$\text{Intervalo de la clase} = \text{Rango} / \text{número de clases deseado}$

Intervalo de la clase

El intervalo de la clase resulta de dividir al rango entre el número de clases deseado.

$$\text{Intervalo de clase} = 19\,637 / 10 = 1963.7$$

Una vez que se conoce el intervalo de clase, se procede a la construcción de las clases (10 en el ejemplo seleccionado). Como regla empírica, el límite inferior de la primera clase es el valor mínimo de los datos. Asimismo se debe garantizar, como se mencionó más atrás, que todos los valores caigan en una y sólo una de las clases.

Para garantizar esto, se acostumbra sumar un número que por su magnitud haga imposible que exista duda de en cual clase debe estar cada observación. Para ello, si el intervalo de clase es entero, se suma un decimal, si el intervalo de clase tiene a un decimal se suman centésimas y así sucesivamente. Es decir que se suma un decimal extra. Si el intervalo de clase es entero sume un decimal, si contiene a un decimal agregue dos, si contiene a dos, agregue tres y así sucesivamente.

Tabla 19. Valores a sumar para calcular el ancho de clase.

<i>Ancho de clase</i>	<i>se suman</i>	<i>Valor a sumar</i>	
con enteros	décimas	0.1	10.1
con decimales	centésimas	0.01	10.01
Con centésimos	milésimos	0.001	10.001
Regla general			
se acostumbra sumar un decimal más, de la cantidad calculada			

Resumiendo la información hasta aquí encontrada, tenemos que:

Tabla 20. Resumen de cálculos para el rango e intervalos

Rango =	19637	
Intervalo de clase	1963.7	Como el intervalo de clase contiene a un decimal
Clases =	10	
Valor a sumar =	0.01	se suman centésimas

De la información ordenada obtenida más atrás en los pasos 3 y 4, sabemos que el menor valor le corresponde al estado de Chiapas con un valor de 2 599. Ese valor es el que se pone en el límite inferior de la clase 1, como el ancho de clase es 1 963.7 ese valor se suma al límite inferior de la clase 1, obteniendo así el límite superior de la clase 1.

Límite inferior de la clase 1 = valor mínimo de los datos VMD = 2599

Límite superior de la clase 1 = VMD + intervalo de clase =
 $= 2599 + 1963.7 = 4562.70$

Tabla 21. Tabla para cálculos de límites inferior y superior.

Clase	<i>Lim. inferior</i>	<i>Intervalo de clase</i>	<i>lim. Superior</i>
1	2599.00	1963.70	4562.70
2	4562.71	1963.70	6526.41
3	6526.42		
4			
5			
6			
7			
8			
9			
10			

Ejercicio

Actividades
Paso a Paso
con Excel:
distribución de
frecuencias



Para completar este ejercicio busque el archivo tabla 3, hoja 10 clases.

La tabla a completar se encuentra en la columna G

Estudie la tabla y revise las fórmulas en cada celda e intente rellenar los datos tecleando fórmulas similares en las celdas cuyos datos están faltantes.

Como los datos se calculados contienen a un decimal, (en el intervalo de clase contiene al valor .7) se agregan dos decimales .01)

Entonces

$$\begin{aligned} \text{Límite inferior de la clase 2} &= \text{Límite superior de la clase 1} + 0.01 = \\ &= 4\,562.70 + 0.01 = 4\,562.71 \end{aligned}$$

El límite superior de la clase 2 se obtiene sumando el intervalo de clase al valor encontrado para el límite inferior de la clase 1. Se sigue con este procedimiento hasta encontrar a las diez clases.

Como ejercicio calcule los datos faltantes y compare sus resultados contra los de la tabla 4.

Es importante señalar que la cantidad a sumar, sólo se debe sumar al límite inferior de la clase 2.

Tabla 22. Distribución de frecuencias. Tasas de víctimas hombres por 100 mil habitantes hombres en México para el año 2004, según resultados de la ENSI-3.

Clase	Lim. inferior	Lim. superior	marca de clase	frecuencia	% de frecuencia	frecuencia acumulada	% acumulado
1	2599.00	4562.70	3580.85	1	3.1%	1	3.1%
2	4562.71	6526.41	5544.56	1	3.1%	2	6.3%
3	6526.41	8490.11	7508.26	7	21.9%	9	28.1%
4	8490.11	10453.81	9471.96	7	21.9%	16	50.0%
5	10453.81	12417.51	11435.66	4	12.5%	20	62.5%
6	12417.51	14381.21	13399.36	4	12.5%	24	75.0%
7	14381.21	16344.91	15363.06	2	6.3%	26	81.3%
8	16344.91	18308.61	17326.76	3	9.4%	29	90.6%
9	18308.61	20272.31	19290.46	1	3.1%	30	93.8%
10	20272.31	22236.01	21254.16	2	6.3%	32	100.0%
Intervalo de clase		1963.70	Observaciones	32			
decimales a sumar		0.01					

De la tabla 17, la marca de clase, para cada clase se obtiene sumando los límites inferior y superior y dividiendo al resultado de dicha suma entre 2.

Para obtener la columna de frecuencia se comparan los resultados de la distribución de frecuencias contra los datos originales. Esta comparación se realiza en la siguiente tabla.

Tabla 23. Tabla para obtención de la columna de frecuencia

entidad_levantamiento	Tasas por 100 mil hom.	Clase	entidad levantamiento	Tasas por 100 mil hom.	Clase
Chiapas	2599	1	Oaxaca	10726	
Veracruz-Llave	4888	2	Campeche	11116	
San Luis Potosí	6846		Baja California Sur	12070	
Tabasco	7525		Chihuahua	12414	
Zacatecas	7541	3	Puebla	13041	
Nayarit	8095		Tamaulipas	13125	
Guerrero	8302		Morelos	14001	
Colima	8410		Aguascalientes	14304	
Nuevo León	8489		Yucatán	15190	
Querétaro de Arteaga	8905	4	Sinaloa	15634	
Guanajuato	8976		Jalisco	16585	
Michoacán de Ocampo	9043		Sonora	16677	
Durango	9307		Quintana Roo	17020	
Hidalgo	9511		México	19706	
Coahuila de Zaragoza	9516		Baja California	21564	
Tlaxcala	9775		Distrito Federal	22236	

Actividades



Para completar este ejercicio busque el archivo tabla 3, hoja resultados.

La tabla a completar se encuentra en la columna J

El valor correspondiente a Veracruz es de 4 888 que se encuentra

entre los valores máximo y mínimo de la clase 2, por lo que a dicha entidad se le asigna la clase 2. El valor que corresponde a Zacatecas (7 541) se encuentra entre los valores mínimo y máximo de la clase 3; asimismo el de Querétaro de Arteaga que se encuentra en la clase 4. **Como ejercicio encuentre las clases en que se encuentra cada entidad de la república y complete la tabla.**

Una vez completada la tabla, cuente cuantos valores están en la clase 1 y el resultado se coloca en la columna correspondiente a frecuencia de la tabla de distribución de frecuencias. Esta columna nos indica cuantos valores caen en cada clase.

La columna % de frecuencia se obtiene dividiendo a la frecuencia (f) de cada clase entre el número total de observaciones y multiplicando por 100, así para la clase 4, el % de frecuencia de dicha clase es:

$$f \text{ de la clase 4} = 7 / 32 = 0.219 * 100 = 21.9 \%$$

Por lo que toca a la columna de frecuencia acumulada esta se deriva de la columna de frecuencias. La frecuencia acumulada de la clase 1 es 1, la frecuencia acumulada hasta la clase 2, es la frecuencia de la clase 2 más la frecuencia acumulada de la clase 1, es decir, $1 + 1 = 2$.

La frecuencia acumulada de la clase 3, es la frecuencia de la clase 3, (7) más la frecuencia acumulada hasta la clase 2, así $7 + 2 = 9$. Se sigue con este procedimiento hasta completar toda la columna.

Observe que el total en la columna de frecuencia no puede ser mayor al total de observaciones (32) hecho que se registra en la frecuencia acumulada de la última clase cuyo valor es 32, es decir, frecuencia de la clase 10 más la frecuencia acumulada hasta la clase 9.

La columna de % acumulado se obtiene dividiendo cada valor de la columna de frecuencia acumulada entre el valor máximo de la frecuencia acumulada o entre la suma de los datos de la columna de frecuencia.

Actividades

Estudie el ejemplo del archivo tabla 3, hoja “resultados”, y vea además la hoja “modelo”



Paso a Paso con Excel

Modelo para obtener una distribución de frecuencias

Ayuda para el maestro



Imagen 26. Modelo con fórmulas en Excel para realizar la tabla de distribución de frecuencias, mostrada en la tabla 17.

Clase	Lim. inferior	Lim. superior	marca de clase	frecuencia	% de frecuencia	frecuencia acumulada	% acumulado
1	2599	=B2+C12	=(B2+C2)/2	1	=E2/\$E\$12*100	=E2	=G2/\$G\$11*100
2	=C2+\$C\$13	=B3+\$C\$12	=(B3+C3)/2	1	=E3/\$E\$12*100	=E3+G2	=G3/\$G\$11*100
3	=C3	=B4+\$C\$12	=(B4+C4)/2	7	=E4/\$E\$12*100	=E4+G3	=G4/\$G\$11*100
4	=C4	=B5+\$C\$12	=(B5+C5)/2	7	=E5/\$E\$12*100	=E5+G4	=G5/\$G\$11*100
5	=C5	=B6+\$C\$12	=(B6+C6)/2	4	=E6/\$E\$12*100	=E6+G5	=G6/\$G\$11*100
6	=C6	=B7+\$C\$12	=(B7+C7)/2	4	=E7/\$E\$12*100	=E7+G6	=G7/\$G\$11*100
7	=C7	=B8+\$C\$12	=(B8+C8)/2	2	=E8/\$E\$12*100	=E8+G7	=G8/\$G\$11*100
8	=C8	=B9+\$C\$12	=(B9+C9)/2	3	=E9/\$E\$12*100	=E9+G8	=G9/\$G\$11*100
9	=C9	=B10+\$C\$12	=(B10+C10)/2	1	=E10/\$E\$12*100	=E10+G9	=G10/\$G\$11*100
10	=C10	=B11+C12	=(B11+C11)/2	2	=E11/\$E\$12*100	=E11+G10	=G11/\$G\$11*100
Intervalo de clase		1963.7	Observaciones	=SUMA(E2:E11)			
decimales a sumar		0.01					

Todo el trabajo hasta aquí realizado representa a una gran cantidad de operaciones y cálculos, afortunadamente los programas como Excel permiten ahorrar tiempo, ya que se puede generar un modelo que sirva para todas o casi todas las ocasiones. Si observa con atención el modelo en la tabla 6, verá que para realizar una tabla de distribución de frecuencias con 10 clases sólo necesita contar con los valores que se muestran en color amarillo, es decir: el valor inicial o el valor mínimo (el límite inferior de la clase 1) el cual se coloca al inicio (para este caso 2 599, celda b2, color amarillo en la tabla 6), el intervalo de clase (1963.7, celda c12, también en amarillo. Para obtener el intervalo de clase recuerde que necesita conocer los valores máximo y mínimo de su distribución de frecuencia desagrupada) los decimales a sumar (0.01, celda c13, en amarillo) y, finalmente, las frecuencias o número de veces que caen las observaciones en cada clase (celdas e2:e12) las cuales se derivan de comparar el valor que posee cada observación contra las clases generadas con los procedimientos aquí descritos.

Histogramas.

Actividades

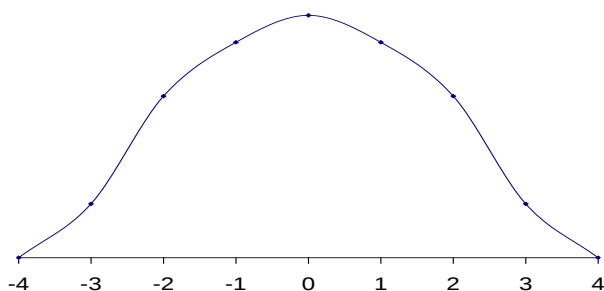


Vaya a la tabla 3, hoja “resultados”, y genere usted mismo el histograma, siguiendo las indicaciones que se dan.

A las distribuciones de frecuencias se les puede representar gráficamente mediante los histogramas. Los histogramas son gráficos de columnas, (como la realizada más atrás) en donde el eje de las x's corresponde a la **marca de clase** y en el eje de las y's se coloca al % de frecuencia

La generación del histograma es importante ya que, cuando se tiene una gran cantidad de información socio- demográfica ésta tiende a adquirir la forma de una campana al representarla mediante un histograma.

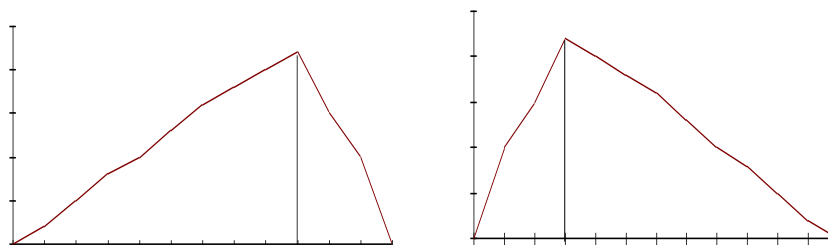
Imagen 27. Curva de distribución normal



Cuando una distribución de frecuencias se asemeja a la campana, se dice que esa distribución es una **distribución normal**. La forma que presente nuestra distribución de frecuencias es muy importante ya que puede darse el caso que esta se encuentre cargada a la derecha y, por ello, ser **negativamente asimétrica** o cargada a la izquierda y, entonces, ser **positivamente asimétrica**.

Al respecto sólo base mencionar que para saber si la distribución es negativa o positivamente asimétrica imagine usted una recta numérica donde el valor cero corresponde exactamente a la posición que ocupa el centro de la gráfica de una distribución simétrica.

Imagen 28. Distribuciones asimétricas



Positivamente asimétrica

Negativamente asimétrica

Observe que a la derecha del centro de la gráfica (valor cero) tenemos valores positivos. Si la gráfica está a la derecha de ese valor cero, imaginario, entonces su distribución será positivamente asimétrica, si está a la izquierda del centro de la gráfica (de ese valor cero imaginario), entonces su distribución será negativamente asimétrica.

Estos dos casos de asimetría son los que se presentan con más frecuencia en el análisis delictivo.

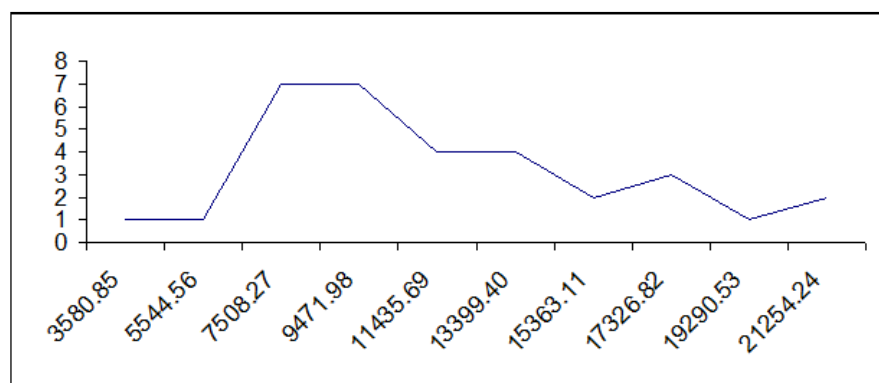
En el caso de que su distribución sea **normal**, implica que la mayoría de los valores de las observaciones se cargan al centro, con pocas observaciones en los valores mínimos y máximos. Si su

distribución es **positivamente asimétrica** implica que la mayoría de las observaciones están cargadas hacia los valores altos o máximos. Cuando la mayoría de las observaciones están cargadas hacia los valores más bajos entonces la distribución es **negativamente asimétrica**.

