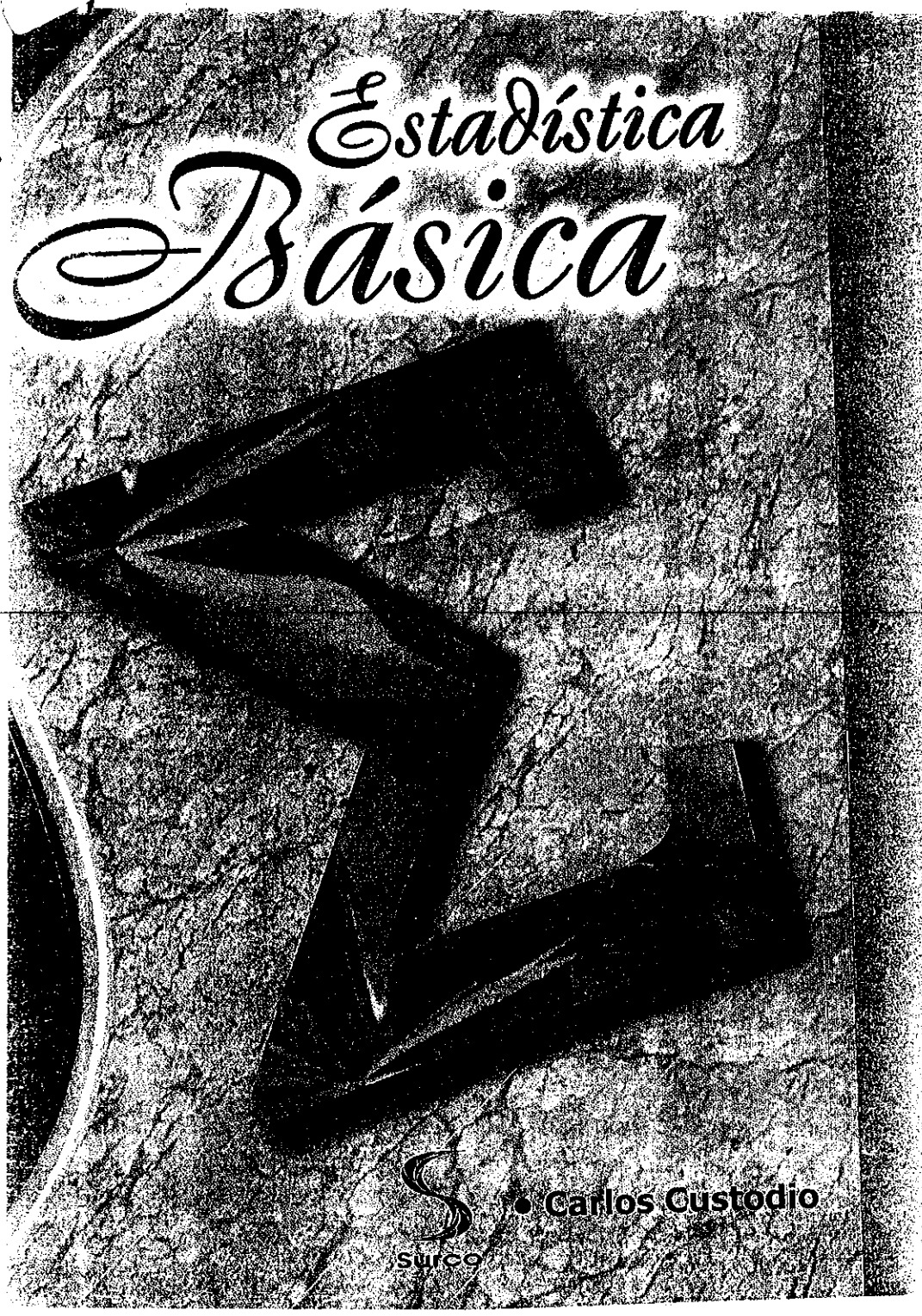
Estadística Básica



Carlos Custodio

Surco

Reservados todos los derechos. Se prohíbe, sin la autorización escrita de Editorial Surco o del autor, la reproducción total o parcial de esta obra por cualquier medio o procedimiento , incluidos la reproducción fotostática y el tratamiento informático.

© Carlos Custodio

Diagramación y composición: **Editora Master Print** 809-290-1368

Impresión: **Editora Búho**

Cuarta edición, 2007

ISBN: 99934-45-7-01

Editorial Surco

Publicaciones de Editorial Surco
César Nicolás Penson 56
Suite 402, Gazcue, Santo Domingo, R.D.
Tels.: 809-688-5854 / 687-9167
E-mail: editorialsurco@hotmail.com

Impreso en la República Dominicana

Índice

Prólogo 11
Introducción 15

I. Aspectos Básicos 19
1.0 Introducción 19
1.1 La estadística y su importancia 19
1.2 Orígenes y desarrollo de la estadística 23
1.3 Estadística descriptiva 24
1.4 Estadística inferencial 25
1.5 Términos y conceptos básicos 26
1.6 Censo 27
1.7 Encuesta 32
1.8 Etapas de una encuesta por muestreo 34
1.8.1 La planeación 34
1.8.2 La elaboración de un diseño de investigación 34
1.8.3 Diseño de la muestra -muestreo- 35
1.8.4 Diseño del cuestionario 35
1.8.5 El trabajo de campo 36
1.8.6 La verificación y la codificación 36
1.8.7 La preparación para el análisis 37
1.8.8 Análisis y preparación del informe 37

II. Muestreo 39
2.0 Introducción 39
2.1 Representatividad de las muestras 41
2.2 Métodos de selección de muestras 42
2.3 Muestreo probabilístico 43
2.3.1 Muestreo aleatorio simple 43

2.3.2 Muestreo sistemático 46
2.3.3 Muestreo estratificado 47
2.3.4 Muestreo de conglomerados 49
2.4 Muestreo no probabilístico 50
2.4.1 Muestreo de juicio 50
2.4.2 Muestreo por cuotas 51
2.4.3 Muestreo decisional 53
2.4.4 Muestreo de agrupación causal 53

III. Los datos estadísticos: su clasificación y organización 55
3.0 Introducción 55
3.1 Datos estadísticos 56
3.1.1 Clasificación de los datos estadísticos 56
3.2 Distribución de frecuencias 59
Ejercicios de Aplicación 67

IV. Presentación de datos estadísticos 73
4.0 Introducción 73
4.1 Presentación escrita 74
4.2 Presentación tabular 74
4.3 Presentación gráfica 78
Ejercicios de Aplicación 89

V. Medidas de tendencia central o de posición 113
5.0 Introducción 113
5.1 Notación sumatoria 114
5.2 La media aritmética 116
5.2.1 La media aritmética para datos no agrupados 118
5.2.2 La media aritmética para datos agrupados 119
5.2.3 La media aritmética ponderada 122
5.2.4 Características de la media aritmética 125
5.2.5 Propiedades matemáticas de la media aritmética 127
5.3 La mediana 134
5.3.1 Mediana para datos no agrupados 134
5.3.2 La mediana para datos agrupados 136
5.3.3 Características de la mediana 139
5.4 La moda 140
5.4.1 La moda para datos no agrupados 141
5.4.2 La moda para datos agrupados 142
5.4.3 Características de la moda 144

5.5 La media geométrica 145
5.5.1 La media geométrica para datos no agrupados 145
5.5.2 La media geométrica para datos agrupados 148
5.5.3 Características de la media geométrica 150
5.5.4 Tasa promedio de crecimiento 151
5.6 La media armónica 153
5.6.1 La media armónica para datos no agrupados 154
5.6.2 La media armónica para datos agrupados 155
5.6.3 Características de la media armónica 156
5.7 Medidas de posición no central 156
5.7.1 Deciles 156
5.7.2 Cuartiles 157
5.7.3 Percentiles 158
5.7.4 Deciles, cuartiles y percentiles para datos agrupados 160
Ejercicios de Aplicación 163

VI. Medidas de dispersión 181
6.0 Introducción 181
6.1 Medidas de dispersión absolutas 182
6.1.1 Recorrido o rango 182
6.1.2 Desviación intercuartílica o cuartílica 183
6.1.3 Desviación media 185
6.1.4 Varianza 189
6.1.5 Desviación típica o estándar 189
6.2 Medidas de dispersión relativas 195
Ejercicios de Aplicación 199

VII. Medidas de asimetría y de apuntamiento o curtosis 205
7.0 Introducción 205
7.1 Asimetría 205
7.1.1 Coeficiente de asimetría de Pearson 208
7.2 Apuntamiento o curtosis 214
7.2.1 Momentos 217
Ejercicios de Aplicación 223

VIII. Números índices 225
8.0 Introducción 225
8.1 Números índices simples 227
8.1.1 Indice de precios 228
8.1.2 Indice de cantidad 228
8.1.3 Indice de valor 229
8.1.4 Relativos de enlace 236
8.2 Números índices compuestos 236
8.2.1 Indice de agregados no ponderados 237
8.2.2 Indice de agregados ponderados 239
8.3 Indices más importantes en economía y administración .. 248
8.3.1 Indice de precios al consumidor 248
8.3.2 Indice de precios al por mayor 252
8.3.3 Promedio industrial Dow-Jones 252
8.3.4 Indice de producción industrial 253
Ejercicios de Aplicación 254

Bibliografía 259

Prólogo

El gran desarrollo experimentado mundialmente por la estadística se debe a la necesidad del uso de sus herramientas en todas las ciencias para describir y resumir un conjunto numérico de datos e interpretar sus resultados. Esto se pone de manifiesto, principalmente, en los campos de la sociología, demografía, econometría, administración, procesos productivos, publicidad, negocios, política, ecología, etc.

De lo que se trata no es de una exaltación del dato estadístico como condición sine qua non del quehacer científico. Resulta más bien de una filosofía y de una práctica científica y tecnológica que nos hace ver en muchos ámbitos del saber, la necesidad de organizar los datos de la realidad tipificándolos, y a base de ellos generar argumentaciones que permiten explicar la realidad e implementar acciones a fin de satisfacer alguna necesidad.

Nada de lo anterior es tan cierto como en la sociología, la economía y la administración, disciplinas científicas preocupadas no solo por entender los grandes procesos sociales, en su organización y cambio, sino también, por la planificación del desarrollo a mediano y largo plazos. Es así, como hacer predicciones o pronósticos sobre las tendencias de población de una zona, región o país, cuyos efectos, producidos por estas tendencias en la dinámica del cambio —como puede ocurrir en la demanda de empleos y servicios y en los mercados de consumidores— debe ser analizado y evaluado continuamente con técnicas estadísticas confiables, con el objeto de definir o redefinir cualesquiera estrategias.

En ese sentido, las estadísticas se hacen vitales, en tanto aportan datos actualizados inherentes a los procesos señalados, que han de ser analizados y evaluados. De esta manera, la estadística como conjunto de métodos contribuye no solo con la ciencia en la elaboración de teorías generales, sino que aporta instrumentos metodológicos y técnicos al servicio de los procesos productivos y para el mantenimiento y mejoramiento de la calidad de la vida.

Por vía de consecuencia, es válido pensar, además, en el interés de la estadística en el complejo mercado de valores, el cual se fundamenta en un conjunto de indicadores estadísticos que sirve de base a la comercialización de acciones industriales y comerciales. De esta manera, podemos ver que no existe una actividad científica, económica y social en el mundo donde la estadística no juegue un papel importante. Por ejemplo, en el ámbito de la pedagogía, los principios fundamentales que la sustentan, aconsejan que la enseñanza de cualquier disciplina debe integrar diferentes estrategias de aprendizaje: conceptuales, metodológicas, instrumentales y valorativas que les permitan al estudiantado desarrollar su capacidad para comprender la realidad y comprendiéndola y explicándola, organizar procesos que incidan significativamente en los ámbitos generales y particulares de la vida humana. Esto supone que el estudiantado habrá alcanzado el nivel de conocimientos y habilidades para construir y manejar conocimientos e informaciones que favorezcan el ejercicio innovador y eficiente de su profesión.

Naturalmente, esto requiere metodología de aprendizaje apropiadas que les permitan al estudiantado superar miedos frente a determinadas disciplinas, como por ejemplo, frente a las matemáticas, en razón de que muchos estudiantes se asumen incapaces para comprenderlas. Ya se sabe que este es un falso problema y que se debe a las dificultades pedagógicas del mundo académico-docente por no relacionar en el acto de aprendizaje las matemáticas con la vida cotidiana. La estadística constituye una sorpresa, principalmente para

aquellos que prefieren estudiar áreas que no estén supuestamente ligadas a las matemáticas. El autor de **Estadística Básica** conocedor de estos inconvenientes, por su vasta experiencia académica, trata de solucionarlos en este texto mediante la utilización de herramientas matemáticas en el análisis de los diferentes capítulos de la obra, desarrollados con un considerable nivel de detalles para así lograr mayor comprensión y mejor manejo instrumental científico del estudiantado.

La obra **Estadística Básica** está concebida justamente, en una perspectiva global, a través de la cual se quiere crear conciencia en el estudiantado de la importancia de la estadística en el mundo productivo y en el desarrollo moderno, a partir de la dimensión innovadora del acto de aprendizaje en la educación superior. Pero también persigue producir cambios en las concepciones sobre el marco operacional de las matemáticas en el aprendizaje.

En consecuencia, en este libro **Estadística Básica**, su autor trata de que se produzca un cambio de actitud frente al uso del dato cuantificable, y se convierta en un saber para la práctica social y por lo tanto, que sea útil para la vida. En este caso, el autor, invita y desafía a la creatividad, incluyendo ejercicios relacionados con la experiencia y la realidad de los estudiantes. También el autor estimula a los estudiantes a obtener experiencia de aprendizaje, mediante la utilización de datos e informaciones provenientes de diferentes contextos.

Ha de reconocerse la contribución de esta obra al acervo bibliográfico de aprendizaje en la educación superior en la República Dominicana. En la bibliografía dominicana no se conocía hasta este momento ningún libro de estadística cuyo contenido se ajustara en su totalidad a los programas universitarios de estadística básica. El texto por lo tanto, contribuirá, en todas sus partes, a resolver el gran problema de la escasez de un libro idóneo que han venido confrontando los estudiantes del nivel básico de estadística en nuestras universidades. En otras palabras, este aporte a la educación

CAPITULO

I

Aspectos Básicos

1.0 Introducción

"El pensar de manera estadística llegará a ser un día algo tan necesario para ser buen ciudadano como el poder leer y escribir", así lo comentó el novelista inglés Herbert George Wells hace más de cien años. Sin lugar a dudas, que la historia, un siglo después ha demostrado que Wells tenía una aguda percepción del futuro.

La estadística se usa como un valioso auxiliar en los diferentes campos del conocimiento y en variadas actividades de las ciencias. Hoy día, difícilmente podría encontrarse una rama del saber en el que las herramientas estadísticas no tengan una preponderante aplicación.

La estadística es un lenguaje que permite comunicar información basada en datos cuantitativos. Así, a diario nos encontramos con una gran variedad de datos numéricos relativos a fenómenos como: la tasa de desempleo, la actividad del mercado de valores, comercio exterior, índices industriales, resultados de encuestas de opinión, datos sobre actividades deportivas, índice de precio al consumidor, índice de precio al mayoreo, etc.

1.1 La Estadística y su importancia

El desarrollo de los métodos y conceptos estadísticos en los últimos decenios ha sido tan importante que constituye uno de los aspectos más relevantes entre los estudiosos de las ciencias. La estadística, por mucho tiempo, se vinculaba principalmente con la recolección de datos y su presentación en tablas y gráficas.

La evolución de la estadística ha llegado al punto en que su proyección se percibe en casi todas las áreas de trabajo. Esto se debe a que la estadística moderna está relacionada de manera directa con el problema de la toma de decisiones en condiciones de incertidumbre.

La estadística abarca la recolección, presentación y caracterización de información para ayudar tanto en el análisis e interpretación de datos como en el proceso de la toma de decisiones.

La característica más importante del gran avance de la estadística consiste en el cambio de los métodos puramente descriptivos por otros que sirven para establecer generalizaciones.

Esto es, un cambio de la estadística descriptiva a la estadística inferencial.

La estadística puede aplicarse en diferentes áreas funcionales, como son:

Administración

- Para describir las características de los empleados dentro de una empresa.
- Para mejorar la calidad de los servicios o de los productos fabricados por la empresa.

Mercadeo

- Para hacer estimaciones de la proporción de clientes que prefieren un producto.
- Para establecer estrategias de publicidad para la comercialización de un producto.

Contabilidad

- Para seleccionar muestras, cuando se pretende hacer una auditoría.
- Para medir variación en los costos de producción.

Economía

- Para crear indicadores de cambios en el precio, cantidad, valor, etc.

- Para desarrollar métodos que permitan pronosticar valores en el futuro.

Educación

- Para expresar coeficientes de aprovechamiento.
- Para los fines de investigación en educación.
- Para la realización de diagnósticos.

En sentido general, la estadística es parte esencial de la formación profesional. O sea, la lógica estadística, el pensar estadístico, y las operaciones estadísticas, son hasta cierto punto una parte necesaria en toda profesión.

"La estadística, comentaba T.M. Porter en 1986, ha llegado a conocerse en el siglo veinte como la herramienta matemática para analizar datos experimentales y basados en la observación. Conservada religiosamente por la política pública como la única base confiable para los juicios en torno a la eficacia de procedimientos médicos o a la seguridad de químicos, y adoptada por las empresas para usos como el control de calidad industrial, está evidentemente entre los productos de la ciencia cuya influencia sobre la vida pública y privada ha sido más penetrante. El análisis estadístico ha llegado a ser visto en muchas disciplinas científicas como indispensable para llegar a conclusiones confiables a partir de resultados empíricos. Desde la invención del cálculo, si acaso, no se ha encontrado un nuevo campo de las matemáticas con tan amplio dominio de aplicación", termina señalando T.M. Porter.

La estadística es un lenguaje que permite comunicar información basada en datos cuantitativos. Así, a diario nos encontramos con una gran variedad de datos numéricos relativos a fenómenos como: la tasa de desempleo, la actividad del mercado de valores, comercio exterior, índices industriales, resultados de encuestas de opinión, datos sobre actividades deportivas, índice de precio al consumidor, índices de precio al mayoreo, etc.

El uso creciente de la estadística en las ciencias administrativas es parte de una tendencia de los últimos años en

basar las decisiones administrativas en un fundamento tan objetivo y científico como resulte posible.