Polígono de frecuencias

Otra forma de presentar a la información de frecuencias es mediante un polígono de frecuencias. Un polígono es aquella figura con líneas unidas una detrás de la otra, hasta el final.

Imagen 29. Polígono de frecuencias

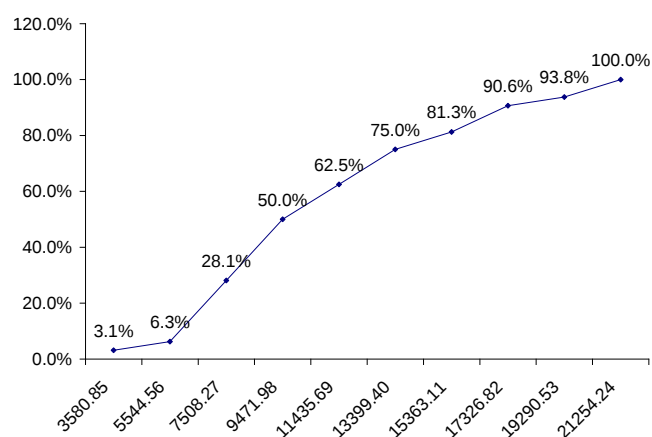


Ojiva o curva de frecuencia acumulada

Otra curva que se deriva de la tabla de frecuencias es la conocida como “ojiva o curva de frecuencia acumulada”. Para la realización de esta gráfica se toma a la columna de “frecuencia acumulada” o a la de “% de frecuencia acumulada” según se desee graficar en datos netos o en porcentaje.

La ojiva o curva de frecuencias acumulada debe verse así:

Imagen 30. Ojiva o curva de frecuencias acumulada



Como interpretar

Las tablas y gráficas obtenidas se repiten a continuación

Imagen 31. Tasas de víctimas hombres por 100 mil habitantes hombres en México para el año 2004, según resultados de la ENSI-3. (Sólo considera a las personas de 18 años y más). Distribución de frecuencias en Excel

	A	B	C	D	E	F	G	H	I
1	Clase	Lim. inferior	Lim. superior	marca de clase	frecuencia	% de frecuencia	frecuencia acumulada	%	
2	1	2599.00	4562.70	3580.85	1	3.13	1	3.13	
3	2	4562.71	6526.41	5544.56	1	3.13	2	6.25	
4	3	6526.42	8490.12	7508.27	7	21.88	9	28.13	
5	4	8490.13	10453.83	9471.98	7	21.88	16	50.00	
6	5	10453.84	12417.54	11435.69	4	12.50	20	62.50	
7	6	12417.55	14381.25	13399.40	4	12.50	24	75.00	
8	7	14381.26	16344.96	15363.11	2	6.25	26	81.25	
9	8	16344.97	18308.67	17326.82	3	9.38	29	90.63	
10	9	18308.68	20272.38	19290.53	1	3.13	30	93.75	
11	10	20272.39	22236.09	21254.24	2	6.25	32	100.00	
12	Intervalo de clase		1963.70	Observaciones		32			
13	decimales a sumar		0.01						

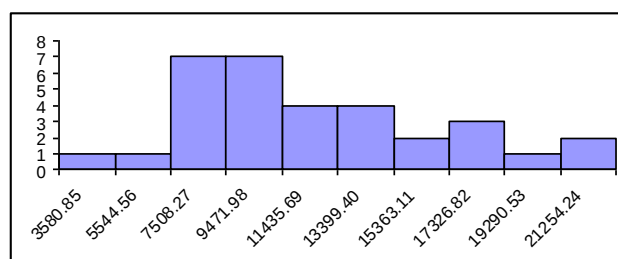
Al reacomodar la información de la tabla 3, como una distribución de frecuencias, y obtener las gráficas asociadas a dicha tabla encontramos que 14 entidades de la república (clases 3 y 4) concentran casi el 44% de las víctimas hombres por cada 100 mil habitantes hombres.

Imagen 32. Histograma, Polígono de frecuencias y Curva de frecuencias acumulada u Ojiva.

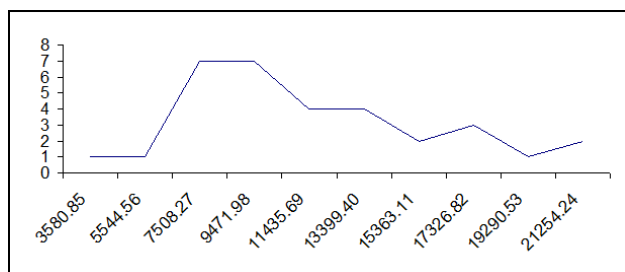
El número de víctimas hombres por cada 100 mil hombres está, en esas entidades, entre 6 526. 42 y 453.83 víctimas.

Son 8 entidades las que cuentan con un número mayor a las 14 381.26 víctimas hombres por cada 100 mil hombres que residen en esas entidades.

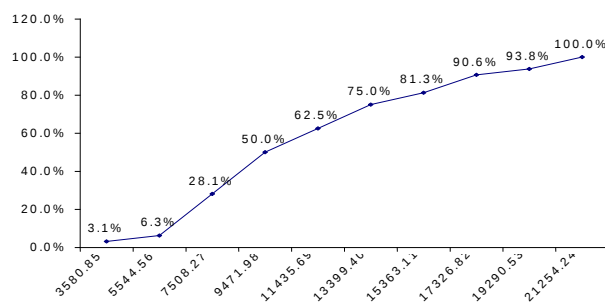
Histograma.



Polígono.



Curva de frecuencias acumulada.



En el 6.25% de las entidades del país , existen más de 20 272.39 víctimas hombres por cada 100 mil hombres. Es decir, que el número de víctimas en esas dos entidades en la clase 10, esta casi 10 veces por encima de la entidad que se encuentra en la clase 1, (Chiapas) y casi 5 veces por encima de la entidad (Veracruz) que se encuentra en la clase 2.

Otro dato interesante es que en la entidad con menor número de víctimas se victimiza a un poco más de 2 hombres por cada 100 hombres residentes en esa entidad , mientras que en las entidades con el mayor número de víctimas la victimización sobre personas del sexo masculino alcanza en promedio (la marca de clase es 21 524.24) a más de 21 hombres por cada 100 hombres residentes en esas entidades. Es decir más de 1 por cada 5 hombres ha sido víctima, en esas dos entidades de la clase superior.

Otra forma de decir lo anterior es que el nivel de victimización sobre personas del sexo masculino es de casi 10 veces en las entidades en la clase más alta sobre la de clase más baja.

Observando a la ojiva, podemos señalar que en el 50% de las entidades del país, el número de víctimas hombres es de 11 435.69 por cada 100 mil hombres residentes en esas entidades, o menos. El 75% de las entidades presentó un nivel de victimización igual o menor a 13 399.40 víctimas del sexo masculino por cada 100 mil residentes hombres en esas entidades.

Actividades



Considere la siguiente tabla.

Tabla 24 Personas de 18 años y más, que consideran que vivir en su municipio actual es inseguro y población de 18 años y más por entidad de residencia y sexo. 2004

Población según sexo de 18 años y más					Población según sexo de 18 años y más				
Entrevistados		Entrevistados		entidad_levantamiento	Entrevistados		Entrevistados		entidad_levantamiento
Hombre	Mujer	Hombre	Mujer		Hombre	Mujer	Hombre	Mujer	
21159	22133	57165	93928	Morelos	42124	46593	221104	290989	
145902	123944	521519	569937	Nayarit	7161	6550	46095	67043	
8618	8767	42926	52130	Nuevo León	57675	59092	351630	445851	
13633	10496	80221	97372	Oaxaca	54402	31589	264405	274935	
17094	22935	432627	556428	Puebla	79293	96336	495146	615034	
91918	77435	507512	598344	Querétaro de Arteaga	10585	12738	116385	151839	
31316	47469	188400	254457	Quintana Roo	33674	27355	154935	184014	
3957	3811	25927	49283	San Luis Potosí	14033	28659	147700	193761	
456876	405697	1736055	2126315	Sinaloa	83200	91869	360627	461948	
18276	16427	94446	135568	Sonora	58903	63927	246702	336446	
85163	74203	472979	553714	Tabasco	24703	33754	270668	305166	
40874	61842	256212	345265	Tamaulipas	85152	67048	340304	396475	
34234	20178	189603	276298	Tlaxcala	12097	7019	77393	87948	
201983	183359	687158	909282	Veracruz-Llave	50376	56823	620292	718423	
665578	399669	2484504	2718324	Yucatán	21689	36829	97311	170393	
37364	75518	314611	406563	Zacatecas	11734	9119	72410	108031	

Fuente: ICESI. 3a. Encuesta Nacional Sobre Inseguridad ENSI- 3
www.icesi.org.mx

En dicha tabla en las columnas de entrevistados se presenta, por sexo, sólo a aquellos quienes dijeron que vivir en su municipio de

residencia en ese momento de la encuesta era inseguro. En la columna de Población se presenta a la población de hombres y mujeres de 18 años y más.

Obtenga la tabla de frecuencias en datos absolutos, de entrevistados hombres, mujeres y en total para los dos sexos, en 10 clases.

Realice los gráficos de histograma, polígono y distribución de frecuencias acumuladas, para cada una de las 3 tablas encontradas en el paso anterior.

Interprete los resultados

Obtenga la tabla de frecuencias en tasas, por cada 100 mil hombres, por cada 100 mil mujeres y por cada 100 mil habitantes hombres y mujeres, en 10 clases.

Realice los gráficos de histograma, polígono y distribución de frecuencias acumuladas, para cada una de las 3 tablas encontradas en el paso anterior

Interprete los resultados.

Una vez, y sólo entonces, que haya realizado los ejercicios vaya a la pagina y en la sección correspondiente al material, abra el archivo:

Tabla 7 resultados

Y compare los resultados por usted obtenidos, contra los aquí proporcionados. Al abrir este archivo aparecerá un mensaje que le indica que éste está protegido y le proporciona un espacio para que usted teclee la clave que lo abre.

Para ello usted tiene que hacer primero el ejercicio de la tabla 7 y teclear el valor obtenido para la tasa de personas hombres que se sienten inseguros en su municipio de residencia habitual, para el estado de Baja California.

Proporcione la clave en enteros (5 dígitos) para ver los resultados.

Esta clave sirve también para auto evaluarse, si el archivo se abre con el valor proporcionado por usted, muy probablemente el ejercicio se habrá realizado en forma adecuada, si no se abre, esto implica que tiene algún error y que debe revisar su trabajo. Pregúntele a su profesor el cual seguramente podrá auxiliarlo en la detección de los errores que haya cometido.

Utilización de los modelos

Mediante la utilización de Excel es demasiado simple realizar cálculos ya sea que exista o no una gran cantidad de información. En este apartado revisaremos tres modelos los cuales se

presentan en el archivo “modelos capítulo 2.xls” que acompaña a este material.

Estos modelos le simplificarán todavía más su trabajo como analista; de tal manera que usted podrá dedicar menos tiempo a la parte mecánica que tiene que ver con la captura de la información y los procedimientos tediosos de cálculo y más tiempo a la parte de análisis.

Para ello abra el archivo “modelos capítulo 2.xls” y encontrará el primer modelo donde existen 4 hojas llamadas:

Distribución de frecuencias,

Por entidad,

Por municipio; y

Lista de municipios.

El primer modelo es universal y puede utilizarse con cualquier variable. Los modelos por entidad y municipio aunque se aplican al caso particular de México; en realidad mientras no se sobrepase el número de datos (32 para el caso del modelo por entidad y 2 455 para el caso del modelo por municipio) pueden aplicarse para cualquier variable que caiga en ese rango.

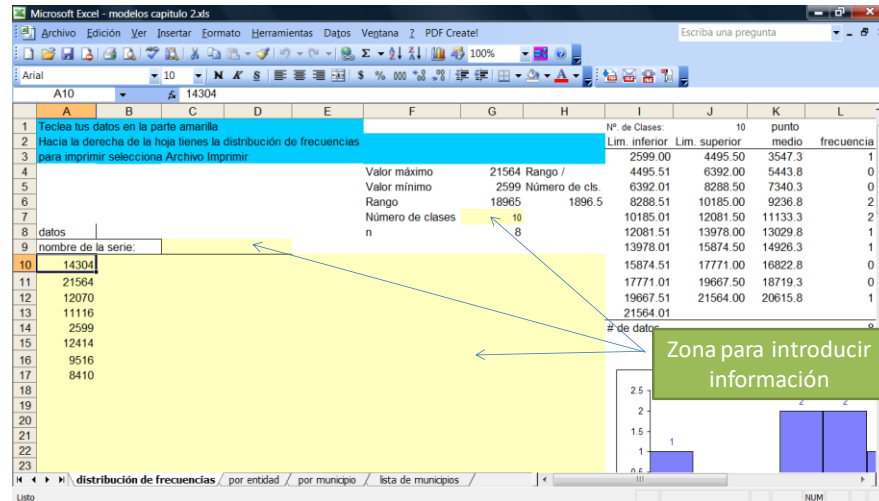
Sólo debe garantizar en caso que se utilice para cualquier otra variable que las filas no utilizadas, queden limpias sin ningún valor.

Es importante aprender a manejar estos modelos, ya que conforme se vaya pasando a otros temas, éstos se irán refinando y aumentando, de tal forma que al final, podamos terminar con un modelo en un solo archivo que nos haga más fácil el análisis y con ello la investigación sobre algún tema en particular.

Modelo de distribución de frecuencias para una variable

En la hoja distribución de frecuencias usted encontrará el modelo de éstas:

Imagen 33. Modelo de distribución de frecuencias para una variable



En este modelo usted encontrará tres zonas sombreadas en color amarillo. En la primera (celda G7) usted debe introducir el número de clases deseado (entre 5 a 10); en la segunda (celdas c9:d9) el nombre de la variable a la que se pretende obtener su distribución de frecuencias; y en la tercera y última zona (celdas a10:h23) usted deberá introducir los valores de las observaciones. No importa el orden de introducción ni en que celda se encuentren los valores, lo único importante es que todos sus datos deben estar dentro de la zona sombreada.

Este modelo es dinámico lo que implica que cada vez que se introduce un nuevo valor u observación, el modelo actualiza la información de forma automática.

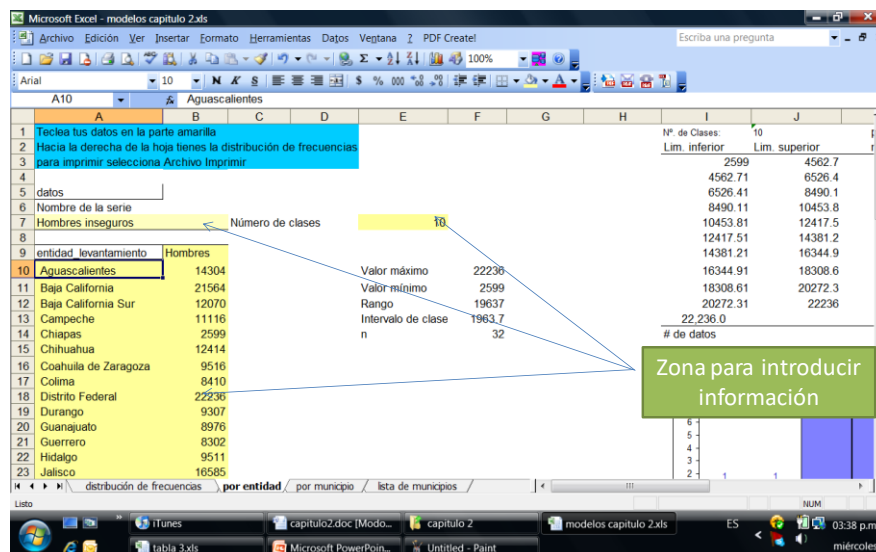
Observe que el modelo tiene los primeros 8 valores de las observaciones de la tabla 3, es decir a los valores correspondientes a las primeras 8 entidades, y observe que los datos se encuentran en desorden.

Como ejercicio, capture la información de las 24 entidades faltantes, en el orden que usted desee, y observe que ante cada nueva observación capturada, los datos se actualizan. Compare los resultados finales contra la distribución de frecuencias obtenida, más atrás, para la tabla 3; los cuales se muestran en la tabla 4.

Modelo para cuando se tiene una variable por entidad.

En la hoja por entidad usted encontrará el modelo para cuando tiene una variable por entidad de la república y se desea conservar la tabla de observaciones.

Imagen 34. Modelo de distribución de frecuencias para una variable por entidad.



El procedimiento para introducir la información es similar al del modelo anterior, con las zonas en amarillo donde se teclean los datos; aunque se puede copiar directamente la información desde alguna otra hoja.

Por ejemplo en el ejercicio donde se utilizó la tabla 7, se pedía realizar 6 distribuciones de frecuencia. La forma más simple de realizar dicho ejercicio era hacerlo mediante la utilización de este modelo.

Abrir el archivo tabla 7.xls

Abrir el archivo modelos capítulo 2.xls

En tabla 7, seleccionar las celdas a6: b37.

Una vez seleccionadas, haciendo clic con el botón derecho del ratón, seleccione copiar en el menú emergente. Observe que en las celdas seleccionadas se muestra un ligero parpadeo.

Haga clic, en la pestaña que aparece en la parte inferior de su pantalla, correspondiente al archivo modelos capítulo 2.xls

Asegúrese de que esta activa (es la que se ve en ese momento en la pantalla) la hoja "por entidad"

Coloque el apuntador del ratón sobre la celda A10 y haga clic con el botón derecho.

Seleccione pegado especial en el menú emergente

Seleccione valores

Haga clic en aceptar.

Teclee el nombre de la serie y el número de clases

Ahora bien, cuando copié las celdas a6: b37, del archivo tabla 7.xls al archivo “modelos capitulo 2.xls” hoja “por entidad” usted copio los nombres de las entidades y el orden que éstas tienen, así como los valores correspondientes a la variable “Personas hombres que consideran que vivir en el municipio de residencia actual es inseguro”.

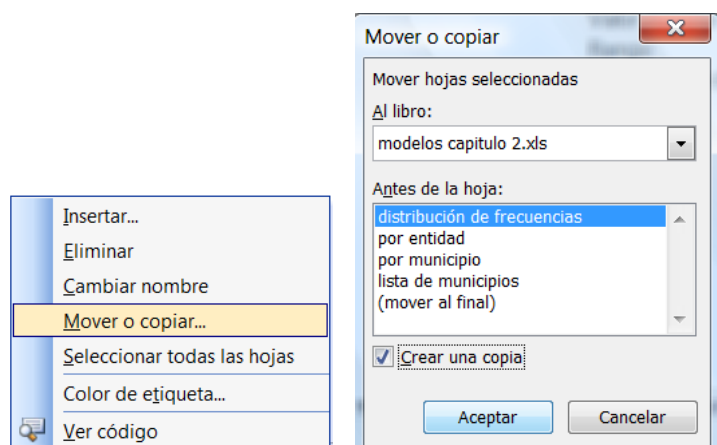
En virtud que ya se realizó la primera distribución de frecuencias del ejercicio solicitado para la tabla 7, ahora para no perder su información:

Coloque el apuntador del cursor en la pestaña correspondiente a la hoja “por entidad”

Haga clic con el botón derecho del ratón

En el menú emergente seleccione “Mover o copiar” con el botón izquierdo del ratón.

Imagen 35. Instrucciones para mover o copiar una hoja seleccionada.

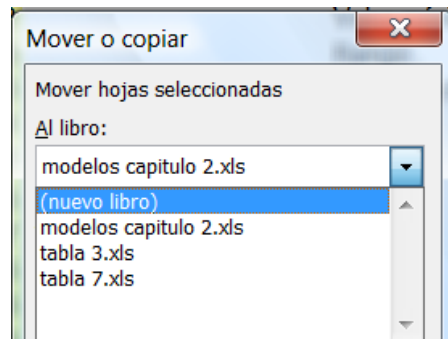


En el menú siguiente habilite “Crear una copia”

Con el apuntador del ratón haga clic en

Seleccione Nuevo Material

Imagen 36. Selección del libro a donde mover o copiar hojas seleccionadas.



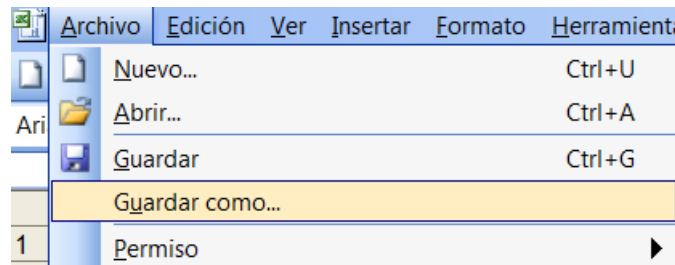
Haga clic en aceptar.

En este momento usted debe tener un archivo llamado “MaterialN”, donde N puede ser cualquier número. Para conservar su información y finalizar el ejercicio de la tabla 7, guarde el nuevo archivo con el nombre de ejercicio modelo.

Para ello, con el botón izquierdo del ratón, haga clic sobre la pestaña correspondiente al archivo materialN (el archivo recién creado por usted) que se encuentra en la parte inferior de su pantalla.

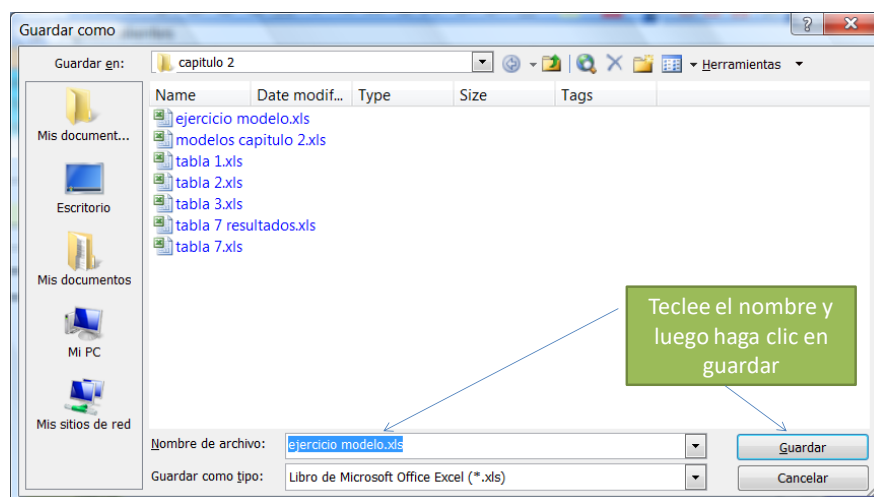
Seleccione Archivo Guardar Como

Imagen 37. Instrucciones para guardar con un nombre distinto



En la ventana emergente introduzca el nombre “ejercicio modelo”

Imagen 38. Como cambiar el nombre de archivo



Haga clic en guardar.

Si revisa el directorio correspondiente al capítulo 2, podrá observar que se ha creado un nuevo archivo llamado “ejercicio modelo.xls”

Esta serie de pasos, garantiza que el archivo modelo va a continuar sin cambios y se podrá regresar a él cada vez que usted lo desee.

El nuevo archivo creado contiene a una sola hoja llamada “por entidad” sin embargo como esta corresponde a la variable “Personas hombres que consideran que vivir en el municipio de residencia actual es inseguro” podemos renombrar la hoja y llamarla “hombres”.

Este modelo puede ocuparse cada vez que se requiera. En virtud que el ejercicio de la tabla 7 requiere que cada vez se ocupe una columna distinta, repita los pasos del modelo para cuando se tiene una variable por entidad las veces que sea necesario.

Modelo para cuando se tiene una variable por municipio.

Finalmente se presenta en las hojas “por municipio” y “lista de municipios” se presenta un modelo similar al de “por entidad” Solo que aquí se obtiene la distribución de frecuencias para cuando se tiene una variable por municipios. El modelo en cuanto tal se presenta en la hoja “por municipio” y la hoja lista de municipios es para el paso intermedio para verificar que se cuenta con la variable para todos los municipios.

Esto es, algunas veces, por ejemplo en el caso de que se tenga la variable “homicidio doloso por municipio”, no todos los municipios cuentan con observaciones y por ello no aparecerán algunos en un posible listado. Sin embargo se copia la información se pone en la hoja “lista de municipios” y se verifica uno a uno que todos los

municipios aparezcan listados. En su caso con operaciones de copiado y pegado, se van insertando aquellos municipios que no aparecen en la lista y que por ello el valor de la observación “homicidio doloso” es cero para ese municipio. No aparece en la lista entonces debe ser que tiene un valor cero.

Cuando se pegue la variable municipios en la hoja que contiene a la lista, observe que ésta está ordenada por entidad y municipio, por las claves numéricas según el catálogo del INEGI, por lo que si se tiene el cuidado de ordenar su variable en ese sentido, deberá ser muy fácil trabajar con su información. En realidad, el garantizar que se tienen todos los municipios para esta variable, insertando aquellos municipios que no están en su listado original con un valor de cero, viene a ser el paso más difícil para trabajar con este modelo; ya que una vez que se completa el listado, se copia la información a la hoja que contiene el modelo y éste se actualizará de forma inmediata y de manera automática.

Ahora bien, como se señala más atrás, al inicio de esta sección, Este modelo puede ser universal, siempre y cuando se cuente con 2 455 observaciones (el número de municipios de México en la fecha de creación de este documento) o menos.

Suponga que existe algún país latino que cuenta con 55 entidades federativas, entonces puede utilizarse el modelo por municipio, teniendo cuidado de que estén llenas sólo las filas correspondientes a esas 55 entidades, y verificando que de la fila que hubiere correspondido a la entidad 56, en adelante no exista información.

El modelo se llenó con información para ayudar a los estudiantes de México a realizar análisis delictivos que se utilizan con frecuencia. Sin embargo, ese modelo puede funcionar para cualquier país, con los ajustes correspondientes.

[illegible]

Principales métodos

Medidas de tendencia central y de variabilidad

Introducción En este apartado se hace una rápida revisión de las técnicas estadísticas utilizadas para realizar inferencias acerca de un conjunto de datos, al cual llamaremos población, basados en una muestra de dicha población.

Existen muchos textos dedicados a la enseñanza de la estadística descriptiva, la cual puede dividirse en dos métodos: gráficos y numéricos. En este texto se enfoca la atención a los métodos numéricos, restringiendo la discusión a aquellas medidas que son más útiles, no sólo con propósitos descriptivos sino de inferencia también.

Las medidas que se obtienen a partir de una muestra se llaman estadísticos, los que se obtienen a partir de una población se llaman parámetros.

Los dos tipos de medidas más importantes son las que localizan el centro de los datos y las que describen o miden la variabilidad o dispersión de los datos con respecto a la medida central. El primer tipo se conoce como Medidas de tendencia central, y el segundo como Medidas de variación o dispersión.

Medidas de tendencia central

Moda: La primera medida de tendencia central que se revisará es la moda.

Definición Moda

La moda de un conjunto de observaciones es el dato que se presenta con mayor frecuencia.

Ejemplo. Halle la moda de los siguientes datos 2, 2, 3, 2, 5, 6.

Solución:

Moda = 2 ya que es el dato que aparece con mayor frecuencia.

La moda puede no existir; o bien puede aparecer más de una moda.

Ejemplo. Halle la moda de los siguientes datos: 2, 2, 2, 3, 5, 6, 6, 7, 8, 10, 10, 10.

Solución:

En el conjunto de observaciones anteriores aparecen dos modas: 2, 10.

Si bien la moda es una de las medidas de tendencia central menos utilizada es útil cuando se pretende identificar productos con mayor demanda. Es decir se emplea cuando se quiere obtener el dato que ocurre con mayor frecuencia.

Ventajas y desventajas de la moda

Como medida de tendencia central, la moda tiene pocas ventajas sobre las demás. La principal es que es muy fácil de obtener debido a que no contiene cálculos aritméticos; en segundo lugar no existe un supuesto sobre el nivel de medición, ya que la moda puede ser utilizada en los niveles nominal, ordinal, de intervalo y de razón.

Por ejemplo, si se consulta en Internet a la tabla de delitos por delegación (en la sección de estadísticas del Distrito Federal, que se encuentra en estadísticas de seguridad pública) puede observarse que la moda en los delitos para el año 2005, de la delegación de Iztapalapa corresponde al delito de Robo a Transeúnte⁸, con 2 786 delitos registrados de este tipo. Es decir, es el delito que más se cometió en ese año.

Una de las principales desventajas de la moda, es que sólo considera a la categoría u observación con la mayor frecuencia, no tomando en cuenta a los demás valores.

Mediana:

La siguiente medida de tendencia central que se revisará será la mediana.

Definición Mediana, para número impar de datos

La mediana para un número impar de datos será el dato de en medio una vez que estos se han ordenado en forma creciente o decreciente. Es decir, es el dato que se encuentra en la posición obtenida por la fórmula siguiente:

$$mediana = \frac{n+1}{2} \quad \text{dato en la posición}$$

Ejemplo

Encuentre la mediana para los siguientes datos 3, 10, 11, 6, 5.

Solución:

Primero se ordenan los datos 3, 5, 6, 10, 11. Y la mediana es el tercer dato en virtud que el número de observaciones es impar.

$$\text{Mediana} = 6$$

Se puede decir que la mediana divide a los datos en dos partes iguales, la mitad de los datos están por abajo de la mediana y la otra mitad por encima.

⁸ No se considera al rubro Otras denuncias, en virtud que esta conformado por diversos delitos, que se agrupan en dicha clasificación.

Definición Mediana para número par de datos

La mediana para datos pares es el promedio de las dos observaciones centrales, cuando los datos están ordenados en forma ascendente o descendente.

Ejemplo

Obtenga la mediana para los siguientes datos. 3, 10, 11, 6, 5, 7.

Solución:

1. Se ordenan los datos en forma ascendente, 3, 5, 6, 7, 10, 11.
2. Se localizan los dos valores de las observaciones centrales:
6, 7
3. Se obtiene el promedio el cual corresponde a la mediana.

$$\text{Mediana} = \frac{6 + 7}{2} = \frac{13}{2} = 6.5$$

Ventajas y desventajas de la mediana

Una de las desventajas de la mediana, es como con la moda, que no se utilizan todas las observaciones, en virtud de que la mediana sólo está interesada en identificar a aquel dato u observación que corresponde a la posición de en medio (el 50 percentil), Así, al obtener la mediana no se considera si existen valores extremos altos o bajos, esto implica que la mediana no se ve afectada por dichos valores.

21	6231
750	6325
1500	6451
6650 Mediana	6650 Mediana
15219	6782
95812	6897
112540	6954

En la tabla anterior se observan dos distribuciones de datos, donde la mediana es la misma, el valor central, el que ocupa la posición de en medio, sin importar que en la primera distribución existen valores mínimos y máximos extremos, mientras que la segunda distribución es más uniforme.

Entonces una de las principales ventajas de la mediana es que no es sensible a los valores extremos, por lo que se utiliza, sobre todo, cuando se tienen distribuciones asimétricas.

Media aritmética

Una de las medidas de tendencia central más usada en el análisis delictivo es la media aritmética o promedio aritmético.

Definición: Media aritmética

La media aritmética para un conjunto de observaciones se define como la suma de éstas dividida entre el número de observaciones

Ejemplo

Encuentre la media aritmética (generalmente se dice la media) de los siguientes valores: 2, 3, 4, 5, 6.