Así, la administración moderna depende cada vez más de datos estadísticos para conseguir la información necesaria acerca de sus operaciones internas y sobre los diferentes campos de negocios y actividades económicas.

Los datos estadísticos son concisos, específicos, capaces de ser analizados objetivamente por diferentes procedimientos y se prestan para hacer comparaciones. Por estas razones, son especialmente útiles en funciones administrativas de gran trascendencia, como establecer metas, evaluar rendimientos, localizar debilidades en las empresas, medir progresos, etc.

Las organizaciones recopilan datos sobre sus operaciones internas, a través del sistema de contabilidad así como a través del sistema de registros de datos.

Asimismo, la estadística es de gran importancia en la investigación científica debido a que:

1. Permite el tipo de descripción más exacto. La descripción de fenómenos es un objetivo de la ciencia, descripción que ha de ser tan precisa y completa que sea útil para quien pueda entenderla al leer los símbolos en términos de los cuales se describen esos fenómenos. La estadística y la matemática son parte de nuestro lenguaje descriptivo.
2. La estadística nos obliga a ser claros y exactos en nuestros procedimientos y en nuestro pensar.
3. Nos permite resumir los resultados de manera significativa y cómoda. Las grandes masas de datos son abrumadoras y sino se ponen en orden, prácticamente carecen de sentido. La estadística nos ofrece medios eficaces para ordenar el caos que produce una gran cantidad de datos desorganizados.
4. La estadística nos permite deducir conclusiones generales. Asimismo, siguiendo pasos estadísticos podemos saber hasta donde se puede tener fe en una conclusión y hasta donde podemos ampliar nuestras generalizaciones.

La estadística comprende la recolección, el resumen, la presentación, el análisis de datos y su interpretación.

Estas actividades y sus diferentes técnicas son propicias durante el proceso de toma de decisiones, lo cual es una función de gran importancia en la empresa.

La eficiencia en el proceso de toma de decisiones va a depender en gran medida de la correcta interpretación de los resultados estadísticos. Los resultados estadísticos pueden ser útiles sólo en la medida en que sean interpretados correctamente.

Con interpretaciones plenas y apropiadas extraídas de los datos, los resultados estadísticos son una fuente de gran sentido y significación.

1.2 Orígenes y desarrollo de la estadística

La estadística tiene sus orígenes en tiempos muy remotos, al igual que otras disciplinas. Sus antecedentes se limitan a meros recuentos de población, de tierras y de bienes.

Durante las civilizaciones egipcia, griega y romana, los datos se obtenían principalmente con propósitos de aplicar impuestos y planificar el reclutamiento militar. O sea, que los orígenes de la estadística están estrechamente ligados a la administración del estado.

En la Edad Media, las instituciones eclesiásticas mantenían registros de nacimientos, muertes y matrimonios. En el siglo XIX Quetelet, en Bélgica, considerado como el padre de la estadística aplicada al estudio social, la emplea en diversas disciplinas sociales, tales como la antropología, la demografía y la criminología.

En América, durante la época colonial, se mantuvieron diversos registros y a partir de 1790, la Constitución de Estados Unidos estableció el levantamiento de censos cada diez años.

Durante el período mercantilista, cuando los Estados pugnaron por crear potencias económicas dirigidas a propósitos políticos, se inició la recopilación de datos sobre aspectos

económicos, como comercio exterior, fabricación y suministro de alimentos. Muchos de los datos recopilados en este período eran tratados como secretos de Estado.

Hasta ese momento, la evolución de la estadística se había circunscripto a las actividades de conteo, medición, descripción, tabulación, ordenamiento y levantamiento censal, las cuales constituyeron los soportes en que se erigió la estadística descriptiva.

Otra corriente de influencia en la evolución de la estadística se originó en las matemáticas de juego de azar y condujo al desarrollo de la estadística inferencial. A principios de siglo, dos grandes exponentes del desarrollo de la estadística inferencial fueron Francis Galton, biólogo, y su discípulo Karl Pearson, biólogo y además matemático.

Uno de los acontecimientos científicos más importantes del presente siglo ha sido el gran desarrollo de la estadística y su incidencia como herramienta básica para la investigación empírica en otras disciplinas.

1.3 Estadística Descriptiva

La parte de la estadística que consiste en **recolectar** los datos **organizarlos, presentarlos, analizarlos** e **interpretar** los resultados recibe el nombre de estadística descriptiva.

Mark L. Berenson y David M. Levine, en su obra **"Estadística Básica en Administración"**, (1998), definen la estadística descriptiva como "aquellos métodos que incluyen la recolección, presentación y caracterización de un conjunto de datos con el fin de describir apropiadamente las diversas características de ese conjunto de datos".

Asimismo, John E. Freund, Frank J. Williams y Benjamín M. Perles, en su obra Estadística para la Administración con enfoque moderno; (1996), señalan que "por estadística descriptiva nos referimos a cualquier tratamiento de datos que esté diseñado para resumir o describir algunas de sus características más importantes sin intentar deducir nada que

escape al alcance de los datos". Cuando obtenemos informaciones a través de un censo, cuando calculamos el crecimiento intercensal de la población de un país, cuando se determina el crecimiento de la producción de una empresa para un período de tiempo determinado; todas estas informaciones pertenecen al campo de la estadística descriptiva.

El objeto de la estadística descriptiva es hacer que los datos se comprendan más fácilmente, que sea más sencillo referirse a ellos y analizarlos. La estadística descriptiva proporciona las herramientas para organizar, simplificar y resumir información básica a partir de un conjunto de datos.

Los diferentes métodos de la estadística descriptiva han sido la base del desarrollo de los métodos de la estadística inferencial y han servido para mejorar la teoría de las probabilidades, que en realidad, ha hecho posible la aplicación de la estadística a todos los campos de la investigación.

El estudio de las diferentes técnicas que nos permiten **recolectar, organizar, presentar, analizar** e **interpretar** resultados de un conjunto de datos son el objeto de este libro.

1.4 Estadística Inferencial

La investigación acerca de cómo obtener mayores beneficios a través de los juegos de azar durante el Renacimiento sirvió de base a la formulación de la teoría de las probabilidades.

Matemáticos como Bernoulli, De Moivre y Karl Gauss fueron los precursores del objeto de la estadística inferencial.

A principios del presente siglo, fue, sin embargo que estadísticos como Pearson, Fisher, Gosset, Neyman, Tukey y Wald sentaron las bases para el desarrollo de los métodos de la estadística inferencial.

La estadística inferencial la podemos definir como aquellos métodos que nos permiten hacer posible la estimación de una característica de la población, sobre la base del estudio de una muestra. Esto es, nos permite hacer generalizaciones en la población a través de los resultados obtenidos mediante el

estudio de una muestra. Por tanto, un aspecto sumamente importante de la estadística inferencial es el proceso de usar las estadísticas de muestras para sacar conclusiones respecto a los parámetros de la población.

La estadística inferencial tiene por objetivo describir las características de un conjunto, pero sin la necesidad de realizar el registro de datos a todos los elementos o unidades del conjunto o población, sino a una parte de ésta. Cuando la selección de los elementos o unidades de la población se hace de acuerdo a los procedimientos que las técnicas del muestreo establecen, entonces se puede inferir o inducir a partir de la parte o muestra estudiada, cuál o cuáles son los indicadores estadísticos de toda la población con mayor o menor aproximación.

La inferencia constituye la base teórica del muestreo, permite conocer el todo con cierta aproximación, a partir del estudio de una parte.

Es preciso señalar que la descripción de un conjunto sobre la base de los principios de la estadística inferencial no nos dan una certeza completa en sus resultados, sino que están sujetas a una probabilidad de error. Una incorrecta selección de las unidades muestrales puede acarrear la imposibilidad de inferir correctamente las características de la población.

Por ejemplo, cuando se desea hacer una estimación sobre la votación que obtendrá un candidato en una elección, el investigador no va a entrevistar a la población completa. Sino que, seleccionará una muestra y, basándose en los resultados obtenidos, extraerá conclusiones referentes a toda la población de votantes.

Asimismo, este investigador anexará a estas conclusiones los indicadores probabilísticos que indican el nivel de confianza que los resultados de esta muestra reflejan con relación al comportamiento de la votación en la población total.

Para una comprensión más efectiva, a continuación vamos a definir y tratar algunos términos y conceptos básicos.

1.5 Términos y conceptos básicos

Población:

- Una población es cualquier grupo completo, ya sea de personas, animales o cosas. Es la totalidad de elementos o cosas bajo consideración. La población se refiere a un grupo finito de elementos.

Elementos de una población:

- Son las unidades individuales que constituyen o conforman una población.

Universo:

- Se refiere al conjunto de sucesos o cosas que no tienen límite numérico.

Muestra:

- Es una porción de la población que se selecciona para fines de análisis. La muestra siempre debe ser representativa de la población de la cual se extrae, o sea, que cada uno de los elementos de la población tenga la misma oportunidad de ser seleccionado en la muestra.

Parámetro:

- Es una medida de resumen que se calcula con el propósito de describir alguna característica de la población.

Estadísticas o estadígrafos:

- Son medidas de resumen que se calculan con el propósito de describir algunas características de una sola muestra de la población. Por ejemplo, la media aritmética de las calificaciones de una muestra de estudiantes universitarios.

1.6 Censo

Es una investigación que cubre a todos los miembros o elementos de una población dada, ya sea una nación, una comunidad o una determinada organización. "El gran atractivo del censo, señalan Charles A. Lininger y Donald P. Warwick, en su obra "La Encuesta por Muestreo: Teoría y Práctica", (1981),

reside en que elimina las interrogantes sobre la representatividad de la información. Sin embargo, señalan a seguidas, dados su costo y otros recursos necesarios, no representa una alternativa realista para la mayor parte de los investigadores sociales y económicos. Además, se está haciendo cada vez más evidente que un censo completo es a menudo innecesario, antieconómico y una molestia para el público y también que es menos efectivo que una encuesta, para recoger ciertos tipos de información".

Periódicamente se levantan diferentes tipos de censos en todo el mundo, entre los más conocidos están: el censo de población y vivienda, censos agropecuarios, censos a las empresas mercantiles y manufactureras, etc.

Estos censos proporcionan datos muy importantes sobre población, vivienda, empleos, población económicamente activa, uso de la tierra, tamaño de las fincas, ganadería, etc.

Un censo de población, según el ***"Manual de Censos de Población y Habitación"*** de la Organización de las Naciones Unidas, 1992: ***"Es el conjunto de las operaciones consistentes en recopilar, procesar, evaluar, analizar y publicar datos demográficos, económicos y sociales relativos a todos los habitantes de un país o una parte bien delimitada de un país en un momento determinado".***

Los datos relativos al tamaño de la población y sus características demográficas, económicas y sociales constituyen una gran contribución para la planificación y ejecución de casi todas las actividades técnico–administrativas de los gobiernos nacionales y municipales, así como para el sector industrial y comercial de un país.

Las características principales de los censos de población y viviendas son la universalidad, el empadronamiento individual, la simultaneidad y la periodicidad definida.

Universalidad:

El censo debe abarcar todo el territorio o parte bien delimitada de un país. O sea, debe cubrir una zona definida con precisión. La utilidad de los datos censales es mayor si se

refieren a una zona bien definida, que por lo general es todo el territorio de un país.

En razón de esta universalidad, se debe incluir en la cobertura del censo a todas las personas que estén presentes o residan dentro del país.

Empadronamiento individual:

El censo supone el empadronamiento por separado de cada persona de cada hogar, así como el registro por separado de sus características.

Esta condición es fundamental para poder extraer del censo los datos demográficos, económicos y sociales correspondientes.

Simultaneidad:

Todas las personas y todas las viviendas deben ser empadronadas con relación a un mismo momento bien determinado y los datos recogidos deberán corresponder a un período de referencia bien definido. No es indispensable que el período de referencia sea idéntico para todos los tipos de datos obtenidos; en la mayoría de los casos, será el día del censo, pero en otros, como las características de la ocupación o el trabajo, puede establecerse en un período anterior. La simultaneidad en un momento o en un período específico, es esencial para asegurar un recuento exacto y obtener datos o características comparables.

Periodicidad definida:

Un censo presenta una imagen de la población en un momento determinado. Es útil disponer de una imagen única que suministre datos básicos, así estos datos serán más provechosos si se tienen a intervalos regulares.

Una serie de censos permite conocer las tendencias, evaluar el pasado, describir el presente y prever el futuro. Por consiguiente, conviene levantar censos a intervalos regulares, a fin de disponer de información comparable en una secuencia fija. Los censos consecutivos son fundamentales para estudiar los cambios y tendencias que aparecen con el tiempo en facetas

importantes de la población, tales como el crecimiento, la distribución, las características laborales, la estructura de edades, la alfabetización y el nivel de instrucción. Una serie de censos permite evaluar la magnitud y las orientaciones de las tendencias demográficas y socioeconómicas.

El objetivo primordial del censo es reunir y facilitar información detalladas sobre características específicas de la población y la vivienda en un momento determinado.

El censo facilita datos sobre importantes características demográficas, sociales y económicas de la población y la vivienda en todos los niveles de la planificación. Estos datos combinados con los datos macroeconómicos de otras fuentes, suscitan las bases para la formulación de políticas de desarrollo económico y social en los planos nacional y regional al paso que los datos censales de nivel microeconómico representan una contribución muy importante en la formulación de programas para el desarrollo, dentro del marco más general de los planes nacionales y regionales a los niveles de zonas pequeñas o comunidades.

Los datos censales son importantes porque sirven de base para la planificación de los diferentes sectores que integran el Estado.

En la República Dominicana, la organización del primer censo de población se remonta al año 1920, durante la ocupación norteamericana, registrándose una población de 894,665 habitantes. En 1935 se realizó el segundo censo de población, cuyos datos arrojaron la existencia de 1,479,417 habitantes. En 1950 se efectuó el tercer censo arrojando una población de 2,135,872 habitantes; el cuarto censo se organizó en 1960 y arrojó unas 3,047,070 personas. En 1970 se organizó el quinto censo de población dando como resultado unos 4,009,458 habitantes; el sexto censo de población se efectuó en 1981 de cuya enumeración resultaron unos 5,545,744 habitantes. El séptimo censo de población se efectuó en 1993, y dio como resultado una población de 7,293,390 habitantes en todo el territorio nacional

En sentido general, en el país se han organizado en total ocho censos de población, siendo el último de éstos el efectuado los días 18, 19 y 20 de octubre del año 2002, registrándose una población de 8,562,541 habitantes.

A continuación se presenta la distribución de la población por provincia y el Distrito Nacional según zona de residencia.

Población de la República Dominicana por Provincia y Zona de Residencia, 2002

Provincia	Zona				Total	%
	Urbana	%	Rural	%		
Distrito Nacional	913,540	10.7	-	-	913,540	10.7
Santo Domingo	1,234,721	14.4	583,033	6.8	1,817,754	21.2
Azua	113,988	1.3	94,869	1.1	208,857	2.4
Baoruco	49,530	0.6	41,950	0.5	91,480	1.1
Barahona	134,714	1.6	44,525	0.5	179,239	2.1
Dajabón	29,218	0.3	32,828	0.4	62,046	0.7
Duarte	165,477	1.9	118,328	1.4	283,805	3.3
El Seibo	33,738	0.4	55,523	0.6	89,261	1.0
Elías Piña	23,203	0.2	40,676	0.5	63,879	0.7
Espaillat	84,970	1.0	140,121	1.6	225,091	2.6
Hato Mayor	58,769	0.7	28,862	0.3	87,631	1.0
Independencia	31,729	0.4	19,104	0.2	50,833	0.6
La Altagracia	119,733	1.4	62,287	0.7	182,020	2.1
La Romana	197,050	2.3	22,762	0.3	219,812	2.6
La Vega	171,947	2.0	213,154	2.5	385,101	4.5
María T. Sánchez	61,238	0.7	74,489	0.9	135,727	1.6
Monseñor Nouel	101,725	1.2	65,893	0.8	167,618	2.0
Monte Cristi	52,909	0.6	58,105	0.7	111,014	1.3
Monte Plata	76,987	0.9	103,389	1.2	180,376	2.1
Pedernales	14,450	0.2	6,757	0.1	21,207	0.2
Peravia	90,901	1.1	78,964	0.9	169,865	2.0
Puerto Plata	157,282	1.8	155,424	1.9	312,706	3.7
Salcedo	23,126	0.3	73,230	0.9	96,356	1.1
Samaná	29,046	0.3	62,829	0.7	91,875	1.1
San Cristóbal	273,018	3.2	259,862	3.0	532,880	6.2
San José de Ocoa	32,920	0.4	29,448	0.3	62,368	0.7
San Juan	114,133	1.3	126,972	1.5	241,105	2.8
San Pedro de Macorís	244,571	2.9	57,173	0.7	301,744	3.5
Sánchez Ramírez	69,218	0.8	81,961	1.0	151,179	1.8
Santiago	602,721	7.0	305,529	3.6	908,250	10.6
Santiago Rodríguez	25,979	0.3	33,650	0.4	59,629	0.7
Valverde	114,153	1.3	44,140	0.5	158,293	1.8
Total	5,446,704	63.6	3,115,837	36.4	8,562,541	100.0

La organización y financiamiento del censo de población que periódicamente se ejecuta en el país recae sobre la Oficina Nacional de Estadística, institución del Estado Dominicano, rectora del sistema estadístico nacional.

1.7 Encuesta

Es una investigación en la cual la información se obtiene de una fracción de la población llamada muestra.

La encuesta es una técnica altamente valiosa para la realización de estudios de importantes temas, tales como opinión pública, comportamiento económico, intención de voto, actitudes y opiniones, etc.

Es un medio apropiado y útil para recoger información bajo tres condiciones:

1.- Cuando los objetivos de la investigación exigen datos cuantitativos.

2.- Cuando la información buscada es razonablemente específica y familiar para los entrevistados, y

3.- Cuando el propio investigador tiene un considerable conocimiento previo de los problemas particulares y de la gran variedad de las posibles respuestas que surjan.

Estas condiciones se cumplen fielmente en las áreas de investigación que hemos señalado en el párrafo anterior; opinión pública, comportamiento económico, intención de voto, actitudes y opiniones.