Solución:

Si se denota a la media aritmética como $\bar{x}$ (léase x barra) entonces:

$$\bar{x} = \frac{2+3+4+5+6}{5} = \frac{\text{suma de valores}}{\text{número de valores}} = 4$$

En general para calcular ésta y otras medidas se hace conveniente expresarlas mediante fórmulas. Para esto se introduce el símbolo $\sum$ (sumatoria) que representará el proceso de suma. Si se consideran las N cantidades a ser sumadas $X_1, X_2, X_3, \dots, X_n$.

Entonces $\sum_{i=1}^n X_i$ representa la suma.

Así

$$\sum_{i=1}^3 X_i = X_1 + X_2 + X_3$$

$$\sum_{i=1}^5 X_i = X_1 + X_2 + X_3 + X_4 + X_5$$

Utilizando esta notación se puede representar la fórmula de la media aritmética como sigue:

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

n = número de datos disponibles

Si bien la media aritmética para una muestra o una población se definen de igual manera. Se utiliza el símbolo $\bar{x}$ para la media muestral y μ para denotar la media poblacional.

Ejemplo

Halle la media de los siguientes datos: 3, 10, 11, 6, 5.

Solución:

$$\bar{x} = \frac{\sum_{i=1}^5 x_i}{n} = \frac{3+10+11+6+5}{5} = \frac{35}{5} = 7$$

Ventajas y desventajas de la media aritmética

La media utiliza todos los datos disponibles, por lo que es más consistente que la moda y la mediana. Otra gran ventaja es que contiene propiedades matemáticas que la vuelven deseable, sobre todo al estudiar estadística más avanzada.

Una de estas propiedades es que la suma de las diferencias entre cada observación y la media de dichas observaciones es cero.

Tabla 25. Delitos registrados por la PGJ DF en Iztapalapa en el año 2005

Delito	$\bar{x}$	$x_i - \bar{x}$	diferencia
Otras Denuncias	9685	9685 - 1707.9	7977.14
Robo A Transeúnte	2786	2786 - 1707.9	1078.14
Robo C/V Vehículo	2492	2492 - 1707.9	784.14
Lesiones Dolosas	2441	2441 - 1707.9	733.14
Robo S/V Vehículo	2231	2231 - 1707.9	523.14
Robo A Transporte	1252	1252 - 1707.9	-455.86
Robo S/V Negocio	969	969 - 1707.9	-738.86
Robo S/V Casa Hab.	839	839 - 1707.9	-868.86
Robo C/V Negocio	533	533 - 1707.9	-1174.86
Despojos	240	240 - 1707.9	-1467.86
Violación	230	230 - 1707.9	-1477.86
Homicidio Doloso	131	131 - 1707.9	-1576.86
Robo C/V Casa Hab.	74	74 - 1707.9	-1633.86
Robo A Banco	7	7 - 1707.9	-1700.86
promedio	1707.9		0.00

Es decir que $\sum(x_i - \bar{x}) = 0$

Esta propiedad se utiliza en análisis estadísticos más complejos.

Paso a Paso con Excel:

Medidas de tendencia central.

Ayuda para el maestro

Media aritmética o promedio

En la pagina internet

www.seguridadpublicaenmexico.org.mx seleccione

Proyecto datos México

Estadísticas sobre seguridad pública

Distrito Federal

Delitos por delegación

Esta secuencia de instrucciones hace que se abra un archivo en Excel

En dicho archivo en el encabezado de la tabla dinámica seleccione el año 2007, de tal forma que se muestre la información como se indica en la siguiente tabla.

Imagen 39. Evolución de los delitos denunciados

Fuente: Elaboración propia con base en información de la Procuraduría General de Justicia del Distrito Federal	
Solicitud de información a través del IFAI	
Evolución de los delitos denunciados	
MES	(Todas)
AÑO	2007
Suma de Datos2	Delito
DELEGACION	Robo C/V Vehículo Robo S/V Vehículo Total general
Alvaro Obregón	262 500 762
Azcapotzalco	214 427 641
Benito Juárez	618 797 1415
Coyoacán	436 687 1123
Cuajimalpa	17 77 94
Cuauhtémoc	380 608 988
Gustavo A. Madero	759 1110 1869
Iztacalco	324 286 610
Iztapalapa	945 938 1883
Magdalena Contreras	41 94 135
Miguel Hidalgo	341 547 888
Milpa Alta	13 7 20
Tlalhuac	93 221 314
Tlalpan	389 352 741
Venustiano Carranza	368 321 689
Xochimilco	130 140 270
Total general	5330 7112 12442

En esta puede verse en las columnas C y D, el delito de robo de vehículo con o sin violencia, por delegación para el año 2007.

Es decir que existen 16 observaciones, una para cada delegación.

Para obtener la media de robo de vehículo con violencia, se obtiene el total de robos (5 330, celda c29) y se divide entre el número de observaciones (16 delegaciones) esto da el valor de 333.1.

Para que Excel haga el cálculo, simplemente coloque el ratón en cualquier celda en blanco, por ejemplo en la celda C32 y escriba en dicha celda =PROMEDIO(C13:C28), presione enter y aparecerá el valor 333.1

Si coloca el cursor en la celda C13 y ordena los datos de mayor a menor, encontrará que 9 delegaciones tienen valores por encima de la media

Imagen 40. Robo de vehículo con violencia, en datos brutos

DELEGACION	Robo C/V Vehiculo
Iztapalapa	945
Gustavo A. Madero	759
Benito Juárez	618
Coyoacán	436
Tlalpan	389
Cuauhtémoc	380
Venustiano Carranza	368
Miguel Hidalgo	341
Iztacalco	324
Alvaro Obregón	262
Azcapotzalco	214
Xochimilco	130
Tlahuac	93
Magdalena Contreras	41
Cuajimalpa	17
Milpa Alta	13
Total general	5330

333.1

Las que se muestran resaltadas en la gráfica.

Pero recuérdese que aquí se ha señalado que todos los cálculos deben hacerse con relación a la base teórica que es probable sea victimizada.

En este caso, el cálculo debería realizarse sobre el total de vehículos en circulación en el año 2007, valor que es muy difícil de averiguar. Así que podría utilizarse como variable aproximada “proxy” al número de vehículos registrados por delegación, o al total de viviendas particulares habitadas que disponen de automóvil o camioneta propia, por delegación.

Se conoce como variable proxy en virtud que no contempla a la totalidad de la población en riesgo, sin embargo aproxima mejor a lo que se desea obtener. Obsérvese que no se toma en cuenta a los negocios en cada delegación que disponen de automóvil o camioneta propia y que teóricamente también están en riesgo, o que tampoco se toma en cuenta a los vehículos que no

perteneciendo a la delegación, en un momento determinado del día circulan por las calles de una delegación en particular.

Algunos analistas no toman en cuenta esta situación de obtener denominadores adecuados y a todos los delitos los dividen entre la población y la multiplican por 100 mil

Hagamos el cálculo, desde diferentes perspectivas.

Tabla 26. Robo de vehículo por 100 mil habitantes y por 10 mil viviendas con vehículo particular.

DELEGACION	Robo de vehículo con violencia				
	Vehículos	Población	Viv. Con veh.	Robo Veh por 100 mil habitantes	Robo Veh por 10 mil Viv con veh
Alvaro Obregon	262	686807	64992	38	40
Azcapotzalco	214	441008	42258	49	51
Benito Juarez	618	360478	68954	171	90
Coyoacan	436	640423	86949	68	50
Cuajimalpa	17	149743	13694	11	12
Cuauhtemoc	380	516255	46647	74	81
Gustavo A. Madero	759	1235542	104892	61	72
Iztacalco	324	411321	36653	79	88
Iztapalapa	945	1773343	128656	53	73
Magdalena Contreras	41	221645	18692	18	22
Miguel Hidalgo	341	352640	45482	97	75
Milpa Alta	13	90772	5484	14	24
Tiahuac	93	301317	20980	31	44
Tlalpan	389	576172	64337	68	60
Venustiano Carranza	368	462806	39121	80	94
Xochimilco	130	364647	27754	36	47
Total general	5330	8584919	815545	62	65
	333			59	58

Obsérvese que hay casos como el de la delegación Benito Juárez, donde al calcular por habitante se obtiene una cifra de 171 vehículos robados con violencia por cada 100 mil habitantes, mientras que si se calcula contra las viviendas particulares con vehículo propio se pasa a 90, cifra que nos parece más adecuada.

En todo caso en la medida de lo posible, se debe intentar realizar los cálculos contra denominadores adecuados, que contemplen a toda la población teórica en riesgo, de otra forma se corre el riesgo de generar modelos de intervención inadecuados.

Obsérvese la siguiente tabla donde se colocan a las delegaciones en orden de mayor a menor, según la cantidad de vehículos robados con violencia en ellas, desde diferentes perspectivas.

Tabla 27. Orden que guardan las delegaciones en el delito de robo de vehículo con violencia, según datos brutos, por 100 mil habitantes o por 10 mil viviendas con vehículo particular

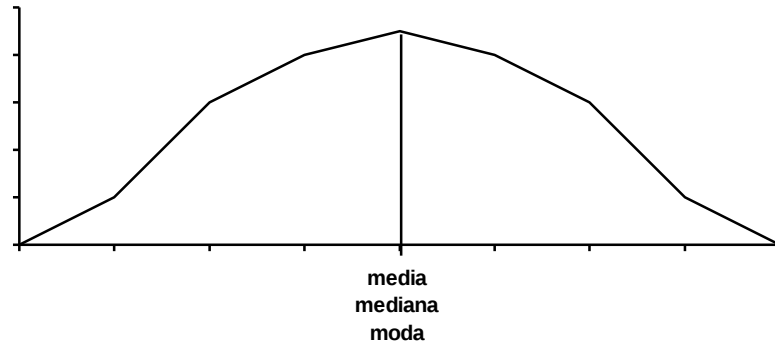
Datos brutos	por 100 mil hab	Por 10 mil veh
Iztapalapa	Cuauhtemoc	Cuauhtemoc
Gustavo A. Madero	Benito Juarez	Venustiano Carranza
Coyoacan	Miguel Hidalgo	Iztacalco
Benito Juarez	Venustiano Carranza	Iztapalapa
Cuauhtemoc	Iztacalco	Miguel Hidalgo
Venustiano Carranza	Coyoacan	Gustavo A. Madero
Miguel Hidalgo	Iztapalapa	Benito Juarez
Iztacalco	Gustavo A. Madero	Azcapotzalco
Tlalpan	Azcapotzalco	Coyoacan
Alvaro Obregon	Tlalpan	Tlalpan
Azcapotzalco	Alvaro Obregon	Tlahuac
Tlahuac	Tlahuac	Alvaro Obregon
Xochimilco	Xochimilco	Xochimilco
Magdalena Contreras	Magdalena Contreras	Magdalena Contreras
Cuajimalpa	Milpa Alta	Milpa Alta
Milpa Alta	Cuajimalpa	Cuajimalpa

En donde puede verse por ejemplo que en datos brutos Iztapalapa es la primera en robo de vehículo con violencia, lo cual es obvio dado que dicha delegación es la que cuenta con el mayor número de viviendas particulares habitadas que disponen de automóvil o camioneta propia, lo cual es lógico ya que es en esa delegación donde vive, proporcionalmente la mayor cantidad de población, ya que más de uno de cada cinco habitantes del D. F. viven en esa delegación.

Relación entre la media, mediana y moda

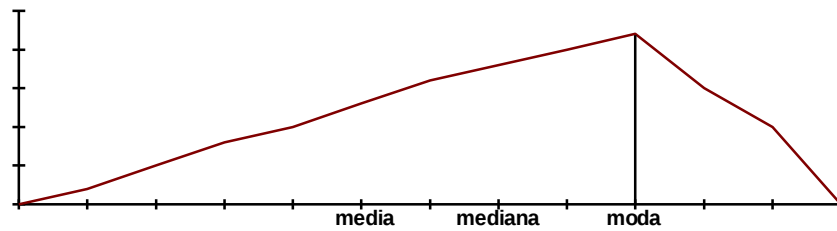
a) Distribución simétrica.

Imagen 41. Media, mediana y moda son idénticas.



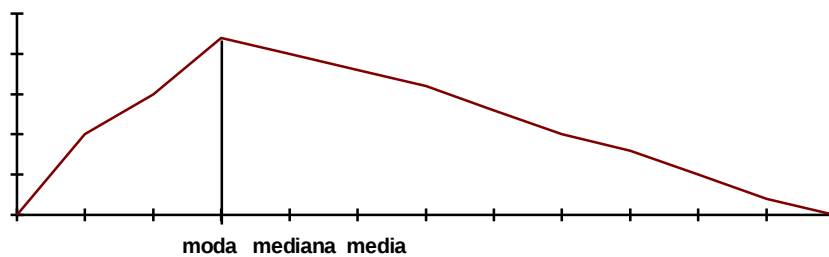
b) Distribución asimétrica cargada hacia la izquierda.
(negativamente asimétrica)

Imagen 42. $\text{media} < \text{mediana} < \text{moda}$.



c) Distribución asimétrica cargada hacia la derecha.
(positivamente asimétrica)

Imagen 43. $\text{Moda} < \text{mediana} < \text{media}$



Ejercicio 8

¿Bajo qué condiciones son iguales la media, la moda y la mediana?

De las tres medidas de tendencia central estudiadas.

- a) ¿Cuál es la más sensible a valores extremos?
- b) ¿Cuál es la menos sensible?

Un aspirante a ingresar a la industria de televisores desea saber:

- a) ¿Cuál es el producto de mayor venta? ¿Estará interesado en la media, mediana y moda?, ¿Por qué?
- b) Le interesa seleccionar la entidad del país donde el producto tiene mayor probabilidad de venderse. ¿Le interesa la moda o la mediana del ingreso familiar? ¿Por qué?
- c) Quiere determinar además la posibilidad de vender en plazos mediante apoyo financiero bancario pagadero por los familiares. ¿Le interesa la moda, la mediana o la media de las tasas de interés cobrada por las diferentes instituciones? ¿Por qué?

Dadas las ventas de autos grandes, ocho cilindros; medianos, 6 cilindros; y chicos, 4 cilindros.

- a) ¿Cómo esperaría encontrar la distribución (simétrica, asimétrica)?
- b) ¿Cómo esperaría encontrar al ingreso por auto vendido (simétrico, asimétrico)?
- c) ¿Cómo esperaría encontrar el consumo de gasolina de los vehículos producidos por Volkswagen (simétrico, asimétrico)?

Considerando a la siguiente tabla:

Tabla 28. Homicidio doloso y robo con violencia a casa habitación según la PGJ DF en el año 2005, en datos brutos.

<i>Delegación</i>	<i>Homicidio Doloso</i>	<i>Robo C/V Casa Hab.</i>
Alvaro Obregon	52	44
Azcapotzalco	23	34
Benito Juarez	34	44
Coyoacan	21	39
Cuajimalpa	13	8
Cuauhtemoc	100	53
Gustavo A. Madero	107	73
Iztacalco	30	21
Iztapalapa	131	74
Magdalena Contreras	16	16
Miguel Hidalgo	36	43
Milpa Alta	15	0
Tlahuac	21	17
Tlalpan	35	40
Venustiano Carranza	40	22
Xochimilco	25	27

Calcule la media, la mediana, la moda, para cada delito.

Señale que delegaciones están por encima o por debajo del promedio para cada delito.

Como seleccionar la medida de tendencia central apropiada

Hasta aquí debe estarse preguntando ¿Cuál será la medida apropiada que debo utilizar para expresar mejor la distribución de datos en análisis?. Esto depende de dos cuestiones: 1. El nivel de medición que tiene la información y; 2. La forma de asimetría que presenta la información.

La moda puede ser utilizada como medida de tendencia central con datos en el nivel nominal, debido a que solo se necesita saber cual categoría es la que tiene mayor frecuencia. Así, podemos señalar que de entre los diferentes delitos registrados por la PGJ DF, como son robo, lesiones, homicidio, etc. el delito de robo es el que presenta mayor frecuencia. En este caso, cuando sólo se requiere saber la categoría más alta, la moda es la adecuada.

Al nivel nominal, la moda es la más adecuada.

La mediana requiere que el nivel de medición de la variable en estudio sea ordinal o mayor. Para localizar la mediana primero debe ordenarse la información, de tal forma que se pueda

determinar que una categoría está por encima o por debajo de las características de alguna otra categoría.

La media requiere que la información este en el nivel de medición de intervalo o de razón. De hecho cuando la información tiene estos dos niveles de medición puede utilizarse cualquier medida de tendencia central. Así la pregunta que surge es ¿cuál medida utilizar?

Si se desea utilizar sólo alguna de las tres medidas aquí estudiadas, cual utilizar depende de la forma que tenga la distribución de nuestra información. Si la forma es perfectamente simétrica como la mostrada en la imagen 41, entonces la moda, la mediana y la media, serán idénticas.

Se puede ver que en este caso de perfecta simetría, que en el punto más alto de la curva se encuentra el dato o categoría con mayor frecuencia (la moda); asimismo ese punto es el que divide a la distribución en dos partes iguales (la mediana) y que es el promedio aritmético de todas las observaciones (la media)

Si la distribución de la información no es simétrica, esto es esta cargada hacia la izquierda (negativamente asimétrica) o a la derecha (positivamente asimétrica) la moda, la mediana y la media no coinciden y dependiendo de hacia que lado se encuentre la asimetría, esto determina la selección de la medida de tendencia central adecuada.

En la imagen 42, se tiene que la “cola” de asimetría esta a la izquierda, lo que implica que la media es menor que la mediana. Es decir que existen valores extremos muy bajos.

En la imagen 43, la “cola” de asimetría esta a la derecha, lo que implica que la media es mayor que la mediana. Lo que implica que existen valores extremos muy altos.

Medidas de variabilidad

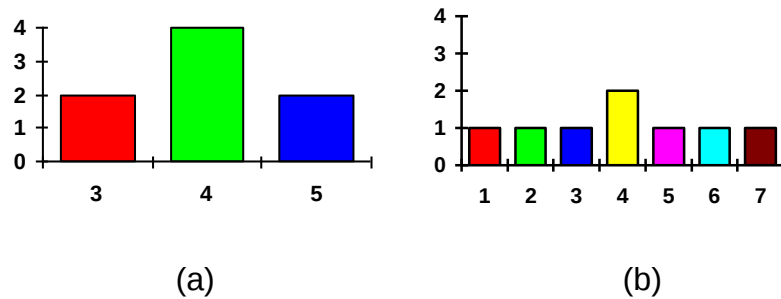
Las tres medidas de tendencia central: media, mediana y moda, nos permiten localizar el “centro” de un conjunto de datos pero no nos dicen nada acerca de la dispersión o variación que dicho conjunto de datos presenta; así, una vez que se localiza el centro en un conjunto de observaciones, se debe dar una medida de la variabilidad que los datos presentan, ya que puede darse el caso de tener una misma medida de tendencia central y una gran dispersión en diferentes conjuntos de datos.

Considérese a las dos series de datos siguientes.

a) 3, 3, 4, 4, 4, 4, 5, 5

b) 1, 2, 3, 4, 4, 5, 6, 7

Imagen 44. Conjuntos de datos con igual media y diferente variabilidad



En el conjunto a) tenemos 8 datos con una media $\bar{x} = 4$ y los datos van de 3 a 5. En el conjunto b) tenemos igualmente 8 datos, con una media $\bar{x} = 4$ pero la variación de los datos es mayor, ya que éstos van de 1 a 7, lo cual puede verse en los gráficos 2.a y 2.b.

Por lo anterior vemos que para describir un conjunto de datos no basta con obtener las medidas de tendencia central sino que además es necesario contar con unos indicadores de cuan concentrados o dispersos se encuentran los datos.

Rango

La medida de variación más simple es el rango.

definición Rango

El rango de un conjunto de observaciones $x_1, x_2, \dots, x_n$, es la diferencia entre el mayor y el menor de los datos

$$\text{rango} = \text{dato con mayor valor} - \text{dato con menor valor}$$

Ejemplo Halle el rango de los siguientes datos: 3, 3, 4, 4, 4, 4, 5, 5

Solución:

Como los datos varían de 3 a 5 entonces $\text{rango} = 5 - 3 = 2$

Ejemplo. Halle el rango de los siguientes datos: 1, 2, 3, 4, 5, 6, 7

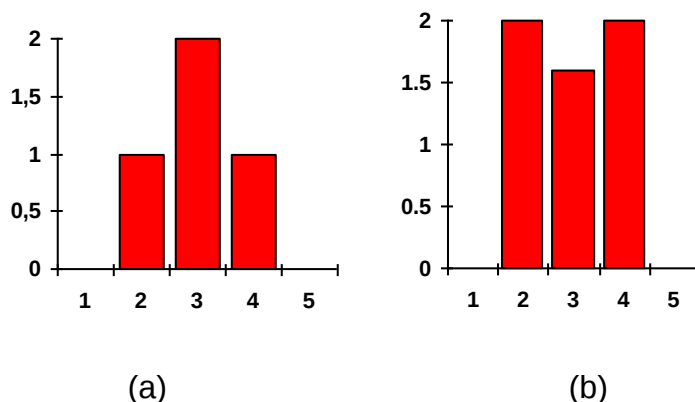
Solución:

Los datos varían de 1 a 7 entonces $\text{rango} = 7 - 1 = 6$

Cuartiles y percentiles

Desgraciadamente el rango no es una medida adecuada ya que podemos encontrar conjuntos de datos como los que se muestran a continuación:

Imagen 45. Conjuntos de datos con rango igual y diferente variabilidad



En ambas distribuciones el rango es el mismo pero en la gráfica 3.b hay menos variabilidad que en la gráfica 3.a. Para reducir este problema se introducen los cuartiles y percentiles, entre otras medidas de variabilidad.

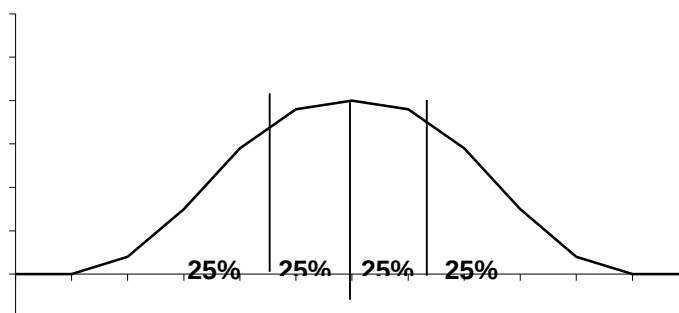
Cuartiles

definición Cuartil menor y cuartil mayor

Si tenemos $x_1, x_2, \dots, x_n$, datos ordenados de forma ascendente, el cuartil menor es el valor de x , en tal forma que, al menos $\frac{1}{4}$ de los datos son menores que x , y al menos $\frac{3}{4}$ de los datos son mayores que x . De igual forma, el cuartil mayor es el valor de x de tal forma que al menos $\frac{3}{4}$ de los datos son menores que x , y $\frac{1}{4}$ de los datos son mayores que x .

Los cuartiles se pueden definir, además, como aquellos valores de x que dividen el área bajo una curva en 4 partes iguales. Sólo se obtienen el primer y tercer cuartil, ya que la mediana es el segundo cuartil o el quincuagésimo percentil.

Imagen 46. Área bajo una curva dividida en partes iguales por los cuartiles menor y mayor, y la mediana



Q1 Mediana

Entonces el cuartil menor Q1 es el valor de x en el cual, al menos el 25% ($\frac{1}{4}$ de los valores) son menores que x , y el 75% ($\frac{3}{4}$ de los valores) son mayores que x . Y a la inversa, el cuartil mayor Q3 es aquel valor de x donde el 75%, ($\frac{3}{4}$ de los valores) será menor que x y el 25%, ($\frac{1}{4}$ de los valores) será mayor que x .

Para calcular los cuartiles se ordenan los datos $x_1, x_2, \dots, x_n$ en forma ascendente y el cuartil menor Q1, es el valor de x en la posición $.25(n + 1)$ y el cuartil mayor Q3 es el valor de x en la posición $.75(n + 1)$.

$Q1 = .25(n + 1)$ valor del dato en la posición

$Q3 = .75(n + 1)$ valor del dato en la posición

Como es obvio, la mediana será el Cuartil 2, siendo aquel valor que divide a los datos en 2, la mitad de los valores de los datos estarán por encima de la mediana y la mitad estará por abajo del valor de la mediana.

$\text{mediana} = Q2 = .5(n + 1)$

Cuando $.25(n + 1)$ y $.75(n + 1)$ no son enteros, entonces los cuartiles se localizan por interpolación.

Ejemplo

Localice la mediana, el cuartil menor y el cuartil mayor de los siguientes datos ($n = 8$)

3, 5, 8, 9, 10, 13, 16, 19

Solución:

La mediana es el valor del dato en la posición $.5(n + 1) = .5(9) = 4.5$, es decir la mediana se encuentra entre los datos que ocupan la posición 4 y 5.

$.5(9 + 10) = 9.5$.

Cuartil menor. El cuartil menor es el valor en la posición

$Q1 = .25(n + 1) = .25(9) = 2.25$

Es decir el cuartil menor esta a .25, $\frac{1}{4}$, de la distancia de los valores del segundo y tercer dato, ordenados en forma ascendente, así

$$Q1 = 5 + .25(8 - 5) = 5 + .25(3) = 5.75$$

Cuartil mayor. El cuartil mayor es el valor en la posición

$$Q3 = .75(n + 1) = .75(9) = 6.75$$

Es decir, el cuartil mayor esta a .75, $\frac{3}{4}$, de distancia entre el sexto y el séptimo dato, entonces

$$Q3 = 13 + .75(16 - 13) = 13 + .75(3) = 15.25$$

Rango intercuartílico. De donde obtenemos el rango intercuartílico

$$Q3 - Q1 = 15.25 - 5.75 = 9.50$$

Cuando las observaciones son muchas la ubicación de los cuartiles hace posible el detectar posibles asimetrías en la distribución de los datos.

Percentiles

definición percentiles

Si tenemos el conjunto de datos $x_1, x_2, \dots, x_n$ ordenados en forma creciente, el p percentil es el valor de x tal que p% de los datos es menor que x, y (100 - p)% es mayor que x.

Por ejemplo, el 30 percentil en un conjunto de datos es el valor x tal que 30 % de los datos están abajo de x y 70 % de los datos están por encima de x. En particular

Q1 = 25 percentil,

Mediana = 50 percentil, y

Q3 = 75 percentil

Para calcular percentiles, primero se ordenan los datos en forma creciente y el percentil p es el valor en la posición $p(n + 1)$. Si $p(n + 1)$ no es entero, entonces el percentil se obtiene por interpolación.

Ejemplo

Si $n = 120$, el 25 percentil esta en la posición

$$(2.6.a)p(n + 1) = 0.25(120 + 1) = 25(121) = 30.25$$

Como 30.25 no es entero, el 25 percentil se encuentra a $\frac{1}{4}$ parte de la distancia entre el valor 30 y el valor 31 y se obtiene como se obtuvieron los cuartiles.

Varianza.

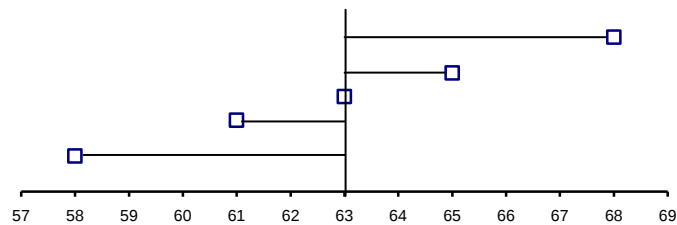
Si bien el rango es simple, no ofrece sensibilidad adecuada para describir bien la dispersión o variabilidad de un conjunto de datos. Los cuartiles y percentiles reducen un poco este problema pero se requieren más medidas para describir en forma adecuada los datos.

Entonces se hace necesario buscar una medida más sensible que nos brinde una buena imagen de los datos en estudio y que describa también cuanta variabilidad hay entre ellos.

Considérese el siguiente conjunto de datos: 58, 61, 63, 65, 68; que representan el número de horas que 5 personas gastan al mes en trasladarse de su casa al trabajo y viceversa. Estos datos podemos describirlos gráficamente en un eje y obtener la media, que puede representarse como perpendicular al eje

Imagen 47. Variabilidad de un conjunto de datos con respecto a su media

$$\bar{x} = 63$$



$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{58 + 61 + 63 + 65 + 68}{5} = \frac{315}{5} = 63$$

La varianza se obtiene a partir de las desviaciones de los datos observados con respecto a su media como (), entonces los datos observados que están arriba de la media producen desviaciones positivas y los que están abajo de la media producen desviaciones negativas.

Por esta razón la suma de las desviaciones siempre es cero.

definición Varianza

La varianza de un conjunto de datos se define como la suma de las desviaciones de los datos, con respecto a su media, elevada al cuadrado y dividida por n-1.

$$\text{Varianza} = s^2 = \frac{\sum_i (x_i - \bar{x})^2}{n - 1}$$

Esta es la medida de variabilidad más usada y en conjunto con la desviación estándar, se tienen medidas numéricas de cómo los datos tienden a variar alrededor de su media.

Desviación estándar

definición Desviación estándar

La desviación estándar es la raíz cuadrada positiva de la varianza.

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

Ejemplo . Encuentre la varianza y la desviación estándar de los siguientes datos:

5, 8, 12, 13, 15

Solución:

Tabla 29. Cálculos para obtener la varianza (ejemplo 2.11)

dato	valor	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
1	5	(5-10.6)=-5.6	31.36
2	8	(8-10.6)=-2.6	6.76
3	12	(12-10.6)=1.4	1.96
4	13	(13-10.6)=2.4	5.76
5	15	(15-10.6)=4.4	19.36
Σ	53 $\bar{x} = \frac{\sum x_i}{n} = \frac{53}{5} = 10.6$	0	65.20

Del cuadro de trabajo anterior se obtiene de inmediato la varianza:

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{65.20}{4} = 16.30$$

por lo que

$$s = \sqrt{s^2} = \sqrt{16.30} = 4.03733$$

La desviación estándar puede ser considerada como una medida de la distancia desde la media y también puede ser calculada como sigue:

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}$$

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

Para conjuntos de datos con la misma media, el conjunto que tenga mayor varianza tendrá mayor dispersión de los datos.

Ejemplo

Obtenga la desviación estándar de los ejemplos cuyos datos se muestran a continuación.

Solución:

Tabla 30. Cálculos para obtener la desviación estándar. Ejemplo (a)

a)

valor	x^2
3	9
3	9
4	16
4	16
4	16
4	16
5	25
5	25
$\Sigma x=32$	$\Sigma x^2=132$

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}$$

$$s^2 = \frac{132 - \frac{(32)^2}{8}}{8-1} = 0.57143$$

$$s = \sqrt{s^2} = \sqrt{0.57143} = 0.75593$$

Tabla 31. Cálculos para obtener la desviación estándar. Ejemplo (b)

valor	x^2
1	1
2	4
3	9
4	16
4	16
5	25
6	36
7	49
$\Sigma x=32$	$\Sigma x^2=156$

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}$$

$$s^2 = \frac{156 - \frac{(32)^2}{8}}{8-1} = 4$$

$$s = \sqrt{s^2} = \sqrt{4} = 2$$

Asimetría

La asimetría se utiliza para saber cuándo una distribución no se encuentra distribuida normalmente. Esto se hace ya que las decisiones de inferencia estadística se basan en el supuesto de distribuciones normalmente distribuidas.