"Las encuestas de opinión han transformado en algunos aspectos decisivos la manera en que los gobiernos y los partidos plantean y desarrollan una campaña electoral. Este instrumento de tecnología política ha sustituido en buena parte a la antigua confianza de los profesionales del poder en su olfato o en las experiencias comparadas.

En efecto, la información obtenida mediante procedimientos estadísticos permite extraer conclusiones bastante más seguras que los dictados por la intuición o el sentido común. Es un hecho sabido que la información es poder y resulta lógico, en consecuencia, que los

estrategas de las compañas electorales se esfuercen por disponer de la mayor cantidad de datos ocultos sobre las actitudes de los futuros votantes, a fin de utilizarlos al servicio de sus propósitos. Las encuestas de opinión, así, pueden mostrar claramente los cambios de actitud de los electores ante buen número de cuestiones y permitir a los partidos reaccionar ante esos nuevos datos mediante la modificación o matización de sus programas y discursos". Editorial Periódico *"El País"*. España, 14/9/82

La encuesta por muestreo disfruta de cinco ventajas que la hacen atractiva frente al censo, aún en los casos donde a éste se le considera prácticamente exclusivo, como son los estudios sobre población.

La primera consiste en que su realización resulta **menos costosa** que el censo. Una encuesta, científicamente adecuada, cubre sólo una pequeña parte de los elementos que se incluyen en el censo y, por lo tanto, requiere de menos personal y de un reducido aparato administrativo.

La segunda ventaja consiste en que la encuesta permite **mayor rapidez** en la recopilación y análisis de los datos. Un censo a nivel nacional requiere de más de un año para diseñar los trabajos previos a los inicios de las acciones de campo, por ejemplo, confección del material cartográfico, diseño de cuestionario, reclutamiento del personal en todo el país, etc.

Una vez terminado el trabajo de campo, por lo regular se requieren meses de intenso trabajo (en algunos países años) para el procesamiento y publicación de los datos.

La tercera ventaja es que el número de **entrevistadores** utilizado en la encuesta **es mucho menor** y por consiguiente es posible ofrecerles **una capacitación más eficaz** y **una supervisión más cuidadosa** para asegurar la obtención de datos de alta calidad.

La cuarta ventaja es el resultado de una mejor calificación y la mayor capacitación del personal de campo de una encuesta, con lo cual se alcanza mayor y especial flexibilidad en los temas cubiertos por la encuesta. Los entrevistadores bien capacitados

pueden dedicar más tiempo a los entrevistados y explorar algunos tópicos con lujo de detalles.

La quinta ventaja consiste en que la encuesta por muestreo es menos notoria que el censo y no constituye una pesadumbre para el público.

Un censo nacional de población requiere de una campaña publicitaria con el objetivo de crear el interés y la colaboración de la ciudadanía. La encuesta, al no requerir prácticamente de publicidad y hacer contactos con una fracción de la población no crea oposición ni saturación hacia la investigación.

Una gran ventaja del censo es que proporciona datos hasta un nivel de subdivisiones geográficas muy pequeñas que no se logra en una encuesta.

1.8 Etapas de una encuesta por muestreo

La encuesta por muestreo o cualquier investigación socio-económica implica varias etapas en su proceso de ejecución. Las etapas en el proceso de investigación son interdependientes, o sea, cada una afecta a las otras.

A continuación mostramos las etapas fundamentales de un proceso de investigación;

1.8.1 La planeación

Esta etapa incluye el establecimiento de los objetivos de la investigación y la creación de la estrategia general para obtener y analizar los datos, incluye además, la elaboración de presupuestos, revisión de literatura, definición de objetivos específicos, etc.

La planeación es la etapa esencial para el éxito de una investigación. Es preferible no llevar a cabo una encuesta, si ésta no ha sido bien planeada.

1.8.2 La elaboración de un diseño de investigación

El diseño de la muestra está relacionado estrechamente con la planeación y por lo regular ocurre en forma simultánea. Consiste en una programación preparada de antemano con el objetivo de recopilar y analizar la información necesaria para

satisfacer los objetivos del estudio, al más bajo costo posible.

El diseño de la investigación debe estar acorde con los objetivos generales del estudio y con los tipos de informaciones necesarias para poder llevarla a cabo. Para ello, se debe tomar en cuenta los diferentes métodos que se utilizarán para recopilar esta información, ya sean entrevistas personales, telefónicas o cuestionarios enviados por correo.

1.8.3 Diseño de la muestra -muestreo-

Es el proceso de seleccionar ciertos elementos de la población. En esta etapa, el investigador debe definir con claridad y cuidado la población que se va a estudiar y la generalización de los datos de la muestra que dicha población permitirá.

El diseño de la muestra consiste en un conjunto de procedimientos para seleccionar los elementos de la población y para convertir las informaciones de la muestra en cálculos relativos al total de la población.

La muestra debe ser representativa de la población de la cual se extrae. Los procedimientos aleatorios para elegir las unidades son la forma más confiable de selección.

1.8.4 Diseño del cuestionario

Es el proceso de adaptar los diferentes objetivos del estudio a preguntas que permitan obtener la información necesaria. Esta etapa incluye la formulación del tipo de pregunta, número de éstas, secuencia y los medios para mantener motivada a la persona encuestada.

En esta etapa, se prueba el cuestionario para evaluar las preguntas, así como la eficiencia de las instrucciones del muestreo, la calidad de las entrevistas, la eficacia de la organización en el campo de trabajo, las posibilidades de contradicciones que pueden surgir en la investigación, duración de la entrevista, etc.

Existen dos objetivos fundamentales en el diseño del cuestionario, que son:

1.- Obtener información aplicable a los propósitos de la encuesta o investigación.

Para asegurar la aplicabilidad de la información, el investigador debe tener una clara idea sobre el tipo exacto de información que requiere el estudio.

Se debe tener una base lógica explícita para cada pregunta del cuestionario, que explique no sólo porqué se formulará la pregunta sino también qué se va a hacer con la información.

2.- Recopilar las informaciones con el máximo de responsabilidad y validez, o sea, con precisión.

Cuando los términos utilizados y la secuencia de las preguntas se diseñan con el objeto de motivar al entrevistado y ayudarle a que recuerde más fácilmente, se intensifica la precisión.

Un requisito obligatorio en el diseño del cuestionario de una investigación es el respeto a la dignidad y a la privacidad del entrevistado.

1.8.5 El trabajo de campo

En esta etapa de la investigación se incluye el reclutamiento de entrevistadores, supervisores y otro personal de campo. Además, incluye la capacitación del personal de campo, preparación de instrucciones escritas para los entrevistadores sobre el uso del cuestionario, la recopilación de los datos en el campo y el control de calidad de las entrevistas, etc.

1.8.6 La verificación y la codificación

La verificación y la codificación son procesos afines diseñados con el propósito de transcribir la información registrada en los cuestionarios a una forma adecuada para el análisis estadístico.

La verificación es un paso preliminar en el que las respuestas son inspeccionadas, corregidas y en algunas ocasiones son precodificadas de acuerdo a reglas establecidas.

El procedimiento de la verificación de las respuestas empieza cuando el entrevistador revisa cada cuestionario en el campo y continúa con las revisiones que realizan el supervisor de campo y el personal de oficina. El objetivo básico de la verificación es eliminar las respuestas incompletas o inconsistentes, como también los errores en el uso del cuestionario.

La codificación es un proceso técnico que consiste en convertir datos cualitativos en datos numéricos que puedan ser almacenados, contados o tabulados con rapidez y facilidad.

1.8.7 La preparación para el análisis

Esta etapa incluye un conjunto de actividades, como: digitación de los datos, chequeos rutinarios para determinar la compatibilidad de las respuestas, plan de tabulación de los datos.

La digitación es el procedimiento utilizado para el almacenamiento de la información y constituye el primer paso después de la codificación para registrar los datos, de manera tal que puedan ser fácilmente recuperados y tabulados.

La actividad de chequear de manera rutinaria los datos que se van digitando se hace con el propósito de descubrir inconsistencias en algunas respuestas, igualmente los posibles errores surgidos durante la digitación.

El plan de tabulación de los datos consiste en diseñar el formato de las diferentes tablas que nos interesan, en función de los objetivos de la investigación. El plan de tabulación implica, además, el cruce de dos o más variables en una tabla.

1.8.8 Análisis y preparación del informe

Consiste en la presentación e interpretación de tablas simples y de múltiples entradas de los datos recopilados en la investigación. Igualmente, involucra un tratamiento estadístico más complejo y una interpretación teórica elaborada de los resultados.

El objetivo de esta etapa es proporcionar un resumen de los datos, capaz de satisfacer los propósitos de la investigación, lo más breve y comprensible posible.

El análisis puede incluir: tablas de porcentajes, medidas de tendencia central, medidas de asociación (coeficiente de correlación, regresión), pruebas de hipótesis, estimaciones, etc.

La presentación de los resultados se incluye en esta etapa y consiste en la preparación de un informe que resume todo el estudio.

Cuando se prepara el informe debemos tener presente a quién va dirigido. Si es preparado principalmente para especialistas, éste deberá contener mucho más información sobre los aspectos metodológicos de la investigación, que si es preparado para el público en general. Pero, independientemente del público al que va dirigido el informe, es importante tener presente tres aspectos básicos en su planeación. Estos son: el estilo en que se va a escribir, la mecánica de presentación del material y la organización de los temas del informe.

CAPITULO

II

Muestreo

2.0 Introducción

Una muestra es una parte de un conjunto mayor llamado población, que se selecciona cuidadosamente para representarla. Es la base fundamental para el diseño de una buena encuesta. Una muestra representativa de la población debe captar la diversidad del total. **Muestreo** es el proceso que se lleva a cabo para seleccionar dicha parte.

En una encuesta o investigación, la muestra se selecciona con la finalidad de obtener información de la cual hacer deducciones sobre el grupo total en estudio. Ello abarca tres procesos relacionados entre sí y que son los siguientes: **muestreo**, el proceso de seleccionar una parte del todo; **medición**, la etapa intermedia que consiste en contar y formular preguntas; y **estimación**, el proceso de hacer deducciones sobre el grupo total partiendo de la información de la muestra.

El muestreo y la estimación constituyen lo que en investigación se conoce como el **diseño de la muestra**.

Nuestros conocimientos, en sentido general, actitudes y acciones del diario vivir están basados en gran medida en muestras. Esto es una realidad tanto en la vida cotidiana como en la investigación científica.

De esta manera, la opinión que tiene una persona acerca de una determinada empresa que durante décadas viene dando un servicio al público, por lo regular se basa en dos o tres contactos realizados durante un determinado tiempo.

Es común observar a consultores que al cabo de diez o quince días en un país extranjero, escriben informes para pautar a los gobernantes e instituciones cómo planificar y administrar el sistema de transporte, cómo equilibrar el presupuesto, cómo establecer un sistema de seguridad social, etc.

En realidad, la diferencia existente entre este consultor y los estudiosos de estas áreas que viven diez o quince años dedicados a estudiarlo es que, el primero basa sus conclusiones en una muestra mucho más pequeña.

Tanto en la ciencia como en los asuntos puramente humanos carecemos de los recursos necesarios para estudiar más de una porción de los fenómenos que pueden enriquecer nuestros conocimientos.

El objeto del muestreo es establecer generalizaciones con respecto a una población total de elementos sin tener que examinarlos uno por uno.

El muestreo es ampliamente usado en los negocios y en el gobierno. En la gerencia de producción, el muestreo del proceso de manufactura se utiliza a menudo para mantener la adecuada calidad de un producto. El material que se recibe como insumo regularmente se muestrea para comprobar su calidad antes de que entre en el proceso de producción.

El muestreo se utiliza en estudios de tiempo y movimientos, en asuntos relacionados con la determinación de la fracción de tiempo, durante la cual una máquina o trabajador no se encuentra en producción.

En la investigación de mercados, se usa el muestreo para determinar las preferencias del consumidor, para probar nuevos productos, para evaluar una campaña publicitaria, para determinar los niveles de consumo de un producto en un mercado real o potencial, para estudiar actitudes de los consumidores, para determinar los mejores métodos de distribución, etc.

El muestreo se usa en muchas áreas de contabilidad y finanzas, por ejemplo, en la auditoría interna de las operaciones

contables de una empresa. Los registros, tales como registros de cuentas, se muestrean para obtener la información necesaria.

Los equipos físicos se muestrean para determinar su condición y su costo de reposición, los cálculos de nómina se muestrean para comprobar la exactitud de los mismos.

En la administración pública, el muestreo se ha usado extensamente en varios aspectos, tales como, empleos, desempleos, datos sobre ingresos, ahorros, compras; los ingresos por impuestos se muestrean para obtener estimaciones relativas a la población de todos los ingresos por impuestos, etc.

2.1 Representatividad de las muestras

El objetivo del estudio de una muestra es generalizar sus resultados, obtenidos a través de los diferentes métodos de análisis estadísticos, en toda la población. Así, el método para seleccionar la muestra es de gran importancia al tomar en cuenta la validez de las inferencias realizadas, a partir de la muestra de la población objetivo.

La muestra debe seleccionarse de una forma deliberada a partir de la población origen, de modo que sea representativa de la población. Es decir, las características de la muestra se aproximan a las de la población con un margen de error conocido. Algunos investigadores con poca destreza en métodos estadísticos, con frecuencia les preocupa más el tamaño de la muestra que su representatividad, lo que constituye un grave error.

Una muestra representativa de 200 personas es preferible a una muestra no representativa de 2,000,000 de personas.

El mejor y más conocido ejemplo de la no representatividad de una muestra o error de selección muestral, según Mark L. Berenson y David M. Levine, en su obra Estadística Básica en Administración, Concepto y Aplicaciones, (1998), ocurrió en el sondeo de 1936 del Literary Digest. En ese año, **"Literary Digest, una revista respetada, predijo que el gobernador Alf Landon de Kansas recibiría el 57% de los votos y que ganaría**

abrumadoramente las elecciones presidenciales". Cuando se contaron los votos reales, Landon recibió sólo 38%, mientras que el presidente Franklin Delano Roosevelt fácilmente se reeligió para un segundo período en funciones. El tamaño del error del estudio de Literary Digest se consideró enorme y sin precedentes, habiendo perdido su credibilidad.

Qué salió mal?

La predicción de Literary Digest se basó en las respuestas de 2.4 millones de individuos, un enorme tamaño de muestra. Una razón importante fue la no representatividad o error en la selección.

En 1936, el país, E.U.A., todavía resentía la gran depresión. Sin embargo, Literary Digest compiló su marco de población de muestras tales como los directorios telefónicos, las listas de membresías de clubes, suscripciones a revistas y registros de automóviles, circunscribiéndose así a los ricos y excluyendo de su lista a la mayoría de la población votante, quienes durante este período de dureza económica, no podían permitirse comodidades tales como teléfono, membresías a clubes, suscripciones a revistas y automóviles. Por tanto, 57% de las estimaciones del voto para Landon pudieron haber estado muy cercanas de la población objeto de estudio, pero no de la población real.

La muestra seleccionada por Literary Digest fue de 12,000,000 de personas la cual fue respondida (por correo) por 2.4 millones de norteamericanos. Aunque la tasa de respuestas fue pírrica, aproximadamente un 20%; los 2.4 millones que respondieron constituye una de las muestras más grande.

Este es un claro ejemplo de que el tamaño de una muestra nunca puede compensar la falta de representatividad de ésta.

2.2 Métodos de selección de muestras

Existen dos métodos de selección de una muestra, según la manera de seleccionar sus elementos, que son: muestreo probabilístico y muestreo no probabilístico.

2.3 Muestreo probabilístico

Es un proceso de selección de la muestra en el cual los elementos son elegidos por métodos aleatorios, o sea, la selección de los elementos para la muestra se efectúa por procedimientos al azar y con probabilidades conocidas de selección.

El muestreo probabilístico es el preferido por los investigadores porque la selección de la muestra es objetiva y el error muestral puede ser medido en términos de probabilidad bajo la curva normal.

Los tipos más comunes de muestreo probabilístico son: muestreo aleatorio simple, muestreo sistemático, muestreo estratificado y muestreo de conglomerados.

2.3.1 Muestreo aleatorio simple

Es un método de selección de muestras en el cual los elementos o unidades se eligen individual y directamente por medio de un proceso aleatorio, en el que cada elemento no seleccionado tiene la misma oportunidad de ser elegido al igual que todos los otros elementos en cada extracción de la muestra. De modo que cada elemento en la población debe tener igual probabilidad de ser seleccionado.

El método más simple para obtener una muestra, mediante el muestreo aleatorio simple consiste en numerar o codificar los elementos de la población. Luego de numerarlos se echan en una bolsa y se extraen uno por uno y al azar cada elemento de la muestra. Este método es de fácil aplicación cuando la población bajo estudio es pequeña.

Este método, por conveniencia de manejo es reemplazado por una tabla de números aleatorios; véase en la página 41.

El procedimiento para determinar una muestra, mediante muestreo aleatorio simple, a través del uso de la tabla de números aleatorios se resume así:

1.- Asignar un solo número del 1 al N (N = tamaño de la población) a las unidades o elementos de la población.

Para una población con N<100, sólo se requieren dos dígitos, por ejemplo, 01, 02, 03, 04, etc. Para poblaciones con N<1000 se requieren solamente números de tres dígitos, por ejemplo, 001, 002, 003, etc.

2.- Decidir la dirección en que se va a leer la tabla (hacia abajo, hacia arriba, hacia la derecha, hacia la izquierda, etc.) entonces a la ciega dejar caer el lápiz sobre la tabla de números aleatorios y comenzar a leer.

3.- Desde el punto de inicio, comience a escoger los números aleatorios requeridos, siempre de acuerdo al tamaño de la población; como se indica en el primer paso.

4.- Los elementos de la población cuyos números asignados coincidan con los números obtenidos aleatoriamente constituyen la muestra aleatoria. Los números aleatorios duplicados y los que exceden a N se ignoran.

Por ejemplo, si queremos seleccionar una muestra de 10 personas (n=10) de una población de 200 personas (N=200), primero asignamos a cada persona un número, desde 001 hasta 200. Inmediatamente, decidimos la dirección en que se va a leer la tabla y preferimos hacia abajo.

Seguido, seleccionamos al azar el punto de partida de la tabla, dejando caer el lápiz en la tabla.

Supongamos que el punto de partida corresponde a la hilera 9 y la tercera columna y la dirección del movimiento es del punto de partida hacia abajo, como lo habíamos establecido.

Como N=200, solamente los tres primeros dígitos de cada cinco dígitos son necesarios. De esta manera, los 10 elementos de la muestra serán: 175, 089, 048, 003, 047, 017, 059, 167, 170, 162.

Tabla de números aleatorios

Las tablas de números aleatorios están compuestas por los dígitos 0, 1, 2, 3, 4, 5, 6, 7, 8, 9; o sea, contienen los 10 dígitos. Este conjunto de dígitos puede ser leído de manera individual o en grupos y en cualquier dirección; hacia abajo, hacia arriba,

hacia la derecha, hacia la izquierda y diagonalmente, y siempre se pueden considerar como aleatorios.

Estas tablas, son de mucha utilidad para el muestreo aleatorio o al azar ya que poseen dos características muy particulares.

TABLA DE NUMEROS ALEATORIOS

Hilera	*(1)*	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	24571	23165	39407	60614	99692	53643	15237	75497
2	13670	32919	85543	04891	95940	36404	76575	21672
3	30051	56205	28399	57818	50250	64143	21454	43778
4	17977	66365	89867	29215	16767	78664	61052	20792
5	92178	52766	05531	89370	29936	73564	89039	87520
6	55136	12504	50905	63482	77089	16116	69540	42617
7	56292	65313	87697	77362	25261	41434	41434	78662
8	28852	23758	99995	89994	80072	16037	09242	02476
9	36575	00384	56044	71864	37692	93583	67871	55693
10	73548	41988	41754	77623	74789	47006	71348	08856
11	72077	69908	26013	89159	29262	21100	13848	26884
12	62212	59442	56691	84042	17000	40994	90372	92380
13	15835	87145	62164	00392	48946	85269	43931	84040
14	11418	34412	57620	27362	40064	36081	14038	56486
15	75712	00893	75595	99815	16218	04983	25848	55323
16	82781	43481	65187	25236	97977	79008	07923	01439
17	11341	31929	46669	91080	53736	98034	03093	82350
18	67303	67361	40344	20562	65616	94776	74398	73140
19	15563	55883	75112	51585	85050	45388	82118	85799
20	24516	67385	30307	70874	29955	71904	07820	24392
21	29072	24881	51692	76856	70769	91318	33021	56804
22	98398	84599	80434	23325	57478	09349	70354	51570
23	76460	07180	17586	89669	51861	94388	38384	55181
24	74656	38853	70503	48664	67571	54981	12207	02247
25	44671	32847	91009	54934	52683	77609	50212	30923
26	27906	28645	45676	35369	65227	48617	74535	12413
27	40996	11863	78692	95991	14621	30490	84343	91097
28	21917	59181	36117	04700	17593	03176	08479	02393
29	71416	97913	42120	91634	87433	42675	21008	72725
30	06475	16974	50487	28922	50273	34026	12902	41020
31	46714	35033	60626	22039	28633	10540	89239	59295
32	44155	67697	73377	80482	52186	90008	20379	21490
33	64558	40620	56281	64703	21641	93937	48274	05923
34	48437	75697	68647	23526	89468	08261	92414	84681
35	48573	61866	08920	23339	55006	41144	53299	19156
36	15915	75596	87992	28897	85916	53472	02117	10983
37	65319	21980	74852	01770	82811	57641	40814	03221
38	16521	15881	35674	05940	79340	40810	89367	85602
39	83865	40476	53331	63211	58895	32626	34054	51839
40	47653	32290	48778	87661	56278	33845	92017	20928

La primera característica es que los dígitos están ordenados de manera tal que la probabilidad de que aparezca cualquiera en un punto dado de una secuencia es igual a la probabilidad de que ocurra cualquier otro.

Esto significa que en una secuencia de números, la probabilidad de que aparezca un dígito cualquiera en un punto de la secuencia es igual a 1/10.

La segunda característica es que las combinaciones de dígitos tienen la misma probabilidad de ocurrir que las otras combinaciones de un número igual de dígitos.

Esta segunda característica significa que todas las combinaciones de dos dígitos, tres y así sucesivamente son igualmente probables.

Así, se puede observar que estas dos características satisfacen los requisitos necesarios para el muestreo aleatorio.

2.3.2 Muestreo sistemático

En un muestreo sistemático, cada elemento K-ésimo, de una población arreglada en cierto orden especificado, se selecciona para la muestra partiendo de un primer elemento determinado al azar entre los elementos K.

Una muestra sistemática se obtiene cuando los elementos se seleccionan en forma ordenada.

Este método de selección de la muestra depende del número de elementos o unidades incluidos en la población y el tamaño de la muestra. Requiere del uso de un listado de todos los elementos de la población.

Si los elementos del listado no están dispuestos en un orden particular, el muestreo sistemático puede dar lugar a un muestreo aleatorio simple, muestreando cada elemento k-ésimo del listado.

El elemento K se obtiene dividiendo el tamaño de la población entre el tamaño de la muestra, o sea,

$$K=\frac{N}{n}$$

Así, si N=300 y n=20, entonces

$$K=\frac{300}{20}=15$$

Esto significa que cada decimoquinto elemento de la población va a ser seleccionado.

El primer elemento de la muestra se selecciona de manera aleatoria, consultando la tabla de números aleatorios para determinar dónde empezar y luego cada K-ésimo elemento será seleccionado.

Por ejemplo, si la tabla de números aleatorios nos indica el número 12, entonces se deberá seleccionar el decimosegundo elemento de la población como primer elemento, después, el elemento número 27 (o sea 12+15=27); el elemento 42 (o sea 27+15=42); elemento numero 57, etc.

Es así como una muestra sistemática puede dar la misma precisión de estimación de una población que una muestra aleatoria simple, cuando los elementos en la población están ordenados al azar.

El muestreo sistemático puede ser modificado ligeramente, evitando el problema de que el intervalo de muestreo coincida con un arreglo periódico en la población considerada y facilitando asimismo la evaluación del error de muestreo.

2.3.3 Muestreo estratificado

Existe una técnica que con gran amplitud es usada para obtener un muestreo más eficiente, o sea, para obtener más precisión al mismo costo que la que obtendríamos con una muestra aleatoria simple, es el muestreo estratificado.

En este tipo de muestreo, la población se divide en cierto número de subgrupos o estratos, cada uno de los cuales se muestrea independientemente.

El proceso, a través del cual se divide la población en subgrupos o estratos, recibe el nombre de estratificación. El objeto de la estratificación es llevar a cabo selecciones separadas en cada uno de los subgrupos o estratos.

Existen **dos razones** de gran importancia para el uso del muestreo estratificado que son: **controlar la representatividad** de la muestra y permitir la aplicación de procedimientos de **selección diferentes en cada uno de los estratos.**

Mediante el ordenamiento de los elementos de la población, que generalmente es heterogénea, en subgrupos o estratos homogéneos, la variabilidad es menor que la de la población total. Por tal razón, se necesitará un tamaño de muestra más pequeño. Así, en la medida en que son más semejantes los elementos en cada uno de los estratos, tanto menor será el tamaño de la muestra.

El proceso de selección dentro de cada estrato o subgrupo, por lo general es al azar, pero a veces es útil la realización de un censo para algunos de los estratos. Por ejemplo, en un inventario, podemos determinar que el 10% de los artículos existentes en almacén representan aproximadamente el 70% del valor del inventario total, y que el otro 90% de los artículos representan el restante 30% del valor del inventario. Como se trata de que son pocos los artículos considerados en la clase de alto valor, resulta más lógico realizar el levantamiento de un censo completo de estos artículos y tomar una muestra al azar de los otros estratos que presentan gran cantidad de artículos de menor valor.

Las estimaciones de la población, basadas en el muestreo estratificado, generalmente presentan más precisión que si la población fuera muestreada por medio del muestreo aleatorio simple.

El número de elementos que componen cada estrato es conveniente que sea proporcional al tamaño del estrato con relación a la población.

De esta manera, por ejemplo, si el tamaño del estrato A, representa el 25% de la población, ese mismo porcentaje es utilizado para seleccionar el número de elementos de la muestra del estrato A. Así, si el tamaño de la muestra es de 400 elementos, el 25% de la muestra van a ser seleccionados del

estrato A, o sea, unos 100 elementos. En cambio si el estrato B representa el 15% de la población, dicho porcentaje se utiliza para seleccionar el número de elementos de muestra de estrato B. De esta manera si el tamaño de la muestra es de 300 elementos, el 15% de la muestra se selecciona del estrato B.

2.3.4 Muestreo de conglomerados

El muestreo de conglomerados es un procedimiento de selección en el cual los elementos para la muestra se escogen de una población agrupada o aglomerada, en lugar de hacerlo de una población aislada. El procedimiento para obtener una muestra de conglomerados consiste en, primero, dividir la población en grupos que son convenientes para el muestreo. A seguidas, se selecciona una porción de los grupos al azar o por el método sistemático. Finalmente, se toman todos los elementos al azar o por el método sistemático de los grupos seleccionados para así obtener la muestra.

Por lo regular, una muestra obtenida, a través de este proceso, produce un mayor error muestral, o sea, proporciona una menor precisión de las estimaciones acerca de la población, que una muestra aleatoria simple del mismo tamaño.

Los elementos individuales dentro de cada conglomerado, normalmente tienden a ser iguales.

Para mayor comprensión del muestreo de conglomerados, vamos a ilustrar el proceso de selección a través de un estudio hipotético. Por ejemplo, nos interesa conocer las condiciones económicas de los estudiantes del nivel medio de una determinada ciudad.

La población en cuestión está constituida por todos los estudiantes del nivel medio inscritos en las escuelas dentro de los límites de la ciudad.

El muestreo aleatorio simple podría ser enfocado para el estudio. Los investigadores podrían usar los registros existentes en las escuelas públicas y privadas y por medio de las técnicas indicadas anteriormente, obtener una muestra aleatoria simple de la lista de la población en cuestión.

Sin embargo, el resultado sería una muestra distribuida por toda la ciudad. Aunque esta muestra podría presentar pocos problemas, resultaría poco conveniente y muy costosa por la alta dispersión de las escuelas por toda la ciudad.

Una alternativa sería tomar una muestra de todas las escuelas y después medir las condiciones económicas de todos los alumnos inscritos en esas escuelas. Este procedimiento de muestreo tendría la ventaja de reducir los costos y simplificar el trabajo de campo.

En el muestreo de conglomerados, los elementos se seleccionan por grupos y no en forma independiente. El efecto de la selección de grupos sobre el error estándar va a depender de la similitud entre los elementos del grupo y aquellos de la población.

En muchos casos, los elementos de la muestra seleccionados en grupos no mostrarán la misma variación que presentaría un número equivalente de elementos seleccionados de manera independiente. La gente que vive en la misma cuadra o en el mismo sector puede ser más parecida en cuanto a características, como por ejemplo, sus ingresos, que la gente de una población más grande.

2.4 Muestreo no probabilístico

El muestreo no probabilístico incluye todos los métodos en que los elementos de la muestra no se seleccionan mediante procedimientos al azar o aleatorios, o con probabilidades de selección conocidas. Algunos procedimientos de selección del muestreo no probabilístico son: muestreo de juicio, muestreo por cuotas, muestreo decisional y muestreo de agrupación causal.

2.4.1 Muestreo de juicio

Este es un proceso a través del cual los elementos se escogen basándose en opiniones informadas que garantizan la representatividad de la población que se estudia; este muestreo se basa en la opinión de expertos. Por ejemplo, un especialista

en planificación de transporte puede decidir que en un estudio sobre calidad del servicio se deben considerar sólo los autobuses que viajen a una región determinada, ya que éstos representan en forma adecuada al país en sentido general.

Otro experto puede determinar que, basándose en las elecciones presidenciales anteriores, la representación electoral de una sola provincia o del Distrito Nacional constituye la base suficiente para hacer predicciones sobre los resultados electorales de todo el país.

Aunque estas selecciones de muestras son prudentes y cautas, proporcionan una base de generalización muy peligrosa. Los juicios que se basan en fundamentos anteriores, por ejemplo, pueden ser invalidados por los cambios que se producen en las condiciones ocurridas entre el pasado y el presente.

Las muestras de juicios proporcionan interesantes estudios de casos y pueden generar hipótesis muy importantes, pero no constituyen bases confiables para estimaciones estadísticas.

2.4.2 Muestreo por cuotas

Es un proceso de selección en el cual los elementos son elegidos en el campo mismo, por los entrevistadores, utilizando categorías prefijadas de elementos de la muestra, para obtener un número predeterminado de casos en cada categoría. Las cuotas se establecen sobre la base de características conocidas de la población en estudio. Por ejemplo, un censo reciente puede mostrar que la población estudiantil de una provincia representa el 25% de la población estudiantil total.

El tipo más sencillo de muestreo por cuotas establecería que el 25% de la muestra provendría de esa provincia y las selecciones reales serían efectuadas por los entrevistadores en el campo.

Las cuotas pueden establecerse de manera satisfactoria, basándose en los datos de un censo y en otras informaciones, pero existen limitaciones prácticas que son impuestas por la clase de información disponible.

En la práctica, el proceso de selección por cuotas se basa, por lo general, en características tales como, edad, sexo, nivel educativo, tipo de vecindario, situación de empleo, etc.

En un estudio, por ejemplo, a un entrevistador se le puede asignar que entreviste 20 estudiantes del sexo masculino, eligiendo la mitad de un sector donde se encuentren las mejores viviendas y la otra mitad de los sectores más pobres; incluso se puede llegar a especificar las edades, por ejemplo que la mitad tenga menos de 30 años y la otra mitad tenga más de 30 años.

La dificultad que se presenta inmediatamente consiste en que el entrevistador debe decidir cuáles casas va a elegir en el mejor sector y cuáles en el más pobre.

Las instrucciones recibidas por el entrevistador pueden sugerir sobre cómo obtener mayor representatividad, pero, a menos que se utilicen métodos probabilísticos, existe mucha posibilidad de caer en arbitrariedades. En fin, la parte más débil del muestreo por cuotas es la selección de los elementos para cumplir las cuotas. Sencillamente, desconocemos si los elementos elegidos representan en realidad las similitudes y diferencias existentes entre los elementos de una categoría determinada.

Los partidarios de este tipo de muestreo pueden sostener que este método produce generalizaciones confiables sobre la población representada, aunque carezca de una base técnica teórica sólida. Además, asegurarían que debido a que regularmente no pueden completarse todas las entrevistas para los casos de muestras seleccionadas por métodos aleatorios, los casos resultantes no pueden calificar como una muestra probabilística.

Sin embargo, la ventaja del muestreo probabilístico es tal, que aún con una tasa de casos incompletos de 10-20%, es posible llegar a estimaciones muy razonables de error de muestreo. Con las muestras por cuotas, aunque se complete la cuota total de entrevistas, no existe base técnica alguna para calcular la escala de errores que puede esperarse de las estimaciones producidas.

2.4.3 Muestreo decisional

En este caso, los elementos de la muestra son seleccionados de una población por los entrevistadores que usan su propio criterio para decidir cuáles son los informantes representativos.

Por ejemplo, se puede instruir a entrevistadores para que entrevisten a prósperos hombres de negocios. Los diferentes entrevistadores pueden tener criterios diferentes de lo que constituye un próspero hombre de negocios.

2.4.4 Muestreo de agrupación causal

Son muestras formadas por elementos que se han reunido ocasionalmente o de acceso fácil, tales como los estudiantes inscritos en una clase que van pasando por una esquina. Dichas muestras no permiten generalizaciones que vayan más allá de las agrupaciones mismas y por lo general no tienen interés científico.

CAPITULO

III

Los datos estadísticos
Su clasificación y organización

3.0 Introducción

Los datos estadísticos han sido usados durante siglos por los gobiernos organizados como forma de ayudar a la toma de decisiones en la administración del estado.

Durante las civilizaciones egipcia, griega y romana, los datos se recopilaban principalmente con el propósito de reclutamiento militar y el cobro de impuestos.

Los datos estadísticos son concisos, específicos y capaces de ser analizados objetivamente por diferentes procedimientos.

Cuando se produce el nacimiento de un niño y éste registra un peso de tres kilos, su peso como un valor aislado tiene poco sentido. Pero cuando se dispone del peso de todos los niños nacidos durante un año,,estos valores pueden ser comparados, analizados e interpretados. Así, el peso de los niños nacidos durante un año constituye un conjunto de datos estadísticos.

En función de sus características los datos se clasifican en cuantitativos y cualitativos; siendo los cuantitativos la base fundamental de estudio de la estadística.

Cuando se dispone de una gran cantidad de datos desorganizados, éstos, de esa forma constituyen un caos estadístico. Sin embargo, mediante su ordenamiento a través de una distribución de frecuencias pueden ser resumidos y su análisis se torna más sencillo.

El uso de la computadora ha hecho posible que los gobiernos, las empresas y otras organizaciones almacenen y procesen grandes cantidades de datos.

3.1 Datos estadísticos

Los datos estadísticos se obtienen mediante un proceso que incluye la observación o medición de conceptos, como calificaciones de exámenes, ingresos anuales de una ciudad, temperatura diaria durante un año de una comunidad, velocidad de circulación de los vehículos por una autopista, etc.

Estos conceptos reciben el nombre de variables, ya que producen una serie de valores que tienden a mostrar cierto grado de variabilidad, al realizarse un conjunto de mediciones de manera sucesiva.

Las ventas anuales de una red de tiendas, el porcentaje de empleados que pagan el impuesto sobre la renta, la cantidad de toneladas de azúcar producidas anualmente, comprenden el análisis e interpretación de números, a los cuales se les conoce como datos.

Estos datos deben ser procesados, ya que si no se procesan pueden carecer de significado. El procesamiento transforma los datos en información, organizándolos y condensándolos en gráficas, en tablas o en valores que indican su esencia.

Existen varios tipos de datos estadísticos, que se agrupan en dos clasificaciones: datos de características cuantitativas y datos de características cualitativas; de los cuales hablaremos a continuación.

3.1.1. Clasificación de los datos estadísticos

Los datos estadísticos se clasifican según sus características en: datos de características cuantitativas y datos de características cualitativas.

Los datos de características cuantitativas son aquellos que se pueden expresar numéricamente y se obtienen a través de mediciones y conteos.

Un dato cuantitativo se puede encontrar en cualquier disciplina; psicología, contabilidad, economía, medicina, educación, etc.

Como ejemplo ilustrativo puedo citar: el coeficiente de inteligencia de una persona, el precio al consumidor de un litro de leche, la calificación en matemáticas de un estudiante del octavo grado.