La asimetría se encuentra mediante la siguiente fórmula:

$$\text{asimetria} = \frac{3(\bar{x} - \text{mediana})}{s}$$

Esta medida toma el valor de cero para una distribución simétrica, negativo para una distribución asimétrica a la izquierda, es decir que se extiende hacia valores más negativos y positivo para una distribución asimétrica a la derecha, que se extiende hacia valores más positivos. Esta medida varía entre ± 3 , aunque en escasas ocasiones va más allá de ± 1 .

Curtosis

Esta medida representa el grado de agudeza (elevación o achatamiento) que tiene una distribución, comparada contra la distribución normal.

definición 0-11 Curtosis

La curtosis se define como la mitad de la amplitud intercuartílica y la amplitud 90-10 percentil

$$\text{curtosis} = \frac{\frac{1}{2}(\text{cuartil}_3 - \text{cuartil}_1)}{\text{percentil}_{90} - \text{percentil}_{10}}$$

Para una distribución muy aguda, este valor se aproxima a 0.5; Para una distribución muy plana, este valor se aproxima 0.0; para una distribución más o menos normal la curtosis se aproxima a .25.

Gráfico de caja (box plot)

La construcción y utilización de un gráfico de caja es importante ya que, al incorporar medidas de tendencia central y de posición, como son la mediana y los cuartiles, nos permite observar mejor la variabilidad que presentan los datos y la concentración en el centro o en los extremos de la distribución.

Ejemplo

Use los datos de la siguiente tabla, para construir un gráfico de caja.

Tabla 32 Averiguaciones previas iniciadas por mes

mes	dato
ene	196
feb	206
mar	245
abr	213
may	235
jun	308
jul	281
ago	261
sep	267
oct	255
nov	286
dic	254

Solución:

1) se ordenan los datos en forma ascendente:

196, 206, 213, 235, 245, 254, 255, 261, 267, 281, 286, 308

2) Se obtiene la mediana

$$Q2 = \text{mediana} = .5(n+1) = .5(12+1) = .5(13) = 6.5$$

Es decir que la mediana está a la mitad del valor entre los datos 6 y 7, y se obtiene por interpolación.

$$\text{mediana} = .5(254+255) = 254.5$$

3) Se obtiene el Q1 y el Q3

$$Q1 = .25(n+1) = .25(13) = 3.25$$

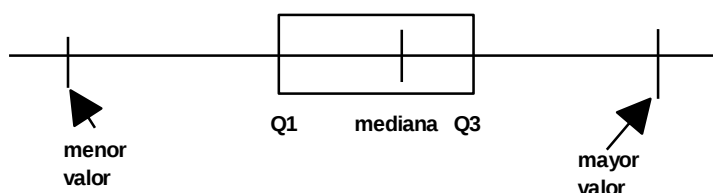
$$Q1 = 213 + .25(235 - 213) = 213 + 5.5 = 218.5$$

$$Q3 = .75(n+1) = .75(13) = 9.75$$

$$Q3 = 267 + .75(281 - 267) = 267 + 10.50 = 277.5$$

Las tres medidas obtenidas, más los valores menor y mayor de los datos nos permiten construir, como sigue, el gráfico de caja.

Imagen 48 Gráfico de caja



a) Se ubican los valores menor y mayor y se tiende una línea entre ellos.

b) Se ubican los valores correspondientes al cuartil 1, a la mediana y al cuartil 3.

c) Se dibuja un rectángulo cuyos extremos son los valores del cuartil 1 y cuartil 3.

Una vez cumplidos estos pasos se tiene un gráfico de caja. En éste podemos observar lo siguiente:

Los cuartiles menor y mayor, Q1 y Q3

$$\text{El rango intercuartílico} = (Q3 - Q1)$$

Los valores extremos menor y mayor

Si existe o no simetría en los datos

Cuando la mediana esta más cercana al cuartil 3, Q3, tenemos una distribución asimétrica cargada a la izquierda, hacia valores

más negativos (mediana > media). Cuando la mediana esta más cercana al cuartil 1, Q1, tenemos una distribución asimétrica cargada hacia la derecha, hacia valores más positivos (mediana < media).

Este gráfico de caja nos sirve también para identificar valores extremos; para esto necesitamos obtener los valores siguientes:

$$\text{frontera interior menor} = Q1 - 1.5(\text{rango intercuartílico}) \quad (a)$$

$$\text{frontera interior mayor} = Q3 + 1.5(\text{rango intercuartílico}) \quad (a)$$

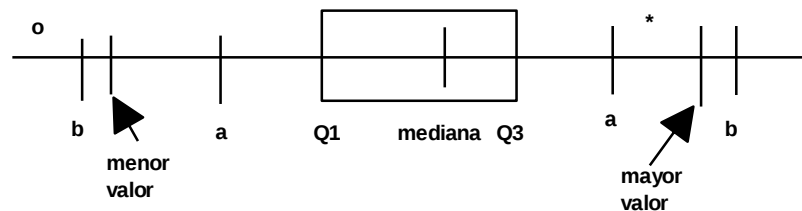
$$\text{frontera exterior menor} = Q1 - 3(\text{rango intercuartílico}) \quad (b)$$

$$\text{frontera exterior mayor} = Q3 + 3(\text{rango intercuartílico}) \quad (b)$$

Medidas que podemos observar en el gráfico de caja (imagen 49) como (a) que corresponde a la frontera interior menor y mayor y (b) que corresponde a la frontera exterior menor y mayor.

Cualquier dato cuyo valor se encuentre entre (a) y (b) está considerado como dato extremo sospechoso, y se señala con un asterisco (*). Cualquier dato cuyo valor este mas allá de (b) se conoce como dato extremo y se señala con un círculo (o).

Imagen 49 Gráfico de caja con fronteras interiores y exteriores, menores y mayores; y con valores extremos



z- score

Esta medida se utiliza para determinar, dato a dato, a cuantas desviaciones estándar se encuentra un valor x de su media $\bar{x}$. Para esto se obtiene primero la distancia existente entre x y $\bar{x}$.

$$\text{distancia} = x - \bar{x}$$

y luego convertimos esta distancia a desviaciones estándar dividiendo entre la desviación estándar

$$z_{score} = \frac{\text{distancia}}{\text{desviacion estandar}} = \frac{x - \bar{x}}{s}$$

Regla empírica

para finalizar con este apartado se enunciará lo que se conoce como regla empírica

Dado un conjunto de datos cuya distribución se aproxima a la normal, el intervalo:

definición

$(\bar{x} \pm s)$ contiene aproximadamente el 68 % de los datos

$(\bar{x} \pm 2s)$ contiene aproximadamente el 95 % de los datos

$(\bar{x} \pm 3s)$ Contiene todos o casi todos los datos

Ahora si es importante
poner atención, ya que
de aquí en adelante
puede ser que las cosas
se compliquen un poco.

Paso a Paso con Excel:

Estadística descriptiva.

Ayuda para el maestro

Media aritmética o promedio

En la pagina internet

www.seguridadpublicaenmexico.org.mx seleccione

Abra el archivo

Ejercicios de estadística descriptiva

Que se encuentra en

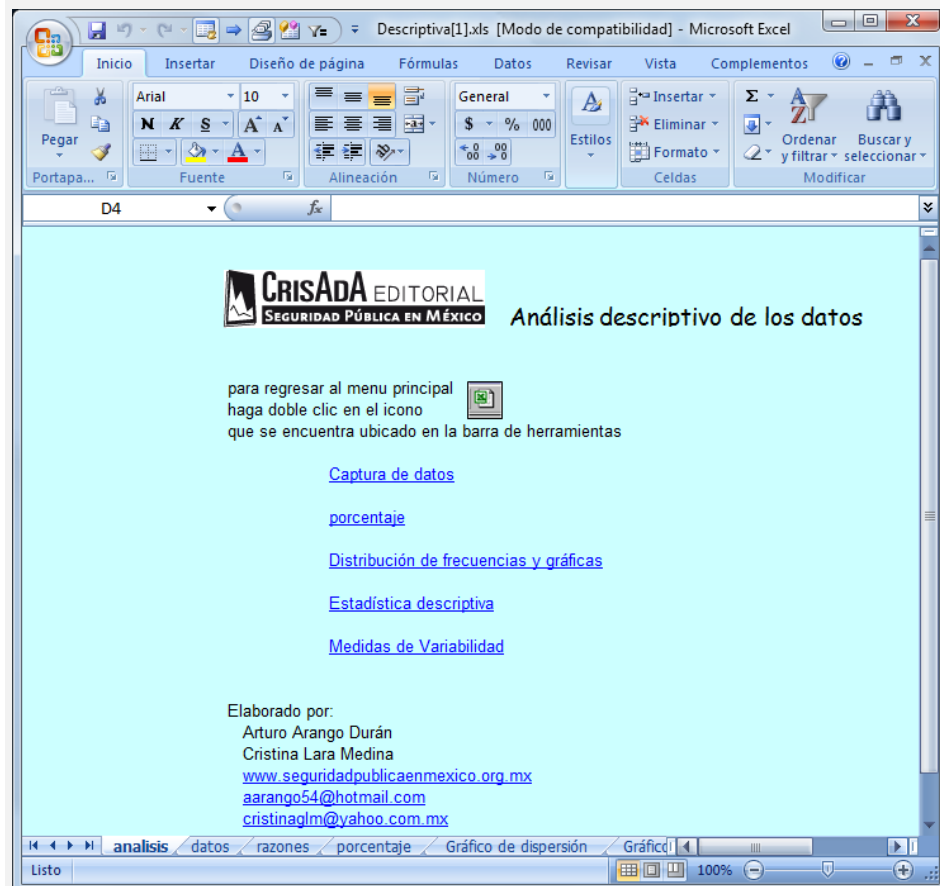
Capacitación

Métodos Estadísticos para Analistas Delictivos y Criminólogos

Cap. I. Introducción

Ejercicios de estadística descriptiva

Imagen 50. Modelo para análisis descriptivo de los datos.



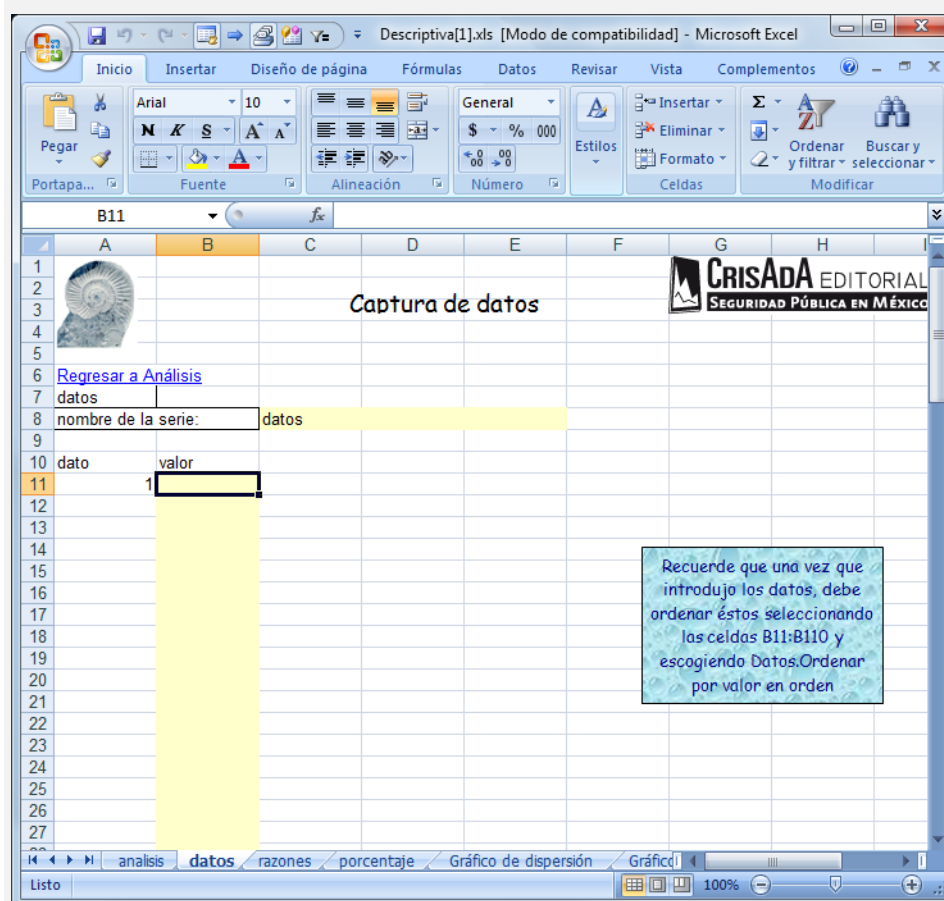
En dicho archivo se encontrará que existen varias hojas llamadas: Análisis, datos, razones, porcentaje, Gráfico de dispersión, Gráfico z score, frecuencias, estadística descriptiva y medidas de variabilidad.

Navegue un poco sobre las diferentes hojas que componen este archivo y encontrara que, con excepción de la hoja razones, las demás no contienen información ya que, o bien están en blanco o bien, tienen valores que no hacen sentido.

Sin embargo, cada una de estas hojas contiene modelos que realizan todos y cada uno de los ejercicios de este último apartado.

La forma de utilizar estos modelos es muy simple, lo único que se debe hacer es colocarse en la hoja datos, en donde se muestra la columna B, celdas b11: b110 sombreadas en color amarillo,

Imagen 51. Modelo para captura de datos



En la celda c8, se teclea el nombre de la serie de valores que se van a capturar y en las celdas b11:b110 se capturan los valores.

A Medida que se van capturando los datos, se va rellenando la columna A, con números consecutivos así como se van realizando los cálculos de forma automática en las hojas de porcentaje, Gráfico de dispersión, Gráfico z score, frecuencias, estadística descriptiva y medidas de variabilidad.

Probemos introduciendo los datos del ejemplo de distribución de frecuencias, que se repiten a continuación, por comodidad.

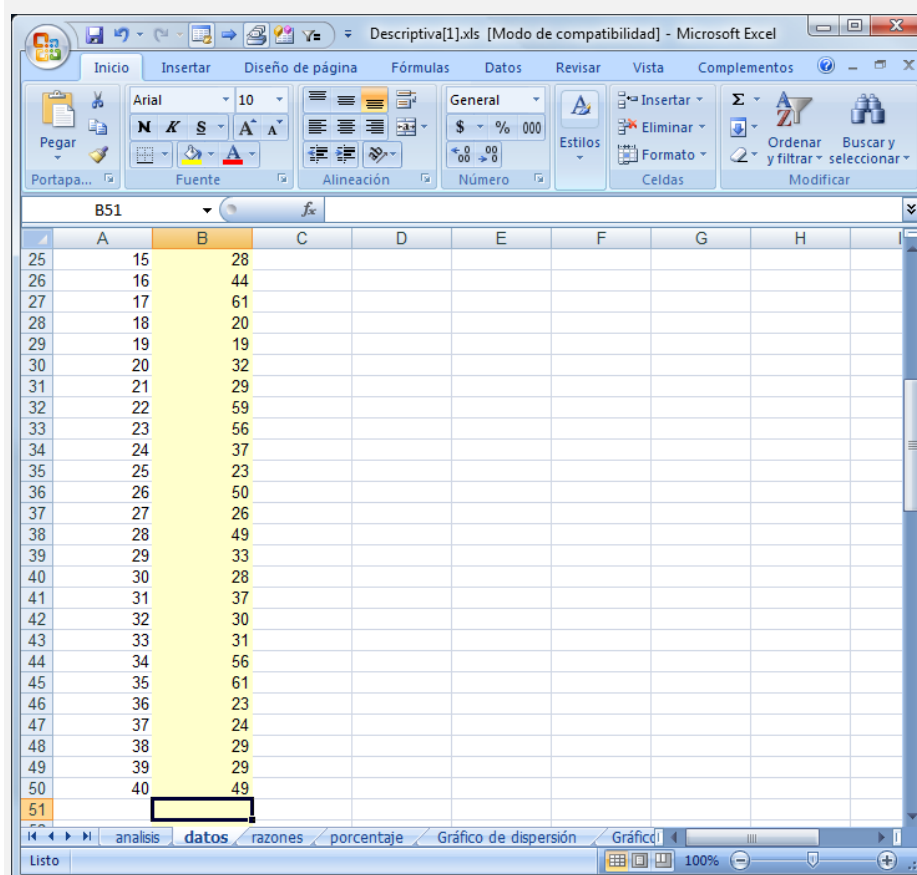
Tabla 33. Robo a transeúnte. Ejemplo.