Cuando estos datos son presentados en conjunto, no en forma aislada, de forma tal que podemos compararlos, analizarlos e interpretarlos, estamos ante un conjunto de datos estadísticos.

Por ejemplo, cuando tenemos las calificaciones de matemáticas de 100 estudiantes del octavo grado, estos datos son susceptibles de ser analizados, comparados e interpretados, no ocurriendo lo mismo con el dato de un solo estudiante.

Los datos de características cuantitativas se clasifican a su vez en: variables continuas y variables discretas.

Variables continuas

Cuando los datos estadísticos se generan a través de un proceso de medición se dice que estos son datos continuos.

Las variables continuas son aquellas que aceptan valores en cualquier punto fraccionario de un determinado intervalo, o sea, que aceptan fraccionamiento en un determinado intervalo.

Como ejemplos de variables continuas se puede señalar; la temperatura de la ciudad de Santiago durante un día de verano, la velocidad de un vehículo que se desplaza por la carretera Internacional, la edad de un niño del sexto curso de una escuela, etc.

Las variables citadas en los ejemplos anteriores (temperatura, velocidad y edad) se obtienen mediante un proceso de medición, y también aceptan fraccionamiento dentro de un determinado intervalo lo que confirma que son variables continuas.

Variables discretas

Contrario a las variables continuas, las variables discretas se generan a través de un proceso de conteo.

Las variables discretas son aquellas que no aceptan valores en puntos fraccionarios dentro de un determinado intervalo, o sea, son aquellas que no aceptan fraccionamiento dentro de un determinado intervalo.

El número de estudiantes de una universidad es un ejemplo de una variable discreta; asimismo, el número de hijos en una familia, el número de facturas por cobrar de una empresa comercial, el número de sentencias falladas por la Suprema Corte de Justicia; constituyen otros ejemplos de variables discretas.

La estadística, como método científico de investigación se nutre de informaciones numéricas, o sea, de datos de características cuantitativas.

Datos de características cualitativas

Los datos de características cualitativas son aquellos que no se pueden expresar numéricamente.

Estos datos se deben convertir a valores numéricos antes de que se trabaje con ellos.

Los datos de características cualitativas se clasifican en:

1.- Datos nominales

2.- Datos jerarquizados

Datos nominales

Los datos nominales comprenden categorías, como el sexo (masculino o femenino), carrera de estudio (administración, contabilidad, derecho, ingeniería, psicología, economía, etc.), material de los pisos (tierra, cemento, granito, mármol, madera, etc.), calificaciones (excelente, muy bueno, bueno, regular, malo, muy malo) etc.

Las características mencionadas no son numéricas por su naturaleza, pero, cuando se aplican, ya sea en una población o una muestra, es posible asignar a cada elemento una categoría y contar el número que corresponde a cada elemento. De esta manera estas características se convierten en numéricas.

Así, los datos nominales se obtienen cuando [illegible] diferentes categorías de un elemento y se cuenta el número de observaciones que corresponde a cada elemento.

Datos jerarquizados

Es un tipo de datos de características cualitativas que se refiere a las evaluaciones subjetivas cuando los conceptos se jerarquizan según la preferencia o logro.

Por ejemplo, las posiciones de una competencia de atletismo se jerarquizan en primer lugar, segundo lugar, tercer lugar, etc.;

Los datos jerarquizados constan de valores relativos asignados para denotar orden: primero, segundo, tercero, cuando, etc.

Tanto los datos nominales como los jerarquizados, que por su naturaleza no son numéricas, se convierten en datos discretos.

3.2. Distribución de frecuencias

Cuando se dispone de una gran masa o cantidad de datos a veces resulta muy difícil responder a ciertos cuestionamientos que sobre una determinada variable se nos hagan. Por ejemplo, si usted tiene las edades de 10 personas y se le pregunta que cuántas tienen 20 años o menos, resultaría muy sencillo contestar; pero si por el contrario, usted dispone de las edades de los estudiantes de cuarto semestre de una carrera de las Universidades de un país y se le formula la misma pregunta, ¿qué usted respondería?, es muy difícil contestar ya que sabemos es una cantidad muy elevada de estudiantes.

Existe una forma en estadística de organizar las informaciones que nos permite responder a este y otros cuestionamientos. A esta forma de organizar las informaciones se le llama **distribución de frecuencias**, y consiste en el ordenamiento de los datos a través de clases y frecuencias.

Cuando los datos se presentan en una distribución de frecuencias se les denomina datos agrupados. Cuando todos los datos observados de una variable se enumeran en forma desorganizada le vamos a denominar datos no agrupados.

Frecuencia simple de clase

Al construir una distribución de frecuencias, se tienen diferentes intervalos de valores que denominaremos clases.

Se define frecuencia simple de clase al número de veces que se repite cada clase. Se le identifica como fi, donde (f) se lee como frecuencia, e (i) define el orden de las clases.

Frecuencia relativa simple

A la suma total de la frecuencia simple de clase le llamamos n.

Cuando c/ la valor de la frecuencia simple de clase se divide entre el tc al de casos u observaciones a este cociente le denominamos frecuencia relativa simple.

La suma de la frecuencia relativa simple siempre será igual a la unidad. Vamos a identificar la frecuencia relativa simple como fr.

Por definición,

$$fr = \frac{fi}{n} = \frac{\text{frecuencia simple de clase}}{\text{número total de observaciones}}$$

Frecuencia acumulada

La suma sucesiva de la frecuencia simple de clase es denominada como frecuencia acumulada.

Al calcular la frecuencia acumulada en una distribución de frecuencias, la acumulada de la primera clase será igual a la frecuencia simple de la misma clase. La segunda acumulada es igual a la primera acumulada más la frecuencia simple de la segunda clase y así sucesivamente.

Así, $F_1 = f_1$

$F_2 = F_1 + f_2$

$F_3 = F_2 + f_3$

$F_4 = F_3 + f_4$

: : :

$Fn = Fn\text{-}1 + fn$

El valor de la última frecuencia acumulada es igual al total de datos.

La frecuencia acumulada se identifica como Fi.

Frecuencia relativa acumulada

Es el cociente que se obtiene al dividir cada frecuencia acumulada entre el total de observaciones. O, la suma sucesiva de la frecuencia relativa simple.

Esta se identifica como Fr.

El procedimiento de acumulación es similar al caso anterior.

$Fr_1 = fr_1$

$Fr_2 = Fr_1 + fr_2$

$Fr_3 = Fr_2 + fr_3$

: : :

$Fr_n = Fr_{n-1} + fr_n$

Recorrido o rango

En una distribución u ordenamiento de datos existe una diversidad de valores que varían de menor a mayor y viceversa.

Se denomina recorrido o rango a la diferencia existente entre el valor máximo observado y el mínimo en una distribución u ordenamiento.

R = Recorrido = Xmáx - Xmín

Intervalo de clase

Una clase está definida por un límite inferior o frontera inferior y un límite superior o frontera superior.

A la diferencia entre el límite superior y el límite inferior de una clase se le llama intervalo de clase. Este indica el recorrido o rango de los valores incluidos en una clase.

Intervalo de clase = Ci

Ci = Límite superior - Límite inferior

Punto medio de clase o marca de clase

Para fines de análisis de datos, los valores de las clases se representan a través del punto medio de clase o marca de clase.

El punto medio de clase se define como la semi-suma de los límites de clase.

El punto medio de clase se identifica como Xi, donde

Xi = 1/2 (Límite inferior + Límite superior)

Pasos para construir una distribución de frecuencias

Conocidos ya todos los elementos teóricos necesarios para la construcción y comprensión de una distribución de frecuencias vamos a proceder a mostrar los pasos requeridos para su ejecución.

1° Determinar el recorrido o rango

$$R = Xmáx - Xmín \qquad 3.1$$

2° Calcular el intervalo de clase, siempre que se conozca el número de clases.

$$Ci = \frac{R}{NC} \qquad 3.2$$

3° Calcular el número de clases, siempre que se conozca el intervalo de clase.

$$NC = \frac{R}{Ci} \qquad 3.3$$

Como se puede observar en el segundo y tercer paso resultaría muy difícil resolver estas ecuaciones por simples métodos matemáticos ya que cada una de ellas presenta dos incógnitas.

Como solución a este problema surge la fórmula de Sturgees que se expresa así,

$$Ci = \frac{R}{1+3.22\log N} \quad \text{donde,} \qquad 3.4$$

R = recorrido

N = número total de valores

Ejemplo 3.1

A continuación se muestran las ventas de un mes, en miles de RD$ (pesos), de 30 vendedores de una compañía de artículos ferreteros.

26	29.4	64.3	56.5	48.5	57
30.4	28.6	70	81.4	63.4	76.4
40.5	32.3	50.4	76.9	68	83
27.8	36.9	82.4	78	73	64.3
44.5	35.6	88	67	22	52.8

a) Construya una distribución de frecuencias

Solución:

Procederemos a dar los pasos citados anteriormente; veamos:

R = Xmáx - Xmín

R = 88 - 22 = 66

Como no conocemos el intervalo de clase ni el número de clases vamos a aplicar la fórmula de Sturgees.

$$Ci = \frac{R}{1+3.22\log N}$$; sustituyendo tenemos,

$$Ci = \frac{66}{1+3.22\log 30} = \frac{66}{5.75} = 11.5$$

Como se puede observar, según los resultados de la fórmula anterior el valor de Ci = 11.5 y el NC = 5.75; sucede que el NC (número de clases) es una variable discreta que no admite fraccionamiento y por consiguiente no puede ser, en la práctica, igual a 5.75, por lo cual debemos aproximar a un valor entero, que en este caso sería, NC = 6.

De esta manera y por aproximación,

NC = 6

Ci = 12

Dado los pasos anteriores procederemos a la construcción de la distribución de frecuencias.

Tabla 3.1 Ventas de un mes de 30 vendedores en miles de RD$

Ventas (Miles RD$)	Vendedores (fi)	Fr	Fi	Fr	Xi
22 - 34	𝍸 || = 7	0.233	7	0.233	28
34 - 46	|||| = 4	0.133	7 + 4 = 11	0.366	40
46 - 58	𝍸 = 5	0.167	11 + 5 = 16	0.533	52
58 - 70	𝍸 = 5	0.167	16 + 5 = 21	0.70	64
70 - 82	𝍸 | = 6	0.20	21 + 6 = 72	0.90	76
82 - 94	||| = 3	0.10	27 + 3 = 30	1.00	88
Total	30	1.00			

Los diferentes indicadores obtenidos anteriormente se interpretan de la manera siguiente. Por ejemplo, si le interesa interpretar el resultado de la frecuencia simple de la quinta clase, puede expresar que: seis vendedores efectuaron ventas durante ese mes entre 70 mil y menos de 82 mil pesos.

Asimismo, puede señalar, al interpretar la frecuencia relativa simple de la cuarta clase, que el 16.7% de los vendedores realizó ventas durante ese mes por montos comprendidos entre 58 mil y menos de 70 mil pesos.

Al interpretar el resultado de frecuencia acumulada de la quinta clase podemos señalar que 27 vendedores durante ese mes realizaron ventas por un monto comprendido entre 22 mil pesos y menos de 82 mil pesos.

De igual manera el 53.3% de los vendedores efectuaron ventas durante ese mes por montos comprendidos entre 22 mil y 58 mil pesos, explicando este resultado el contenido de la frecuencia relativa acumulada de la tercera clase de la distribución.

En lo referente al punto medio de cada clase, este es usado para representar mediante un solo valor el recorrido de cada clase y sirve además para los fines de análisis estadístico de los datos.

Es importante señalar en relación a la construcción de una distribución de frecuencias que el lector o usuario tiene plena libertad en la pre-escogencia del intervalo de clase, en función de la naturaleza de los datos y de su conveniencia técnica.

Ejemplo 3.2

Construir una distribución de frecuencias con las puntuaciones del coeficiente de inteligencia de 70 personas tomadas al azar. Considere un intervalo de clases igual a 10.

64	142	107	67	125	124	91
102	135	105	82	114	126	93
107	128	123	91	116	133	100
151	51	104	101	118	123	98
90	67	112	107	123	121	115
93	73	116	105	107	118	153
109	78	98	104	140	117	147
121	89	95	98	143	114	118
123	102	89	83	135	101	120
84	76	158	78	136	107	105

Solución:

Tal y como señalamos anteriormente, para construir una distribución de frecuencia es preciso conocer:

1.- R = Xmáx = Xmín

R = 158 - 51 = 107

2.- Ci = 10 (conocido previamente)

3.- Nc = número de clases

$$NC = \frac{R}{Ci} = \frac{107}{10} = 10.7 = 11$$

Tendremos una distribución de frecuencias con un intervalo de clases igual a diez, (Ci = 10), y unas once clases.

Tabla 3.2 Coeficiente de inteligencia de 70 personas

Coeficiente de inteligencia	Frec, simple (fi)	Frec, relativa simple (fr)	Frec. Acumulada (Fi)	Frec. Relativa Acumulada (Fr)	Punto Medio (Xi)
50 – 60	| = 1	1/70 = 0.014	1	0.014	55
60 – 70	||| = 3	3/70=0.043	1+3= 04	0.014+0.043=0.057	65
70 – 80	|||| = 4	4/70 =0.057	4+4=08	0.057+0.057=0.114	75
80 – 90	𝍸 = 5	5/70=0.072	8+5=13	0.114+0.072=0.186	85
90 – 100	𝍸 |||| = 9	9/70=0.129	13+9=22	0.186+0.129=0.315	95
100 - 110	𝍸 𝍸 𝍸 | =16	16/70=0.228	22+16=38	0.315+0.228=0.543	105
110 – 120	𝍸 𝍸 =10	10/70=0.143	38+10=48	0.543+0.143=0.686	115
120 – 130	𝍸 𝍸 | =11	11/70=0.157	48+11=59	0.686+0.157=0.843	125
130 – 140	|||| = 4	4/70=0.057	59+4=63	0.843+0.057=0.90	135
140 – 150	|||| = 4	4/70=0.057	63+4=67	0.90+0.057=0.957	145
150 – 160	||| = 3	3/70=0.043	67+3=70	0.957+0.043=1.00	155
Total	70	1.00			

a) Interprete el valor de la frecuencia acumulada de la sexta clase.

Significa que 38 personas tienen un coeficiente de inteligencia por debajo de 110.

b) Interprete el valor de la frecuencia relativa de la octava clase.

Significa que el 15.7% de las personas tienen un coeficiente de inteligencia comprendido entre 120 y 130.

EJERCICIOS DE APLICACIÓN

3.1 A un grupo de 52 estudiantes universitarios se les cuestionó sobre su edad, sexo, estado civil, carrera que cursa en la universidad y su condición laboral; a continuación se presentan las respuestas ofrecidas por éstos:

Edad (años)	Sexo	Estado civil	Carrera que estudia	Condición laboral
25	Masculino	Soltero	Pedagogía	Trabaja
26	Masculino	Casado	Pedagogía	Trabaja
19	Femenino	Soltero	Mercadeo	Trabaja
22	Femenino	Casado	Administración	No trabaja
21	Femenino	Soltero	Mercadeo	No trabaja
25	Femenino	Unión libre	Pedagogía	Trabaja
21	Femenino	Soltero	Contabilidad	No trabaja
31	Femenino	Casado	Pedagogía	Trabaja
18	Femenino	Soltero	Ingeniería	No trabaja
20	Masculino	Soltero	Mercadeo	No trabaja
24	Femenino	Casado	Administración	No trabaja
19	Masculino	Soltero	Ingeniería	No trabaja
22	Masculino	Soltero	Ingeniería	No trabaja
19	Masculino	Soltero	Ingeniería	No trabaja
22	Femenino	Soltero	Contabilidad	Trabaja
31	Masculino	Casado	Pedagogía	Trabaja
30	Femenino	Divorciado	Pedagogía	Trabaja
23	Femenino	Soltero	Derecho	No trabaja
24	Femenino	Soltero	Derecho	Trabaja
25	Femenino	Casado	Derecho	Trabaja
24	Femenino	Soltero	Mercadeo	No trabaja
28	Femenino	Divorciado	Pedagogía	Trabaja
27	Femenino	Unión libre	Pedagogía	No trabaja
24	Masculino	Soltero	Contabilidad	Trabaja
26	Femenino	Unión libre	Contabilidad	Trabaja
25	Femenino	Casado	Contabilidad	Trabaja
21	Femenino	Soltero	Mercadeo	No trabaja
20	Femenino	Soltero	Administración	No trabaja
25	Femenino	Casado	Administración	No trabaja
26	Femenino	Casado	Administración	No trabaja
23	Masculino	Soltero	Administración	Trabaja
27	Masculino	Casado	Ingeniería	Trabaja
28	Masculino	Casado	Administración	Trabaja
29	Masculino	Unión libre	Derecho	Trabaja
27	Femenino	Casado	Derecho	Trabaja
26	Femenino	Casado	Pedagogía	Trabaja
25	Femenino	Soltero	Pedagogía	No trabaja
23	Masculino	Soltero	Administración	No trabaja
22	Femenino	Soltero	Ingeniería	No trabaja
24	Femenino	Soltero	Ingeniería	No trabaja
30	Masculino	Divorciado	Pedagogía	Trabaja
22	Masculino	Soltero	Contabilidad	No trabaja
21	Femenino	Soltero	Contabilidad	No trabaja
24	Masculino	Soltero	Contabilidad	No trabaja
25	Masculino	Soltero	Mercadeo	No trabaja
26	Femenino	Casado	Mercadeo	Trabaja
22	Femenino	Soltero	Mercadeo	No trabaja
24	Masculino	Soltero	Pedagogía	No trabaja
23	Masculino	Soltero	Pedagogía	No trabaja
19	Femenino	Soltero	Mercadeo	Trabaja
21	Masculino	Soltero	Contabilidad	No trabaja
24	Femenino	Casado	Contabilidad	No trabaja

Se pide:

a) Construya una distribución de frecuencias de las edades de los estudiantes con intervalo de clase igual a 2 años.

b) Interprete el resultado de la frecuencia relativa acumulada de la cuarta clase.

c) Construya una distribución de los estudiantes según la carrera que cursan.

d) Qué porcentaje son estudiantes de pedagogía.

3.2 Con los datos del ejercicio 3.1 que se refiere a un grupo de 52 estudiantes universitarios acerca de su edad, sexo, estado civil, carrera que cursa y condición laboral, se pide:

a) Construya una distribución de los estudiantes según el estado civil.

b) Qué porcentaje de los estudiantes son casados.

c) Construya una distribución de los estudiantes según carrera y sexo.

d) Calcule el porcentaje de estudiantes para cada carrera según el sexo.

3.3 El salario (RD$) promedio semanal que devengaban los operarios de parques de zonas francas durante el año 2001 se presenta a continuación:

700.00	1,159.50	692.11	761.68	730.16
724.96	783.33	742.44	1,150.00	855.23
738.21	866.73	720.88	736.33	704.87
592.82	667.58	900.00	575.00	838.82
667.00	700.00	821.14	963.88	700.00
860.15	1,000.00	680.92	755.00	743.43
719.36	784,33	854.00	617.00	1,270.00
601.33	711.75	650.00	575.00	802.00
610.00	880.00	744.25	764.92	780.00
797.00	660.51	650.00	721.14	792.30

Fuente: Consejo Nacional de Zonas Francas, informe estadístico 2001.

a) Construya una distribución de frecuencias con intervalo igual a RD$ 100.00

b) ¿Cuántos parques de zonas francas pagan a sus operarios menos de RD$ 975.00 en promedio por semana?

c) ¿Qué porcentaje de parques de zonas francas pagan a sus operarios RD$ 875.0 o más en promedio por semana?

3.4 Las siguientes informaciones se refieren al número de pasajeros que abordaron 100 autobuses que salieron de su terminal durante un día de trabajo.

42	45	49	50	30	38	50	42	48	64
52	51	37	32	41	46	47	52	53	57
46	48	56	48	47	47	42	57	36	60
52	47	47	51	35	53	56	53	55	65
31	64	41	45	54	50	45	55	68	56
51	50	49	51	45	40	54	50	53	49
52	55	57	44	43	56	51	54	43	48
54	49	38	53	49	43	39	51	38	52
59	43	50	55	42	48	58	37	46	57
63	57	53	46	52	51	38	46	44	63

a) Construya una distribución de frecuencia con intervalo de clase igual a 5 pasajeros.

b) ¿Cuántos autobuses salieron de su terminal con 55 pasajeros o más?

c) ¿Qué porcentaje de autobuses salieron de su terminal con 54 pasajeros o menos?

3.5 El siguiente conjunto de valores se refiere a las horas-hombre necesarias para que una compañía de brillado de pisos termine esta actividad en 50 casas de un proyecto habitacional.

25	28	76	110	46	145	98	75	56	157
35	36	92	115	62	138	83	92	69	120
40	46	72	32	84	125	85	126	81	100
62	80	75	21	93	155	76	131	121	105
74	84	100	38	150	132	68	110	143	89

a) Construya una distribución de frecuencias con intervalo igual a 20 horas-hombre.

b) ¿Cuántos pisos son brillados en menos de 80 horas-hombre?

c) ¿Qué porcentaje de pisos son brillados entre 40 y menos de 60 horas-hombre?

3.6 El número de accidentes de tránsito por día ocurridos en una ciudad, durante un mes, aparecen a continuación:

8	0	9	6	1	12
3	2	4	2	8	2
10	4	5	4	11	3
3	3	0	3	8	4
7	5	13	7	9	5

a) Construya una distribución de frecuencias con intervalo de clase igual a dos accidentes.

3.7 Las calificaciones finales de los 50 estudiantes de Estadística Básica del semestre 2006-1 se presentan a seguidas:

90	34	50	78	68	64	39	61	89	45
60	77	89	77	70	75	94	75	81	57
76	47	90	74	88	83	44	67	90	81
58	79	86	78	72	66	58	52	56	93
73	85	82	98	76	72	63	48	83	80

a) Construya una distribución de frecuencias con intervalo de clase 10 puntos.

b) Cuántos estudiantes obtuvieron menos de 70 puntos.

c) Que porcentaje de estudiantes obtuvo 70 puntos ó más.

3.8 Los montos de 50 préstamos personales (en miles de RD$) desembolsados por un banco comercial de Santo Domingo fueron los siguientes:

635	1000	375	2225	945
935	560	1180	950	2450
520	970	305	2125	2700
475	665	1700	450	1750
1900	725	1525	785	985

1225	1375	1875	630	750
2530	850	590	2990	875
575	335	940	1475	1500
1650	1450	875	650	775
1185	660	650	725	1675

a) Construya una distribución de frecuencias con intervalo de clases igual a RD$400,000.00.

b) Construya una distribución de frecuencias con intervalo de clases igual a RD$300,000.00.

3.9 Con los datos del ejercicio 3.1 de la pag. 67 determine el porcentaje de estudiantes universitario por carrera. Así mismo, diga que carrera es la que posee la mayor participación porcentual de estudiantes.

3.10 Las puntuaciones siguientes se refieren al coeficiente de inteligencia de 70 estudiantes en una escuela del nivel medio .

80	125	92	143	100	95	120	140	84	110
90	128	120	149	110	105	117	130	90	105
105	110	119	135	135	119	110	125	117	119
100	106	114	140	126	130	95	115	118	125
96	113	132	126	130	145	90	110	116	100
85	115	135	90	113	142	84	105	120	141
75	89	140	95	118	135	89	95	130	126

a) Construya una distribución de frecuencias con intervalo de clase igual a 10.

b) Explique el resultado obtenido en la quinta clase de la frecuencia relativa simple.

c) Explique el resultado obtenido en la cuarta clase de la frecuencia acumulada.

3.11 En la ejecución de un estudio para determinar la situación de empleo de un grupo de profesionales, se seleccionó una muestra de éstos de manera aleatoria para conocer la composición por categoría profesional.

- Administrador	- Economista	- Abogado	- Ingeniero
- Contador	- Abogado	- Administrador	- Médico
- Abogado	- Contador	- Ingeniero	- Economista
- Administrador	- Abogado	- Contador	- Ingeniero
- Abogado	- Médico	- Médico	- Contador
- Químico	- Maestro	- Químico	- Administrador
- Maestro	- Contador	- Maestro	- Abogado

a) Determine la cantidad de profesionales por cada categoría mediante la tabla correspondiente.

b) ¿Qué porcentaje eran maestros?

c) ¿Qué porcentaje eran ingenieros?

CAPITULO

IV

Presentación de datos

4.0 Introducción

La presentación de datos estadísticos constituye en sus diferentes modalidades uno de los aspectos de más uso en la estadística descriptiva.

A diario podemos visualizar a través de los diferentes medios escritos y televisivos de comunicación masiva la presentación de los datos estadísticos sobre el comportamiento de las principales variables económicas y sociales, nacionales e internacionales.

Observar una gráfica que muestre con rapidez el comportamiento de las exportaciones de bienes y servicios de la República Dominicana durante un período de tiempo; analizar con alto nivel de precisión una tabla o cuadro que contiene los datos sobre la producción agropecuaria según rubros durante un período de tiempo determinado o simplemente usar la palabra escrita como forma de describir los hechos cuando una serie contiene pocos datos, son los conceptos que sobre estadística descriptiva vamos a estudiar en el presente capítulo.

Existen tres formas diferentes de presentar los datos estadísticos, que son:

1.- Presentación escrita

2.- Presentación tabular

3.- Presentación gráfica

4.1 Presentación escrita

Esta forma de presentación de informaciones se usa cuando una serie de datos incluye pocos valores, por lo cual resulta más apropiada la palabra escrita como forma de describir el comportamiento de los datos. Asimismo, mediante la forma escrita, se resalta la importancia de las informaciones principales.

4.2 Presentación tabular

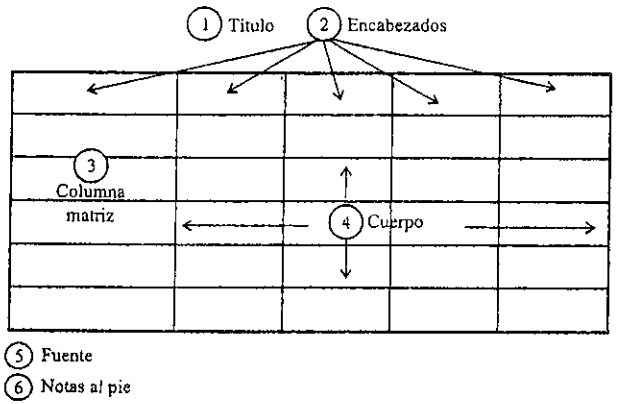
Cuando los datos estadísticos se presentan a través de un conjunto de filas y de columnas que responden a un ordenamiento lógico., a este resultado le llamamos forma de presentación tabular o simplemente tabla o cuadro estadístico.

La forma tabular de presentar los datos estadísticos es de gran uso e importancia para el usuario ya que constituye la forma más exacta de presentar las informaciones.

Una tabla o cuadro consta de varias partes, las principales son las siguientes:

1.- Título
2.- Encabezados
3.- Columna matriz
4.- Cuerpo del cuadro
5.- Fuente
6.- Notas al pie

Figura 4.1 Presentación esquemática de las partes de una tabla.



Título:

Es la parte más importante del cuadro y sirve para describir todo el contenido de éste. Puedo señalar que el título es como una fotografía donde aparece de manera completa y compacta todas las variables contenidas en la tabla.

Un título completo y correcto siempre debe responder las preguntas: qué?, cómo?, cuándo?, dónde?

Ejemplo 4.1

Alumnos matriculados del sector público según nivel, año lectivo 2005-2006

Veamos:

Qué estamos estudiando? Los alumnos matriculados

Dónde? En el sector público

Cómo? Según nivel

Cuándo? Año lectivo 2005-2006

Como se puede observar, este título es correcto ya que responde fielmente a las preguntas señaladas anteriormente.

Encabezados

Los encabezados son los diferentes subtítulos que se colocan en la parte superior de cada columna.

Una tabla, en el caso más sencillo puede tener sólo dos encabezados y por consiguiente dos columnas. En caso de cuadros o tablas más complejas habrá más de dos encabezados y en algunas ocasiones tendrán encabezados principales y sub-encabezados.

Columna matriz

Es la columna principal del cuadro. Esta columna se encuentra al lado izquierdo de la tabla y colocaremos en ella la variable principal bajo estudio.

Cuerpo

El cuerpo contiene todas las informaciones numéricas que aparecen en la tabla.

Los datos que aparecen en el cuerpo responden a las descripciones y clasificaciones de los encabezados y la columna matriz.

Fuente:

La fuente de los datos contenidos en la tabla indica la procedencia de éstos. Cuando se usa una tabla con informaciones que ya han sido elaboradas por otros, es un deber señalar en la fuente la institución que las recopiló y presentó. Sin embargo, si las informaciones fueron recopiladas y presentadas por la misma persona o institución no es necesario señalar la fuente.

Notas al pie:

Las notas al pie son usadas para hacer algunas aclaraciones sobre aspectos que aparecen en la tabla o cuadro y que no han sido explicados en otras partes.

Para ilustrar de manera práctica la forma operativa de cómo funcionan cada una de las partes del cuadro, vamos a realizar el formato de un cuadro.

Ejemplo 4.2

Construya el formato de un cuadro con las informaciones que aparecen en el siguiente título:

Cantidad de alumnos matriculados en Básica y en Media del sector público nacional, según provincias, 2005-2006.

Tabla 4.1 Alumnos matriculados (formato)

Provincia	Matrícula		
	Total	Básica	Media
Total	xxxxx	xxxxx	xxxxx
Distrito Nacional	xxxx	xxxx	xxxx
Azua	xxx	xxx	xxx
Bahoruco	xxx	xxx	xxx
Barahona	xxx	xxx	xxx
..	..	..	..
..	..	..	..
..	..	..	..
..	..	..	..
..	..	..	..
Monte Plata	xxx	xxx	xxx
Hato Mayor	xxx	xxx	xxx

Fuente: Secretaría de Estado de Educación.

Ejemplo 4.3

Intercambio comercial de la República Dominicana, (exportaciones e importaciones nacionales) según valor de éstas 1992-1999.

Tabla 4.2 Intercambio comercial

Año	Valor FOB en miles de US$		
	Intercambio Comercial	Exportaciones generales	Importaciones generales
1992	2,738,039	563,439	2,174,600
1993	2,622,429	504,029	2,118,400
1994	3,189,015	625,915	2,563,100
1995	4,036,300	872,100	3,164,200
1996	4,526,200	945,500	3,580,700
1997	5,209,300	1,017,300	4,192,000
1998	5,776,900	880,300	4,896,600
1999	6,251,800	6,251,800	5,379,600

Fuente: Banco Central de la República Dominicana

Ejemplo 4.4

La población del Distrito Nacional en 1950 fue de 239,464 habitantes, en 1960 de 466,830 habitantes; en 1970 de 813,420 habitantes; en 1981 fue de 1,540,786 y en 1993 fue de 2,193,046 habitantes.

En 1950, la población de Azua fue de 50,608 habitantes; en 1960 de 74,400 personas; en 1970 de 90,590 personas; en 1981 de 140,914 y en 1993 de 199,684 habitantes.

La provincia de Barahona tenía en 1950 unos 62,166 habitantes; en 1960 unos 80,030 habitantes; en 1970 tenía 111,162 habitantes; en 1981 unos 141,313 habitantes; en 1993 tenía 164,835 habitantes.

La provincia Duarte en 1950 tenía una población de 164,400 habitantes; en 1960 de 162,370 habitantes; en 1970 eran unos 200,498 habitantes; asimismo en 1981 habían unos 227,798 habitantes y en 1993 la población era de 281,879 habitantes.

Asimismo, la provincia de Puerto Plata en 1950 tenía 133,652 habitantes; en 1960 tenía 163,960; en 1970 tenía

186,112; en 1981 tenía 201,893 habitantes y en 1993 tenía una población de 261,485 habitantes.

Estos datos fueron obtenidos de la revista República Dominicana en cifras, 1993, que publica la Oficina Nacional de Estadística.

Con los datos anteriores, construya el cuadro o tabla correspondiente.

Solución

Población de algunas provincias de la República Dominicana, según censos, 1950-1993.

Tabla 4.3 Población de algunas provincias de República Dominicana

Provincias	Censos				
	1950	1960	1970	1981	1993
Distrito Nacional	239,464	466,830	813,420	1,540,786	2,193,046
Azua	50,608	74,400	90,590	140,914	199,684
Barahona	62,166	80,030	111,162	141,313	164,835
Duarte	164,400	162,370	200,498	227,798	281,879
Puerto Plata	133,652	163,960	186,112	201,893	261,485

Fuente: Oficina Nacional de Estadística

4.3 Presentación gráfica

La presentación gráfica de las informaciones estadísticas le proporciona al lector o usuario mayor rapidez en la comprensión de los datos. Esta forma de presentación le da al usuario un valor aproximado de la información; a diferencia de la presentación tabular que le ofrece exactitud.

Una gráfica podríamos definirla como una expresión artística usada para representar un conjunto de datos.

Su uso es bastante amplio en estadística, para presentar datos o para comparar varios grupos de datos, dado el hecho de que casi todas las informaciones de características cuantitativas pueden ser expresadas a través de gráficas.

De acuerdo al tipo de variable que vamos a representar, las principales gráficas son las siguientes:

1.- Histograma
2.- Polígono de frecuencias
3.- Gráfica de barras
4.- Gráfica lineal
5.- Gráfica de barra 100%
6.- Gráfica circular
7.- Ojiva

A seguidas vamos a tratar cada una de estas gráficas.

Histograma

Un histograma es un conjunto de barras o rectángulos unidos uno de otro, en razón de que lo utilizamos para representar variables continuas.

La altura de cada barra o rectángulo viene dada por la frecuencia simple de clase.

La base de cada rectángulo o barra será igual al valor del intervalo de cada clase.

Cuando los intervalos de clase son iguales, las bases, por consiguiente también serán iguales; cuando ocurre lo contrario entonces será necesario realizar los ajustes de lugar.

Para mayor ilustración le presentamos a continuación un histograma.

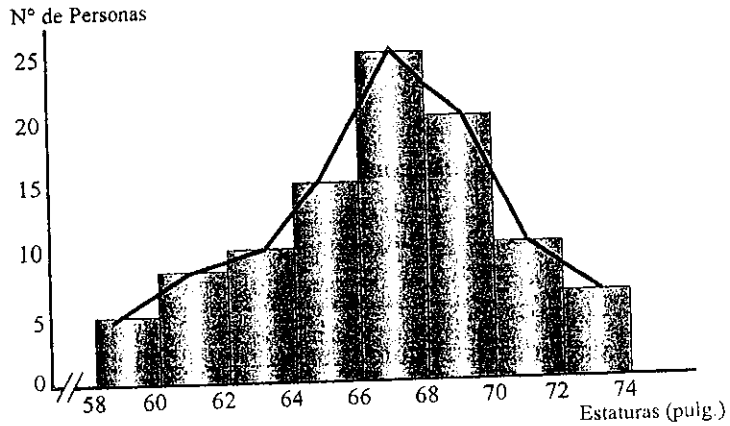
Ejemplo 4.5

Las estaturas (pulgadas) de cien personas que realizan transacciones durante un día en un banco de Santo Domingo son las siguientes

Tabla 4.4 Estaturas de 100 personas.

Estaturas (Pulg.)	No. De Personas
58-60	5
60-62	9
62-64	10
64-66	15
66-68	25
68-70	20
70-72	10
72-74	6
Total	**100**

Figura 4.2 Estaturas de personas que realizaron transacciones durante un día en un banco de Santo Domingo.



Polígono de frecuencias

Esta gráfica se usa para representar los puntos medios de clase en una distribución de frecuencias.

El procedimiento para construir un polígono de frecuencias consiste en determinar cada punto medio de clase y luego se marca en la base superior de cada rectángulo.

Nota: Lo representaremos usando el histograma anterior.

Gráfica de barras

Es un conjunto de rectángulos o barras separadas una de otra, en razón de qué se usa para representar variables discretas.

La altura de cada una de las barras o rectángulos viene dada por la frecuencia de clase.

Asimismo, las barras deben ser de igual base o ancho y separadas a igual distancia.

Las barras o rectángulos en una gráfica de barras pueden disponerse en forma vertical y horizontal, dependiendo de la preferencia de quien la construye.

Dentro del mismo concepto de las gráficas de barras vamos a tener además gráficas de barras compuestas y gráficas de barras componentes.