53	32	61	23	31
54	31	20	50	56
37	60	19	26	61
58	36	32	49	23
56	61	29	33	24
53	28	59	28	29
55	28	56	37	29
46	44	37	30	49

Para ello coloque el cursor en la celda b11 y teclee el valor 53, luego en la celda b12, el valor 54 y así sucesivamente hasta introducir todos los valores.

La pantalla debe verse así:

Imagen 52. Pantalla de captura de datos

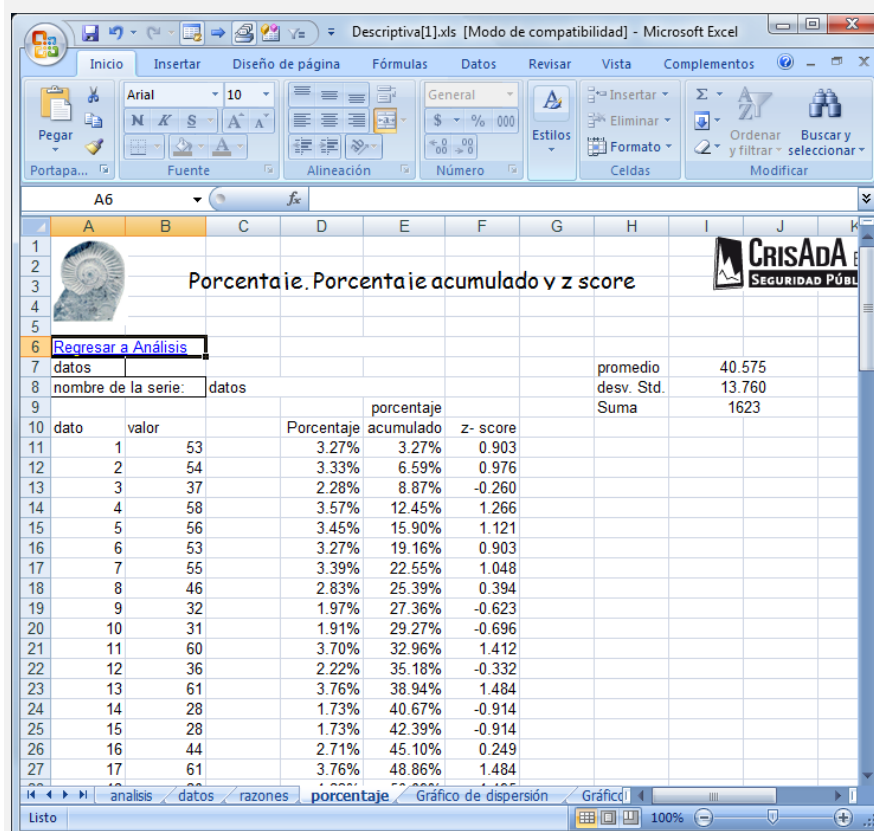


Ahora seleccione cualquiera de las hojas de porcentaje, Gráfico de dispersión, Gráfico z score, frecuencias, estadística descriptiva y medidas de variabilidad y observe que la información contenida en

ellas corresponde a la información tecleada en la hoja datos, pero ya cada una contiene las diferentes fórmulas que hemos revisado en todo este material.

Así la hoja porcentaje se ve así:

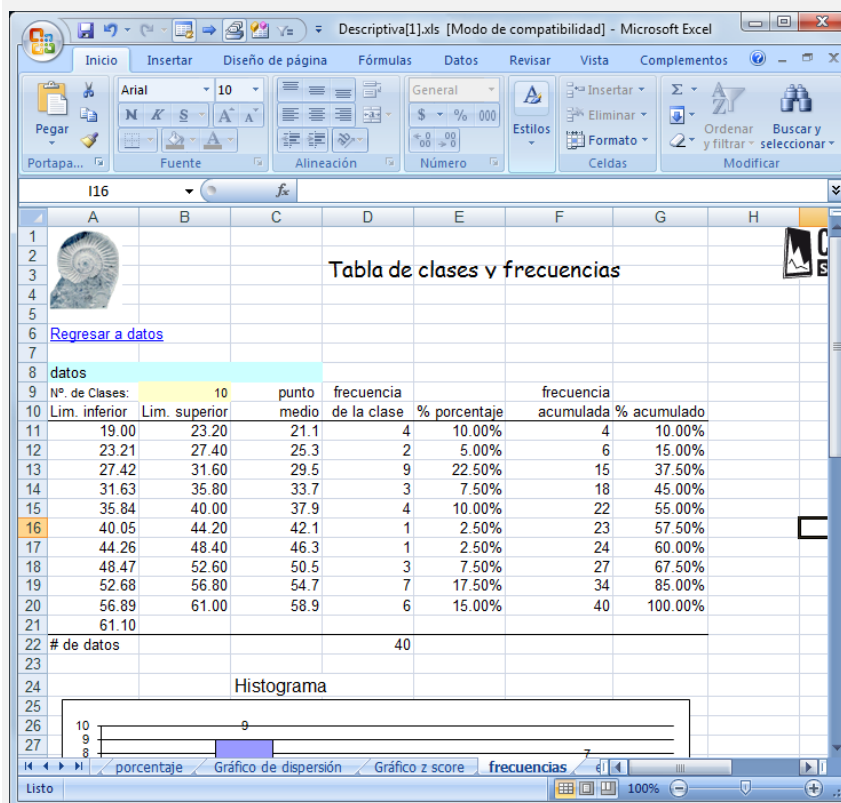
Imagen 53. Modelo de porcentaje acumulado y z-score



Porcentaje, Porcentaje acumulado y z score				
Regresar a Análisis				
datos				
nombre de la serie: datos				
promedio 40.575				
desv. Std. 13.760				
Suma 1623				
dato	valor	Porcentaje	Porcentaje acumulado	z-score
1	53	3.27%	3.27%	0.903
2	54	3.33%	6.59%	0.976
3	37	2.28%	8.87%	-0.260
4	58	3.57%	12.45%	1.266
5	56	3.45%	15.90%	1.121
6	53	3.27%	19.16%	0.903
7	55	3.39%	22.55%	1.048
8	46	2.83%	25.39%	0.394
9	32	1.97%	27.36%	-0.623
10	31	1.91%	29.27%	-0.696
11	60	3.70%	32.96%	1.412
12	36	2.22%	35.18%	-0.332
13	61	3.76%	38.94%	1.484
14	28	1.73%	40.67%	-0.914
15	28	1.73%	42.39%	-0.914
16	44	2.71%	45.10%	0.249
17	61	3.76%	48.86%	1.484

Y la de frecuencias, ya contiene información que hace sentido, como se muestra a continuación:

Imagen 54. Modelo de distribución de frecuencias



Lo interesante de este modelo, es que se puede utilizar cualquier tipo de información en línea o se puede utilizar como base para realizar cualquier análisis por complicado que éste sea; ya que lo único que se requiere es localizar en la red Internet la información que se desea utilizar, copiarla y pegarla en la hoja datos, sobre las celdas amarillas.

El modelo se encarga del resto.

Resumen

Para saber el monto promedio de sentencia otorgada a los delincuentes por el delito de robo a transeúnte, se realizó una muestra en donde se estudio a 30 delincuentes y las sentencias, obteniendo éstas en meses. Esto implica que el nivel de medición es de razón, de tal forma que se pueden realizar todos los cálculos hasta aquí estudiados, sobre estas sentencias.

Sentencia en meses (datos hipotéticos)

102 93 102 102 81 93 65 93 81 67 65 93 72 67 65 72 81 66
93 66 65 72 65 72 66 72 72 63 62 66

Paso 1. Organizar los datos.

En virtud que existen 30 observaciones, es fácil organizar estos datos en clases, como se muestra:

X	f
62	1
63	1
65	5
66	4
67	2
72	6
81	3
93	5
102	3
N = 30	

Paso 2. Encontrar la moda

La moda es la frecuencia que más se repite. En este caso y dado que la sentencia es en meses (X) la distribución es unimodal, es decir que existe una sola moda. La sentencia en meses que más se repite es la de 72 meses. La moda es 72 meses.

Paso 3. Encontrar la mediana

La mediana es aquel dato que divide a las observaciones exactamente en dos. En virtud que se tiene a 30 observaciones, la mediana es el promedio de la observación 15 y 16. Para ello se ordenan los datos en orden ascendente y encontramos que el valor 72 es el que se encuentre en dichas observaciones, por lo que la mediana es 72 meses de sentencia.

Paso 4. Encontrar la media

La media corresponde al promedio aritmético y puede encontrarse con la fórmula siguiente:

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

Utilizando Excel, encontramos que la suma de los meses es de 2294, por lo que $\bar{x} = \frac{2294}{30} = 76.5$

Es decir que en promedio se sentencia a 76.5 meses de cárcel.

Paso 5. Encontrar el rango

Para encontrar el rango reste el valor mínimo del valor máximo.

Es decir $102 - 62 = 40$. Por lo que el rango es de 40 meses de sentencia.

Paso 6. Encontrar la varianza.

La varianza es la suma de las desviaciones de los datos, con respecto a su media, elevada al cuadrado y dividida por $n-1$.

$$Varianza = S^2 = \frac{\sum_i (x_i - \bar{x})^2}{n-1}$$

Utilizando Excel se pueden realizar rápidamente los cálculos.

Observación	Sentencia	$x - \bar{x}$	$(x - \bar{x})^2$
1	62	62 - 76.5	209.3
2	63	63 - 76.5	181.4
3	65	65 - 76.5	131.5
4	65	65 - 76.5	131.5
5	65	65 - 76.5	131.5
...			
...			
...			
27	93	93 - 76.5	273.4
28	102	102 - 76.5	652.0
29	102	102 - 76.5	652.0
30	102	102 - 76.5	652.0

Los cálculos que se muestran son aproximados a un decimal, por lo que la varianza es aproximada.

La suma de las desviaciones al cuadrado es de

5245.9

de tal forma que

$$Varianza = S^2 = \frac{5245.9}{29} = 180.89$$

En Excel se puede utilizar la fórmula predefinida para el cálculo de la varianza de forma automática, con la siguiente sintaxis:

VAR(número1;número2; ...)

Número1, número2,... son de 1 a 255 argumentos numéricos correspondientes a una muestra de una población

Utilizando el comando de insertar función.

Paso 7. Encontrar la desviación estándar

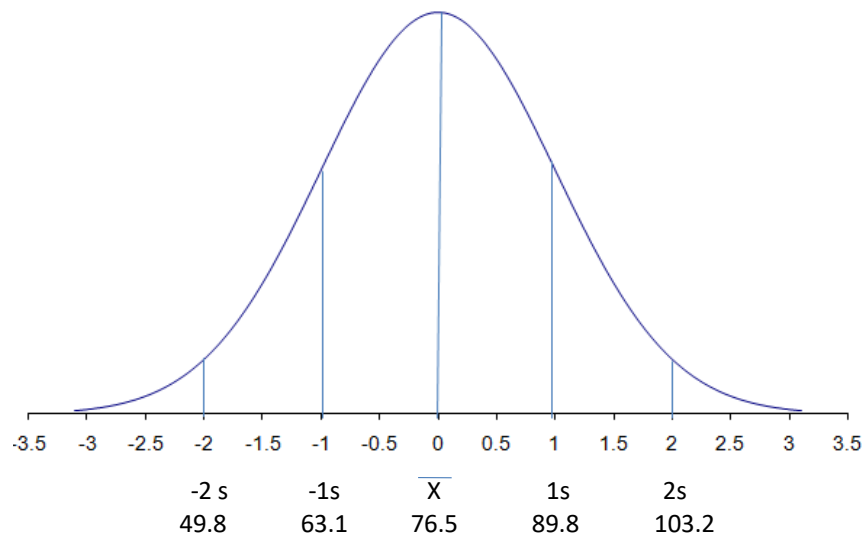
En virtud que la desviación estándar es la raíz cuadrada positiva de la varianza, entonces

$$s = \sqrt{180.89} = 13.4$$

Esto implica que hay una desviación promedio de 13.4 meses entre las sentencias de cada delincuente y el promedio de 76.5 meses de sentencia.

Paso 8. Describir las observaciones

Encontramos que las sentencias están casi normalmente distribuidas (simétricamente) ya que la moda es de 72 meses, la mediana es de 72 meses y la media es de 76.5 meses, con una desviación estándar de 13.4 meses que nos indica que la dispersión de una anchura aceptable.



Sentencia en meses

Así encontramos que 21 delincuentes sentenciados caen entre -1 a 1 desviaciones estándar con respecto de la media; sólo 1 delincuente cae por debajo de -1 desviaciones estándar con respecto de la media (-1.1), mientras que los ocho delincuentes restantes de los 30 en la muestra se encuentran todos entre 1 a 2 desviaciones estándar con respecto de la media (entre 1.2 a 1.9)

La distribución es ligeramente negativa, ya que esta cargada hacia la izquierda. ya que 19 de los 30 delincuentes(63%) recibieron sentencias por debajo de la media, mientras que solo 11 (37%) recibieron sentencias por encima de la media.

[illegible]

Capítulo 2: Estadísticas disponibles y sus interpretaciones

Panorama general

Introducción

Objetivo

Al término del capítulo, el participante conocerá las principales estadísticas disponibles.

Tabla de
contenidos

A continuación se enlistan los temas incluidos en este capítulo

1. Estadísticas centro de llamadas.
2. Estadísticas incidentes.
3. Estadísticas despacho.
4. Estadísticas cruzadas con bases de datos externas.

Estadísticas centro de llamadas.

Generalidades

Estas estadísticas permitan conocer la actividad del centro de llamadas los distintos tipos de llamadas y las cargas relativas según las horas según el día de la semana.

Además permite analizar las eficiencias del servicio de toma de llamada y de su personal.

Mediante éstas se pueden detectar los cambios de comportamiento de las víctimas en cuanto a su facilidad de reportar los incidentes.

Este curso será completado cuando la nueva versión de ECHO esté disponible.

Estadísticas incidentes

Generalidades.

Estas estadísticas permitan de conocer las evoluciones de los incidentes según las horas según el día de la semana.

El análisis por zona permite crear la cartografía delictiva.

Asociado a otros datos externos, por el uso de las técnicas de minería de datos permite de analizar los riesgos y prever algunos comportamientos delictivos.

Este curso será completado cuando la nueva versión de ECHO esté disponible.

Notas del Participante y Reflexiones:

Estadísticas despacho

Generalidades

Estas estadísticas permiten conocer las eficiencias del servicio de despacho y de medir los progresos en las intervenciones.

Permite también detectar falta de medios y de repartir los medios función de la evolución de los delitos y de la necesidad de agregar nuevos tipos de medios.

Este curso será completado con la nueva versión de ECHO/LIMA.

Estadísticas cruzadas con bases de datos de datos externos.

Generalidades

Esas funciones serán desarrolladas en fase 2 del proyecto.

El uso de minería de datos o de análisis cruzadas de datos no están previstas en fase1.