Estos tipos de gráficas de barras nos permiten hacer comparaciones entre una o varias series de datos. Estas comparaciones pueden ser expresadas gráficamente en valores absolutos (valores reales de la variable) o en valores relativos (porcentajes).

Para mayor ilustración, vamos a presentar a los lectores diferentes ejemplos de las gráficas de barras antes señaladas.

Ejemplo 4.6

Cantidad de estudiantes universitarios según provincias del Cibao Central, censo 1993.

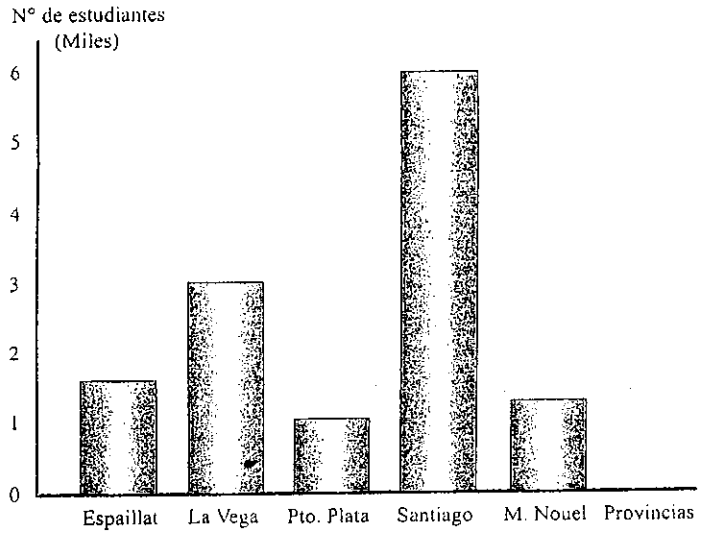
Tabla 4.5 Estudiantes Universitarios del Cibao Central

Provincias Cibao Central	Estudiantes Universitarios
Espaillat	1,625
La Vega	2,977
Puerto Plata	1,066
Santiago	5,990
Monseñor Nouel	1,333

Fuente: Oficina Nacional de Estadística

Construir una gráfica de barras con los datos anteriores.

Figura 4.3 Cantidad de estudiantes universitarios según provincias del Cibao Central, censo 1993. (Miles de estudiantes)



Fuente: Oficina Nacional de Estadística

Ejemplo 4.7

Ingresos y gastos del gobierno central de la República Dominicana, 1994-1999. (Miles de Millones de RD$)

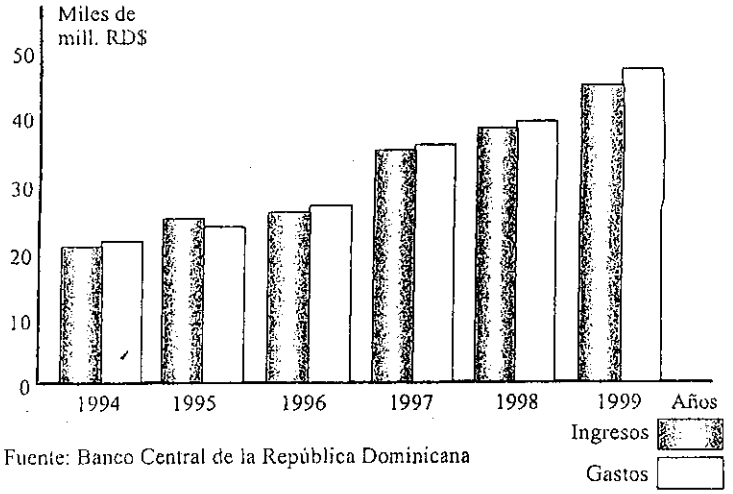
Tabla 4.6 Ingresos y gastos del Gobierno Central

Años	Ingresos	Gastos
1994	22,497.5	22,933.1
1995	25,423.2	24,107.4
1996	26,888.6	27,691.1
1997	35,308.2	35,661.8
1998	39,526.9	39,955.2
1999	46,692.1	48,758.7

Fuente: Banco Central de la República Dominicana

Construir una gráfica de barras compuestas con los datos anteriores.

Figura 4.4 Ingresos y Gastos del gobierno central de la República Dominicana, 1994-1999. (Miles de Millones de RD$)



Fuente: Banco Central de la República Dominicana

Ejemplo 4.8

Participación en el producto bruto interno de los sectores: Construcción, transporte y comunicaciones, 1995-1999.

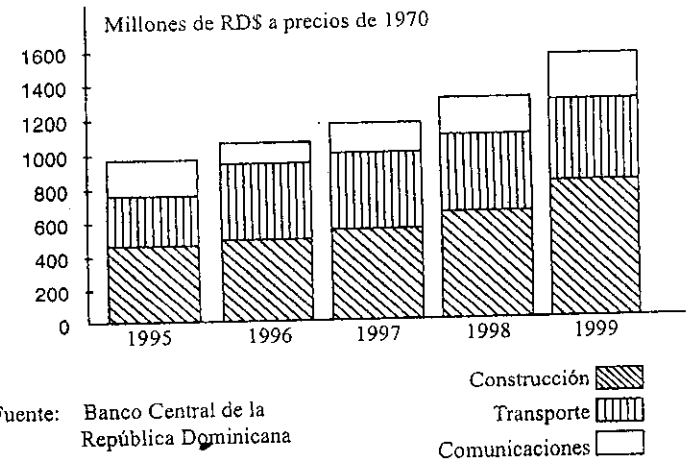
Tabla 4.7 Participación (en millones de RD$ a precios de 1970)

Años	Construcción	Transporte	Comunicaciones
1995	442.3	310.6	159.7
1996	501.6	335.3	185.7
1997	587.2	367.9	221.5
1998	702.1	397.9	267.0
1999	830.9	431.8	308.7

Fuente: Banco Central de la República Dominicana

Construir una gráfica de barras componentes con los datos anteriores.

Figura 4.5 Participación en el PBI de los sectores construcción, transporte y comunicaciones, 1995-1999. (En millones de RD$. A precios de 1970)



Fuente: Banco Central de la República Dominicana

Gráfica lineal

Las gráficas de líneas son usadas principalmente para representar datos clasificados por cantidad o tiempo. O sea, se usan para representar series de tiempo o cronológicas.

Veamos a continuación un ejemplo de la construcción de una gráfica lineal.

Ejemplo 4.9

Valor de las importaciones nacionales a República Dominicana según años, 1992-1999. (Valor FOB en miles de US$)

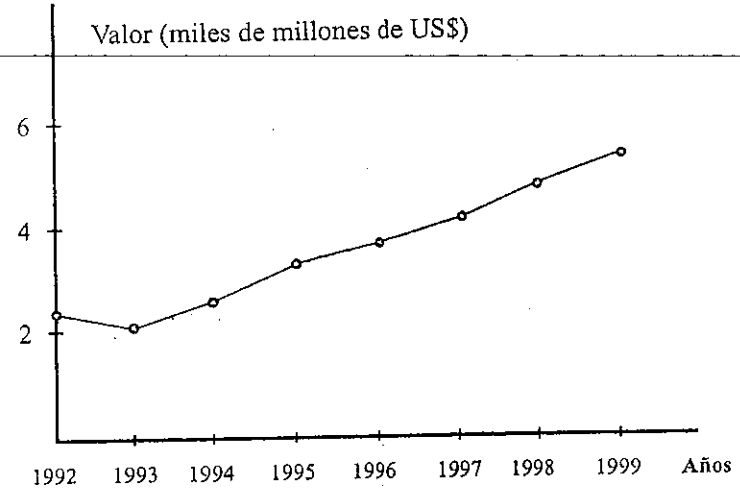
Tabla 4.8 Valor importaciones

Años	Valor de las importaciones (en miles de US$)
1992	2,174,600
1993	2,118,400
1994	2,563,100
1995	3,164,200
1996	3,580,700
1997	4,192,000
1998	4,896,600
1999	5,379,600

Fuente: Banco Central de la República Dominicana

Construir una gráfica de líneas con los datos anteriores.

Figura 4.6 Valor de las importaciones nacionales a República Dominicana según años, 1992-1999. (Valor FOB en miles de millones de US$)



Fuente: Banco Central de la República Dominicana

Gráfica de barra 100% y Gráfica circular

Las gráficas de barra 100% y la gráfica circular se usan especialmente para representar las partes en que se divide una cantidad total. En el caso de la gráfica de barra 100% los valores se convierten a porcentajes y en la gráfica circular se trabaja con grados tomando como base que el perímetro de la circunferencia mide 360 grados.

Ejemplo 4.10

Volumen de producción azucarera en la República Dominicana, según empresas, zafra 1996-1997. (En toneladas métricas)

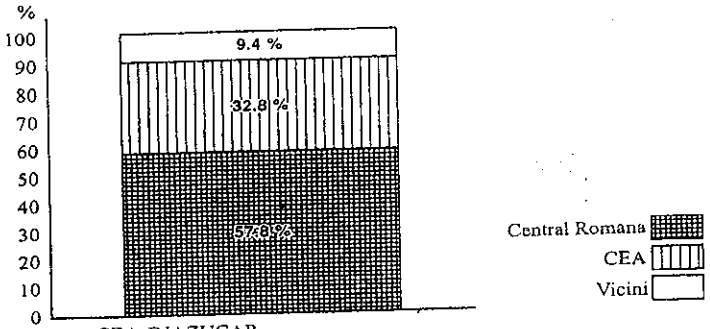
Tabla 4.9 Volumen producción azucarera

Empresas	Volumen de Producción (toneladas métricas)	%
Consejo Estatal del Azúcar	205,651	32.8
Central Romana	362,876	57.8
Vicini	58,967	9.4
Total	627,494	100.0

Fuente: CEA-INAZUCAR

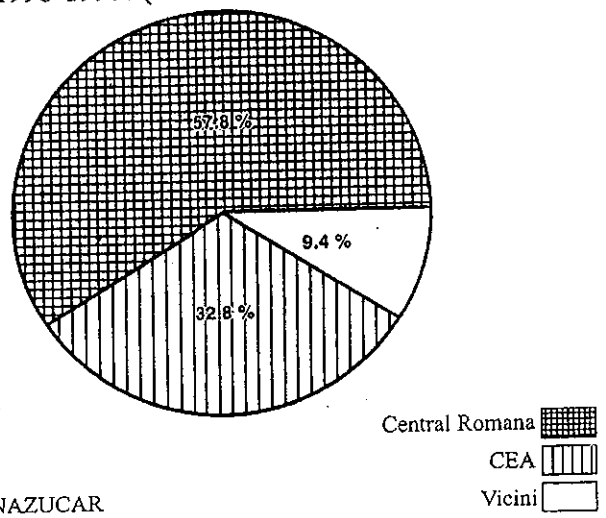
Construir una gráfica de barra 100% y otra circular con los datos referentes al volumen de producción azucarera antes citados.

Figura 4.7 Volumen de producción azucarera en la República Dominicana, según empresas, zafra 1996-1997. (En toneladas métricas)



Fuente: CEA-INAZUCAR

Figura 4.8 Volumen de producción azucarera en la República Dominicana, según empresas, zafra 1996-1997. (En toneladas métricas)



Fuente: CEA-INAZUCAR

La ojiva

Esta gráfica consiste en la representación de las frecuencias acumuladas de una distribución de frecuencias.

La ojiva puede construirse de dos maneras diferentes; sobre la base "menor que" o sobre la base "o más".

Mediante la construcción de la ojiva se puede determinar el valor de la mediana de la distribución.

Ejemplo 4.11

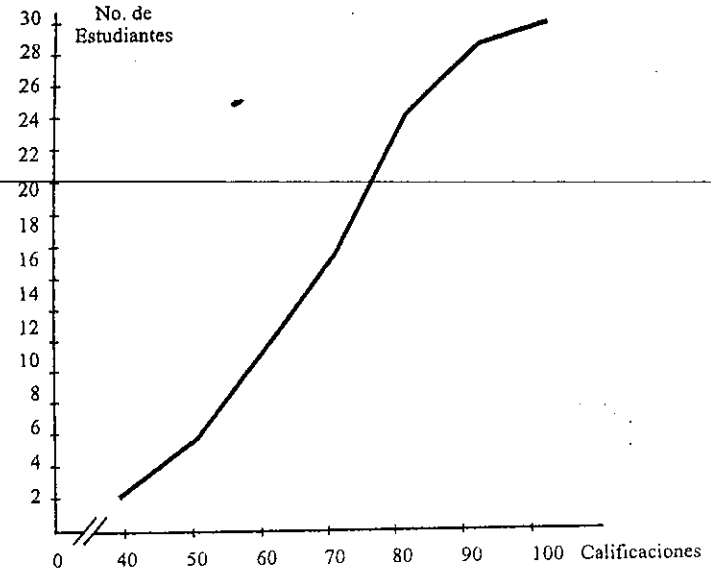
Las calificaciones de 30 estudiantes de estadística están dadas en la siguiente tabla:

Tabla 4.10 Calificaciones de 30 estudiantes

Calificaciones	No. de Estudiantes	Fi
Menos de 40	2	2
40-50	3	2+3= 5
50-60	5	5+5=10
60-70	6	10+6=16
70-80	8	16+8=24
80-90	4	24+4=28
90-100	2	28+2=30
Total	**30**	

Construir una ojiva con los datos anteriores.

Figura 4.9 Calificaciones de 30 estudiantes de estadística



EJERCICIOS DE APLICACIÓN

4.1 La llegada mensual de pasajeros por vía aérea a la República Dominicana y su condición de residencia durante el año 2001 se presenta a continuación:

Durante el mes de ***enero*** arribaron un total de 28,980 pasajeros residentes y 278,081 pasajeros no residentes. En el mes de ***febrero*** llegaron 20,469 pasajeros residentes y 282,470 pasajeros no residentes. Asimismo, durante el mes de ***marzo*** un total de 22,152 pasajeros residentes arribaron al país y 294,231 pasajeros no residentes.

El mes de ***abril*** llegaron 28,733 pasajeros residentes y 256,837 pasajeros no residentes. Durante el mes de ***mayo*** unos 25,068 pasajeros residentes y 197,044 pasajeros no residentes penetraron por vías aéreas al país. En ***junio*** 23,625 pasajeros residentes y 217,525 pasajeros no residentes arribaron por vía aérea.

Igualmente, durante el mes de ***julio*** llegaron unos, 30, 576 pasajeros residentes en el país y unos 290,143 pasajeros no residentes. En ***agosto*** la llegada de pasajeros fue como sigue: residentes 37,152; no residentes 241,445. durante el mes de ***septiembre*** arribaron 25,757 pasajeros residentes y 154,120 pasajeros no residentes.

Asimismo, en ***octubre*** del 2001 llegaron al país por vía aérea unos 20,103 pasajeros residentes y 155,528 pasajeros no residentes. En el mes de ***noviembre*** unos 21,723 pasajeros residentes y 169,100 pasajeros no residentes arribaron al país. Finalmente, durante el mes de ***diciembre*** del año 2001 llegaron al país por vía aérea 26,409 pasajeros residentes y 241,279 pasajeros no residentes.

Nota: Estos datos fueron suministrados por el Banco Central de la República Dominicana.

Se pide:

a) Construir una tabla con los datos anteriores

b) Determinar el porcentaje de pasajeros residentes y no residentes según el mes de llegada.

4.2 La cantidad de vehículos existentes en las principales provincias de República Dominicana y resto del país se detallan, por años, a continuación:

Distrito Nacional: en 1995 tenía 197,440 vehículos, en 1996 tenía 304,801 y en 1997 tenía 382,246.

Santiago: en 1995 tenía unos 39,883 vehículos; en 1996 existían unos 50,255 y en 1997 unos 59,096 vehículos.

La Vega: en 1995 tenía 8,743 vehículos; en 1996 tenía 10,804 y en 1997 tenía unos 11,881 vehículos.

Barahona: en 1995 registró un total de 1,223 vehículos; en 1996 unos 1,461 vehículos y en 1997 unos 1,594 vehículos.

Duarte: en 1995 tenía 6,075 vehículos registrados, en 1996 unos 7,043 y en 1997 tenía 7,161 vehículos.

Espaillat: registró en 1995 unos 5,334 vehículos, en 1996 unos 6,609 vehículos y en 1997 unos 8,171 vehículos.

Puerto Plata: en 1995 tenía 4,842 vehículos, en 1996 tenía 6,130 vehículos y en 1997 tenía registrado unos 6,417 vehículos.

San Cristóbal: registró en 1995 unos 4,121 vehículos; en 1996 unos 4,611 y en 1997 unos 5,464 vehículos.

La Romana: en 1995 tenía 4,338 vehículos; en 1996 tenía 5,469 y en 1997 tenía 6,633 vehículos.

Peravia: en 1995 tenía 3,626 vehículos, en 1996 tenía 3,896 vehículos y en 1997 tenía 4,265 vehículos.

San Pedro de Macorís: en 1995 registró 3,688 vehículos, en 1996 unos 4,519 y en 1997 tenía 4,962 vehículos.

Monseñor Nouel: en 1995 tenía 3,080 vehículos, en 1996 tenía 3,711 vehículos y en 1997 unos 4,097 vehículos registrados.

Resto del país: en 1995 tenía 18,840 vehículos, en 1996 tenía 21,966 vehículos y en 1997 tenía 26,118 vehículos.

Nota: Estos datos fueron suministrados por la Dirección General de Rentas Internas de la República Dominicana.

a) Construya una tabla con los datos anteriores.

b) Calcule el porcentaje de vehículos por provincias y resto del país para cada año.

4.3 Cantidad de estudiantes matriculados en el nivel inicial del sector privado, por provincias, según sexo, República Dominicana 2005 – 2006.-

Fuente: Secretaría de estado de Educación.

a) Construya el formato de la tabla correspondiente.

4.4 El producto bruto interno de la República Dominicana (en millones de RD$ a precios de 1970) se presenta conjuntamente con el producto bruto interno del *sector* agropecuario para el periodo 1995-1999.

Año 1995: PBI total = RD$4,587.2 millones; PBI Agropecuario: RD$ 577.0 millones

Año 1996: PBI total = RD$4,919.8 millones; PBI Agropecuario: RD$ 629.8 millones

Año 1997: PBI total = RD$5,323.8 millones; PBI Agropecuario: RD$ 650.4 millones

Año 1998: PBI total = RD$5,713.8 millones; PBI Agropecuario: RD$ 657.4 millones

Año 1999: PBI total= RD$6,188.2 millones; PBI Agropecuario: RD$ 702.3 millones.