Anexo I. Índice de imágenes, tablas y gráficas

Introducción En este anexo se enlistan las diferentes imágenes utilizadas en este documento.

En este anexo se indica la ubicación exacta de cada imagen dentro de este manual.

Imágenes

Imagen 1. Relación positiva y negativa	15
Imagen 2. Pantalla para selección	31
Imagen 3. Selección del tema que contiene al capítulo 1	32
Imagen 4. Selección del ejercicio 1	32
Imagen 5. Selección del libro que contiene a la tabla 1	33
Imagen 6. Víctimas por grupo de edad	33
Imagen 7. Algunos cálculos.....	34
Imagen 8. Datos para gráfico de pastel.....	53
Imagen 9. Selección de gráficos	53
Imagen 10. Comando para selección de gráfico circular	54
Imagen 11. Gráfico circular	54
Imagen 12. Comando para agregar etiquetas a un gráfico	55
Imagen 13. Como dar formato a las etiquetas de datos.....	55
Imagen 14. Formato de etiqueta de datos. Como mostrar nombre y porcentaje ..	56
Imagen 15. Paso 1 para no mostrar la leyenda en un gráfico	56
Imagen 16. Paso 2 para no mostrar la leyenda en un gráfico	57
Imagen 17. Como ordenar la información en Excel.....	59
Imagen 18. Fórmulas para obtener el porcentaje.....	59
Imagen 19. Selección de columnas no contiguas	60
Imagen 20. Comando para gráficos de columna.....	60
Imagen 21. Gráfico de columnas.....	61
Imagen 22. Como eliminar la leyenda en un gráfico	61
Imagen 23. Gráfico de columnas donde no se muestra la leyenda.....	62
Imagen 24. Selección de datos para ordenar en orden ascendente.	65
Imagen 25. Datos ordenados en orden ascendente.....	65
Imagen 26. Modelo con fórmulas en Excel para realizar la tabla de distribución de frecuencias, mostrada en la tabla 17.....	71
Imagen 27. Curva de distribución normal	72
Imagen 28. Distribuciones asimétricas.....	72
Imagen 29. Polígono de frecuencias	73
Imagen 30. Ojiva o curva de frecuencias acumulada.....	73
Imagen 31. Tasas de víctimas hombres por 100 mil habitantes hombres en México para el año 2004, según resultados de la ENSI-3. (Sólo considera a las personas de 18 años y más). Distribución de frecuencias en Excel	74
Imagen 32. Histograma, Polígono de frecuencias y Curva de frecuencias acumulada u Ojiva.....	74
Imagen 33. Modelo de distribución de frecuencias para una variable	78

Imagen 34. Modelo de distribución de frecuencias para una variable por entidad.	79
Imagen 35. Instrucciones para mover o copiar una hoja seleccionada.	80
Imagen 36. Selección del libro a donde mover o copiar hojas seleccionadas.	81
Imagen 37. Instrucciones para guardar con un nombre distinto.	81
Imagen 38. Como cambiar el nombre de archivo.	82
Imagen 39. Evolución de los delitos denunciados.	90
Imagen 40. Robo de vehículo con violencia, en datos brutos.	91
Imagen 41. Media, mediana y moda son idénticas.	94
Imagen 42. media < mediana < moda.	94
Imagen 43. Moda < mediana < media.	94
Imagen 44. Conjuntos de datos con igual media y diferente variabilidad.	98
Imagen 45. Conjuntos de datos con rango igual y diferente variabilidad.	99
Imagen 46. Área bajo una curva dividida en partes iguales por los cuartiles menor y mayor, y la mediana.	100
Imagen 47. Variabilidad de un conjunto de datos con respecto a su media.	102
Imagen 48. Gráfico de caja.	107
Imagen 49. Gráfico de caja con fronteras interiores y exteriores, menores y mayores; y con valores extremos.	108
Imagen 50. Modelo para análisis descriptivo de los datos.	110
Imagen 51. Modelo para captura de datos.	111
Imagen 52. Pantalla de captura de datos.	112
Imagen 53. Modelo de porcentaje acumulado y z-score.	113
Imagen 54. Modelo de distribución de frecuencias.	114

Tablas

Tabla 1. Datos hipotéticos de habitantes y policías para 10 localidades.	19
Tabla 2. Muestras de tamaño 2, obtenidas de la tabla 1.	20
Tabla 3. Datos hipotéticos del delito de robo de vehículo y robo a casa habitación para 5 localidades.	22
Tabla 4. Obtención de muestras a partir de una población.	22
Tabla 5. Víctimas en México para el año 2004, según resultados de la ENSI- 3 por sexo y grupo quinquenal de edad.	35
Tabla 6. Delito de robo a transeúnte registrado por la PGJ DF para las delegaciones de Iztapalapa y Tlalpan en los años 1997 ^a 2006.	40
Tabla 7. Edades de las víctimas del delito de robo a transeúnte en el centro histórico, en un día en particular.	44
Tabla 8. Distribución de frecuencias para las edades de las víctimas en el centro histórico mostradas en la tabla 2.	45
Tabla 9. Edades de las víctimas del delito de robo a transeúnte. Como selecciona valores máximos y mínimos.	46
Tabla 10. Límites de clase. Como calcularlos.	47
Tabla 11. Ejemplo de conteo de observaciones para saber la frecuencia de cada clase.	47
Tabla 12. Conteo para obtener la frecuencia de cada clase.	48
Tabla 13. Ejemplo de clases no mutuamente exclusivas.	49

Tabla 14. Obtención de frecuencia y frecuencia acumulada	49
Tabla 15. Cálculos para obtener la distribución de frecuencias.	51
Tabla 16. México: denuncias ante las agencias del ministerio público del fuero común por delitos patrimoniales, según modalidad, para el año 2006.	52
Tabla 17. México: denuncias ante las agencias del ministerio público del fuero común por delitos de robo, según modalidad, para el año 2006.	58
Tabla 18. Tasas de víctimas hombres por 100 mil habitantes hombres en México para el año 2004, según resultados de la ENSI-3, por entidad de ocurrencia del delito. (Sólo considera a las personas de 18 años y más)	64
Tabla 19. Valores a sumar para calcular el ancho de clase.	67
Tabla 20. Resumen de cálculos para el rango e intervalos	67
Tabla 21. Tabla para cálculos de límites inferior y superior.	68
Tabla 22. Distribución de frecuencias. Tasas de víctimas hombres por 100 mil habitantes hombres en México para el año 2004, según resultados de la ENSI-3.	69
Tabla 23. Tabla para obtención de la columna de frecuencia	69
Tabla 24. Personas de 18 años y más, que consideran que vivir en su municipio actual es inseguro y población de 18 años y más por entidad de residencia y sexo. 2004	75
Tabla 25. Delitos registrados por la PGJ DF en Iztapalapa en el año 2005	89
Tabla 26. Robo de vehículo por 100 mil habitantes y por 10 mil viviendas con vehículo particular.	92
Tabla 27. Orden que guardan las delegaciones en el delito de robo de vehículo con violencia, según datos brutos, por 100 mil habitantes o por 10 mil viviendas con vehículo particular.	93
Tabla 28. Homicidio doloso y robo con violencia a casa habitación según la PGJ DF en el año 2005, en datos brutos.	96
Tabla 29. Cálculos para obtener la varianza (ejemplo 2.11)	103
Tabla 30. Cálculos para obtener la desviación estándar. Ejemplo (a).....	104
Tabla 31. Cálculos para obtener la desviación estándar. Ejemplo (b).....	105
Tabla 32. Averiguaciones previas iniciadas por mes.....	106
Tabla 33. Robo a transeúnte. Ejemplo.	112

Gráficas

Gráfica 1. Denuncias ante las agencias del ministerio público del fuero común por delitos patrimoniales, según modalidad, para el año 2006.....	52
Gráfica 2. México: denuncias ante las agencias del ministerio público del fuero común por delitos de robo, según modalidad, para el año 2006	58
Gráfica 3. México: denuncias ante las agencias del ministerio público del fuero común por delitos de robo, según modalidad, para el año 2006. En barras.	63

Bibliografía

Autor (es)	<i>Documento</i>
TSA (Thales Software Asia)	Echo User Guide V0.8
TSA (Thales Software Asia)	Lima User Guide. V0.4
Cybertech	Manual de Usuario. V5.0
TSS (Thales Security Systems)	SSS (Especificación del Sistema Segmento) # 47139873-305-rev 1 Fase 1 Proyecto Bicentenario Ciudad Segura.
TSS (Thales Security Systems)	CONOPS. (Conceptos de Operación) V1.5
ARALIA Systems LTD	User Manual ILEX V2.2
ARALIA Systems LTD	User Manual ASTER V2.2
FEDETEC	Manual de Operador ICCS. V5.0 – Ed 1

Fórmulas usadas:

- (1) media aritmética $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$
n= número de datos disponibles.
- (2) $mediana = \frac{n+1}{2}$ dato en la posición
- (3) rango = dato con mayor valor - dato con menor valor
- (4) $Q_1 = .25(n + 1)$ valor del dato en la posición
- (5) $Q_3 = .75(n + 1)$ valor del dato en la posición
- (6) mediana = $Q_2 = .5(n + 1)$
- (6.a) $p(n + 1)$
- (7) Varianza = $s^2 = \frac{\sum_i (x_i - \bar{x})^2}{n - 1}$
- (8) desviación estándar $s = \sqrt{s^2} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$
- (9) varianza $s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n - 1}$
- (10) coeficiente de variación $cv = \frac{s_x}{\bar{x}} 100$
- (11) asimetría = $\frac{3(\bar{x} - mediana)}{s}$
- (12) $curtosis = \frac{\frac{1}{2}(Cuartil_3 - Cuartil_1)}{percentil_{90} - percentil_{10}}$
- (13) El rango intercuartílico ($Q_3 - Q_1$)
- (14) frontera interior menor = $Q_1 - 1.5(\text{rango intercuartílico})$
- (15) frontera interior mayor = $Q_3 + 1.5(\text{rango intercuartílico})$
- (16) frontera exterior menor = $Q_1 - 3(\text{rango intercuartílico})$
- (17) frontera exterior mayor = $Q_3 + 3(\text{rango intercuartílico})$
- (18) distancia = $x - \bar{x}$
- (19) $z_{score} = \frac{\text{distancia}}{\text{desviacion estandar}} = \frac{x - \bar{x}}{s}$
- (20) $asimetria1 = \frac{\sum (x_i - \bar{x})^3}{s^3}$

$$(21) \quad \frac{Abs(\text{suma desvios})}{n} = \frac{\sum |x_i - \bar{x}|}{n}$$

$$(22) \quad a) = \text{suma de desvios}^2 = \sum (x_i - \bar{x})^2;$$

$$(23) \quad b) = \text{media (a)} = \frac{\sum (x_i - \bar{x})^2}{n}; \text{ y,}$$

$$(24) \quad c) = \text{raíz (b)} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

$$(25) \quad \text{autocovarianza}_{\text{rezagado } k} = \frac{\sum (x_t - \bar{x})(x_{t-k} - \bar{x})}{n - k - 1}$$

$$(26) \quad \text{autocorrelacion}_{\text{rezagado } k} = \frac{\sum (x_t - \bar{x})(x_{t-k} - \bar{x})}{\sum (x_t - \bar{x})^2}$$

$$(27) \quad \text{error estandar} = \frac{s}{\sqrt{n}}$$

Referencias

- Mendenhall, Reinmuth, Beaver.** Statistics for Management and Economics. PWS-KENT, 6a. edición.
- Sincich terry.** Business Statistics by Example. Maxwell Mac Millan Internacional Editions. 3a. Edición.
- Groebner David F y Patrick W. Shanon** Business Statistics. A Decisión Making Approach.. Maxwell. Mac Millan. 3a. Edición
- Ott, Mendenhall.** Understanding Statistics.. pws-kent, 5a. edición.
- Ott, Larson, Rexroat, Mendenhall.** Statistics, a Tool for the Social Sciences. PWS-KENT, 5a. edición.
- Ya-lun Chou.** Análisis Estadístico. Interamericana. 2a. edición.
- Makridakis, Wheelwright, McGee.** Forecasting, Methods and applications. WILEY, 2a edición.

Glosario

Término	Descripción
Estadística.	La estadística esta asociada a los métodos que se aplican en la recolección, organización, presentación, análisis e interpretación de datos numéricos.
Estadística descriptiva.	Es una de las ramas de la estadística, que se ocupa de la descripción y análisis de un grupo de información específico, sin obtener conclusiones particulares sobre un grupo mayor, común mente llamado población.
Estadística inferencial.	Representa la segunda rama de la estadística, y se ocupa de predecir (inferir) u obtener conclusiones particulares sobre un grupo mayor que es la población, con base en las características de una muestra de la misma.
Población.	Representa el total de observaciones o características de interés al investigador.
Muestra.	Representa un subconjunto de las observaciones o características de la población de interés.
Parámetros poblacionales.	Son los principales aspectos que quieren conocer de una población, y están representados por las medidas de tendencia central y las medidas de dispersión.
Constante.	Se le denomina así, cuando la observación de un fenómeno no cambia en sucesivos ensayos.
Variable.	Es cuando las observaciones de un fenómeno cambian de un ensayo a otro.
Variable continúa.	Se deriva de una medición y puede tomar cualquier valor decimal de una recta.
Variable discreta.	Se deriva de un conteo y es la que puede tomar cualquier valor entero.
Variable cuantitativa.	Esta representada en cantidades numéricas, los cual significa que pueden ser discretas o continuas.
Variable cualitativa.	Es la que se expresa mediante características.
Nivel de medición de intervalo.	Tienen el cero arbitrariamente situado y éste no representa la ausencia de la característica que se esta midiendo.
Nivel de medición nominal.	Representa un nivel de medición donde únicamente la clasificación se realiza según la característica o características del objeto de estudio.

Porcentaje. Representan una relación que se establece de una de las partes con respecto al todo o total multiplicado por cien. En otras es una proporción multiplicada por 100.

Proporción. Es el número de elementos que tiene una característica específica dividida por el total de elementos.

Gráfica de barras.

Es una gráfica para describir datos cualitativos elaborada por medio de barras, donde cada barra representa una categoría de la variable.

Gráfica de pastel.

Representa la gráfica de mayor utilización dentro de la descripción de datos cualitativos. Es un círculo dividido en diferentes secciones y cada sección representa un porcentaje.

Frecuencia de clase.

Es el número de observaciones que se identifican para una clase determinada.

Distribución de frecuencia.

Es la forma de organizar los datos cuantitativos dentro de un intervalo o la agrupación de los datos dentro de un intervalo de clase. La frecuencia se muestra para cada intervalo de clase.

Diagrama de dispersión.

Los diagramas de dispersión sólo muestran en forma gráfica por medio de puntos la forma en que se distribuyen los datos recolectados.

Histograma. Un histograma consiste en la representación en forma de barras de la distribución de frecuencias.

Polígono de frecuencia.

El polígono de frecuencias es una forma gráfica alternativa para presentar resultados relativos a una distribución de frecuencias, en su elaboración se utilizan líneas.

Rango. Se obtiene de la diferencia entre el dato mayor y el dato menor del conjunto de observaciones de una muestra. Representa el campo de variación.

Ojiva. Es la representación gráfica de la frecuencia relativa acumulada.

Tabla de frecuencias.

Es otra forma de llamar a la distribución de frecuencias.

Cuartil Menor. Representa el percentil 25.

Cuartil mayor. Representa el percentil 75.

Media. Es la suma de las mediciones (observaciones o datos) dividido por el total de las misma.

Mediana. Se ubica a la mitad de un conjunto de datos e indica el percentil 50.

Moda. Es la observación que se presenta con mayor frecuencia.

Percentil. Representan las proporciones reales de las observaciones acumuladas hasta un valor determinado para una muestra.