Fuente: Banco Central de la República Dominicana.

a) Construya una tabla con el Producto Bruto Interno (PBI) total y el Producto Interno (PBI) Agropecuario

b) Calcule la participación porcentual del producto bruto interno agropecuario con relación al producto bruto interno total.

4.5 Construya una tabla con los datos que se presentan a continuación:

Durante el año 1991 los ingresos del Gobierno Central de la República Dominicana fueron como sigue: tributarios RD$ 11,423.3 millones; no tributarios RD$ 1,724.0 millones; y los ingresos extraordinarios ascendieron a RD$ 744.1 millones.

Durante el año 1992 los ingresos del gobierno central fueron como sigue: Tributarios RD$ 15,600.4 millones; no tributarios RD$ 1,680.7 millones e ingresos extraordinarios RD$ 1,270.5 millones. Asimismo, en 1993 los ingresos tributarios del gobierno central fueron RD$ 18,027.6 millones; los no tributarios ascendieron a RD$ 1,384.8 millones y los ingresos extraordinarios sumaron unos RD$ 1,299.2 millones

En 1994 los ingresos tributarios ascendieron a RD$19,219.8 millones; los no tributarios ascendieron a RD$1,363.7 millones y los ingresos extraordinarios ascendieron a RD$1,914.0 millones. Igualmente en 1995, los ingresos se distribuyeron así: tributarios, RD$22,370.9 millones; no tributarios RD$2,344.1 millones e ingresos extraordinarios RD$1,722.2 millones.

En 1996 los ingresos tributarios ascendieron a la suma de RD$ 24,031.0 millones; los no tributarios RD$ 1,970.7 millones y los extraordinarios a RD$ 1,780.7 millones. De igual manera, en 1997 los ingresos del gobierno central se distribuyeron así: tributarios RD$ 31,547.5; no tributarios RD$ 2,633.6 y extraordinarios, RD$ 2,208.3 millones.

Durante el año 1998 los ingresos tributarios alcazaron la suma de RD$ 36,260.4 millones; los no tributarios ascendieron a RD$ 2,129.5 millones y los extraordinarios sumaron RD$ 1,607.4 millones. Finalmente, durante el año 1999 los ingresos fueron como sigue: tributarios, RD$ 40,197.2 millones, no tributarios RD$ 2,476.3 y extraordinarios, RD$ 645.4.

Asímismo, en el año 2000, los ingresos tributarios totalizaron RD$ 47,617.0 millones; los no tributarios ascendieron a RD$ 3,586.1 millones; mientras que los extraordinarios alcanzaron la suma de RD$ 3,456.1 millones.

Finalmente, durante el año 2001, por concepto de ingresos tributarios, el gobierno central recibió RD$ 56,995.6 millones; igualmente, los ingresos no tributarios totalizaron RD$ 2,646.0 millones y los ingresos extraordinarios ascendieron a RD$ 3,108.6 millones.

Nota: Estos datos fueron suministrado por el Banco Central de la República Dominicana.

4.6 Construya el formato de una tabla con el siguiente título.

Valor de las remesas familiares y sociales recibidas en República Dominicana durante el período 1998-2005.

Fuente: Banco Interamericano de Desarrollo.

4.7 Cantidad de estudiantes universitarios de la República Dominicana por universidades según sexo, 2006.

Fuente: Secretaría de estado de Educación Superior Ciencia y Tecnología.

Construir el formato de una tabla con las variables del título anterior.

4.8 La población estimada al 1 de julio del año 2005 en la República Dominicana según estructura de edad se presentan a continuación. Se estima que había 220,224 niños de **menos de un año** a esa fecha. La población existente de **uno a cuatro años** (1-4 años) de edad era de 813,556 personas.

Así mismo de **cinco a catorce años** (5 – 14 años) la población estimada era de 2,052,091 habitantes.

La población de **quince a sesenta y cuatro años** (15 – 64 años) era de 5,503,791 personas.

Del mismo modo, la población de **sesenta y cinco años** (65 y más) al 1 de julio del 2005 era de 510, 521 habitantes.

Nota: Estos datos fueron elaborados por el Consejo Nacional de Población y Familia.- CONAPOFA.

a) Construir una tabla de la población estimada por grupos de edad.

b) Calcule los porcentajes de la población estimada para cada población para cada grupo de edad.

4.9 Durante el período enero-septiembre de 1994 diferentes empresas públicas y privadas realizaron negociaciones con la Bolsa de Valores de Santo Domingo. Los montos (en millones de RD$) de las negociaciones y las empresas aparecen a continuación:

Zona Franca San Isidro, realizó negociaciones por RD$59.7 millones. El Banco Central de la República Dominicana efectuó negociaciones por RD$12.5 millones; asimismo, Mercantil del Caribe negoció con la Bolsa de Valores de Santo Domingo unos RD$12.1 millones. La empresa de telecomunicaciones Tricom, realizó negociaciones por RD$9.4 millones; Helados Bon efectuó negociaciones por RD$9.1 millones; de igual manera la Unión Hotelera realizó negociaciones por RD$8.4 millones y la Urbanizadora Thomén hizo negociaciones con la Bolsa de Valores de Santo Domingo por RD$8.0 millones.

Estas informaciones fueron obtenidas de la Revista Económica del Listín Diario, de la quincena del 30/10/94 al 12/11/94.

a) Construya una tabla del valor de las negociaciones por las empresas señaladas con la Bolsa de Valores de Santo Domingo, período enero-septiembre, 1994.

b) Calcule la participación porcentual de cada empresa con relación al total negociado en el período.

4.10 Según los resultados de la Encuesta Nacional de Hogares de Propósitos Múltiples, realizada por la Oficina Nacional de Estadística en el año 2005, las percepciones de los (as) jefes(as) de hogar sobre los problemas y la gravedad de éstos en su barrio o paraje se presentan a continuación:

El 62.7 % considera que la **pobreza** es un problema muy serio, el 24.6 % entiende que es serio, 9.3 % que no es muy serio, el 1.5 % señala que no existe y el 1.9 % dice que no sabe.

Otro problema señalado es el **desempleo,** considerado por un 56.8 % como muy serio, el 27.8 % cree que es un problema serio, el 9.7 % entiende que no es muy serio, el 1.8 % considera que no existe el desempleo y el 3.9 % no sabe.

Al cuestionar a los (as) jefes de hogar sobre su percepción sobre la **energía eléctrica,** el 46.9 % considera que es un problema muy serio; el 30.7 % señala que es un problema serio, el 18.1 % considera que no es muy serio, el 2.6 % cree que no existe el problema y el 1.7 % no sabe.

La **delincuencia** es percibida por un 26.5 % como un problema muy serio, el 24.4 % dice, que es serio, un 29.5 % lo percibe como no muy serio, el 14.6 % indicó que la delincuencia no existe y el 5 % no sabe.

Un 26.5 % de los (as) jefes de hogares considera como muy serio el problema de las **drogas,** el 18.6 % cree que el problema de las **drogas** es serio, el 19.1 % señala que el problema es muy serio, el 18.1 % señala que no existe y el 17.7% no sabe.

El **consumo de alcohol** como problema social es considerado por un 26.4 % de los encuestados como muy serio, el 28.5 % dice que es serio el problema, el 30.6 % percibe como no muy serio dicho problema, mientras que el 9.0 % señaló que no existe este problema y el 5.5 % no sabe.

Así mismo, **el pandillaje,** como problema del barrio o paraje es considerado por el 12.8 % como muy serio, el 13.2 % entiende que es un problema serio, mientras que el 22.1 % dice que no es muy serio, en el 42.6 % de los parajes o barrios no existe el pandillaje y el 9.2 % no sabe.

La prostitución como problema es calificada por el 12.1 % como muy serio, el 13.9 % cree que es un problema serio, mientras que el 26.2 % considera que no es muy serio, el 34.0 % señaló que dicho problema no existe y el 13.7 % no sabe.

Las riñas callejeras es un problema calificado como muy serio por el 9.7 % de los encuestados, mientras que el 12.4 % creen que es un problema serio. Así mismo, el 30.1 % percibe como no muy serio el problema de las riñas callejeras, el 41.0 % consideran que no existe dicho problema y el 6.9 % no sabe.

4.11 El consumo total de energía eléctrica en megavatios / hora en República Dominicana según sectores, durante los años 2003, 2004 y 2005 se presentan a continuación:

Durante el año 2003, el sector residencial tuvo un consumo de 2,795,298 megavatios/ horas; el sector industrial consumió 2,160,678 megavatios / horas; el sector comercial consumió 556, 838 megavatios/horas y el sector gobierno tuvo un consumo durante el año 2003 de 709,485 megavatios/horas.

Así mismo, durante el año 2004, el consumo de energía eléctrica por el sector residencial fue de 2,312,358 megavatio/horas; el sector industrial tuvo un consumo de 1,655,282 megavatios/horas; el sector comercial consumió 408,969 megavatios/hora. De igual manera, el sector gobierno, tuvo un consumo de 611,781 megavatios/hora.

Durante el año 2005, el consumo de energía eléctrica fue como sigue: el sector residencial consumió 2,390,256 megavatios/horas; el industrial tuvo un consumo de 1,699,047 megavatio/hora.

Así mismo, el sector comercial consumió 442,470 megavatios/hora y el consumo del gobierno durante el año 2005 fue del 690,209 megavatios/horas.

Nota: Estos datos fueron suministrados por el Banco Central de la República Dominicana

* Cifras preliminares.

a) Construir una tabla del consumo de energía eléctrica por años, según sectores.
b) Determine los porcentajes del consumo de energía eléctrica para cada año, por cada sector

4.12 Con los datos del ejercicio 3.1 de la página 67, que se refiere a los datos sobre edad, sexo, estado civil, carrera que cursan y condición laboral, recopilados entre un grupo de estudiantes universitarios. Se pide:

a) Construir una tabla con el estado civil de los estudiantes investigados según el sexo.-

b) Calcule el porcentaje de estudiantes según el estado civil, por sexo.-
c) Qué porcentaje de estudiantes de sexo masculino son casados?

4.13 La ejecución del gasto público social en República Dominicana (en millones de RD$) durante el año 2000, según sectores, se presentan en la siguiente tabla:

Sector	Gastos (Millones de RD$)
Educación	8,522.0
Salud	5,846.0
Asistencia Social	3,222.0
Infraestructura Básica*	3,501.0
Vivienda y otros**	1,439.0
Total	22,530.0

Fuente: Oficina Nacional de Presupuesto

* Incluye: Agua y saneamiento, así como otros servicios comunitarios y municipales.

** Incluye: Vivienda, trabajo, recreación y deportes.

a) Construya una gráfica de barras 100%
b) Construya una gráfica circular

4.14 Con los datos del ejercicio 3.1 de la página 67, que se refiere a los datos sobre edad, sexo, estado civil, carrera que cursan y condición laboral, recopilados entre un grupo de estudiantes universitarios. Se pide:

a) Construir una tabla con los datos relativos a la carrera que cursan los estudiantes según su condición laboral.
b) Determinar el porcentaje de estudiantes según la carrera que cursan por su condición laboral.
c) De los estudiantes que trabajan, que porcentaje son de pedagogía?

4.15 La matrícula por niveles y modalidades del sector educativo público de la Republica Dominicana para el año escolar 2004-2005 se presenta en la siguiente tabla:

Nivel es y modalidades	Matrícula
Inicial	112,523
Básico	1,313,438
Medio	334,780
Técnico Profesional	28,383
Educación de Adultos	85,574
Total	**1,874,698**

Fuente: Secretaría de Estado de Educación

a) Construya una gráfica circular

b) Construya una gráfica de barras 100%

4.16 Con los datos del ejercicio 3.1 de la página 67, que se refiere a los datos sobre edad, sexo, estado civil, carrera que cursan y condición laboral, recopilados entre un grupo de estudiantes universitarios. Se pide:

a) Construir una tabla con los datos relativos a la carrera que cursan los estudiantes según el sexo.

b) Determine el porcentaje de estudiantes según la carrera que cursan por el sexo.

c) En qué carrera existe la mayor participación porcentual de estudiantes del sexo femenino

d) En qué carrera existe la menor participación porcentual de estudiantes del sexo masculino.

4.17 Usando los siguientes rangos de edades de los estudiantes universitarios del ejercicio 3.1 de la página 67, construya una tabla con las edades de éstos según del sexo.

De 18 a menos de 21 años

De 21 a menos de 24 años

De 24 a menos de 27 años

De 27 hasta 31 años

4.18 La tasa de inflación (%) por grupos de bienes y servicios durante el año 2005 se presenta en la siguiente tabla.

Grupos de bienes y servicios	Tasa de Inflación (%)
Alimentos, bebidas y tabaco	0.49
Vestuario y calzados	6.50
Vivienda	9.28
Muebles, accesorios y equipos domésticos	0.46
Salud	6.34
Transporte	24.72
Diversión, entretenimiento y cultura	3.92
Educación	9.28
Hoteles, bares y restaurantes	1.87
Bienes y servicios diversos	2.41

Fuente: Banco Central de la Republica Dominicana

a) Construya una gráfica de barras.

4.19 Las tasas de desocupación (%) de la República Dominicana durante el periodo 1991-2005 se presenta en la siguiente tabla.

Tasa de desocupación 1991-2005

Año	Desocupación
1991	19.6
1992	20.3
1993	19.9
1994	16.0
1995	15.8
1996	16.7
1997	16.0
1998	14.4
1999	13.8
2000	13.9
2001	15.6
2002	16.1
2003	17.0
2004	18.4
2005	17.9

Fuente: Encuesta Nacional de Fuerza de Trabajo (ENFT). Banco Central de la Rep. Dom.

a) Construya una gráfica de barras.

4.20 Las recaudaciones por impuestos sobre el consumo de combustibles durante el período 1996-2004 en la República Dominicana se presentan en la siguiente tabla.

Año	Recaudaciones (mill. de RD$)
1996	3,317.1
1997	5,228.6
1998	6,260.3
1999	4,161.7
2000	2,772.2
2001	7,511.3
2002	8,100.9
2003	8,277.6
2004	11,082.3

Fuente: Banco central de la República Dominicana

a) Construya una gráfica de barras.

b) Construya una gráfica lineal.

4.21 La evolución de los precios del barril de petróleo durante el año 2005 por mes se presenta a continuación

Mes	Precio promedio Mensual * (US$)
Enero	42.9
Febrero	44.6
Marzo	50.9
Abril	50.6
Mayo	47.8
Junio	53.9
Julio	56.4
Agosto	61.9
Septiembre	61.6
Octubre	58.2
Noviembre	55.0
Diciembre	56.4

Fuente: Banco Central de la Republica Dominicana
* Promedio Brent, Dubai y WTI en US$ por barril .

a) Construya una gráfica de barras

b) Construya una gráfica lineal.

4.22 Construya un histograma y un polígono de frecuencias con los datos del ejercicio 5.19 de la página 168 que se refiere a las calificaciones obtenidas por 75 estudiantes universitarios en un curso de matemáticas.

4.23 Indicadores de equipamiento de los hogares de la República Dominicana, por zona, 2005.

(En porcentaje).

Indicador / Hogar	Total	Zona	
		Urbana	Rural
Televisor	75.8	81.4	65.5
Nevera	64.2	72.1	49.5
Radio y/o Equipo. de música	61.5	64.5	56.1
Lavadora de Ropa	58.2	64.5	46.6
Teléfono residencial	25.8	33.6	11.4
Computadora	8.7	11.5	3.5
Aire acondicionado	4.4	5.9	1.6
Motocicleta o passola	20.8	18.6	24.6
Carro privado	11.5	14.4	6.3
Bicicleta	4.6	5.2	3.4
Camioneta	3.9	4.1	3.6
Yipeta	1.6	2.2	0.6

Fuente: Encuesta Nacional de Hogares de Propósitos Múltiples (ENHOGAR -2005); ONE

a) Construya una gráfica de barras con el total de cada indicador.

b) Construya una gráfica lineal que compare los porcentajes por zona de residencia.

c) Construya una gráfica lineal con el total de cada indicador.

4.24 El valor (FOB) de las exportaciones e importaciones nacionales (en millones de US$) de la República Dominicana para el periodo 1990-2002 se presentan a continuación.

Año	Valor exportaciones (millones de US$)	Valor importaciones (millones de US$)
1990	734.5	1,792.8
1991	658.3	1,728.8
1992	563.4	2,174.6
1993	504.0	2,118.4
1994	625.9	2,563.1
1995	872.1	3,164.2
1996	945.5	3,580.7
1997	1,017.3	4,192.0
1998	880.3	4,896.6
1999	805.2	5,206.8
2000	966.1	6,416.0
2001	794.7	5,952.9
2002	847.7	6,237.7
2003	1,064.0	5,095.9
2004	1,250.7	5,368.1
2005*	1,397.9	7.206.6

Fuente: Banco Central de la República Dominica
*cifras preliminares

a) Construya una gráfica lineal que le permita comparar el valor de las exportaciones y el valor de las importaciones.

b) Construya una gráfica de barras compuestas.

4.25 Alumnos de nuevo ingreso en el sistema educativo nacional, por nivel y modalidad, año lectivo 1998-1999.

Nivel y modalidad	Alumnos nuevo ingreso (1998-1999)
Inicial	133,312
Básica	291,978
Adultos	19,165
Media	130,947
Total	**575,402**

Fuente: Secretaría de Estado de Educación.

a) Construya una gráfica de barra 100%.

b) Construya una gráfica circular.

4.26 Alumnos matriculados en el nivel inicial del sistema educativo nacional, por sector, según año lectivo.

Año Lectivo	Público	Privado	Total
1994-95	72,296	71,179	143,475
1995-96	99,330	80,021	179,351
1996-97	112,027	77,058	189,085
1997-98	103,723	86,818	190,541

Fuente: Secretaría de Estado de Educación.

a) Construya una gráfica de barras componentes.

4.27 Estudiantes matriculados al inicio del año escolar en el sistema educativo nacional, según sector y condición del alumno, 1997-1998.

Condición	Público	Privado	Total
Nuevo	223,063	41,911	264,974
Promovido	930,495	202,297	1,132,792
Repitente	65,280	2,846	68,126
Reentrante	22,092	1,534	23,626

Fuente: Secretaría de Estado de Educación.

a) Construya una gráfica de barras componentes.

4.28 Alumnos del sector público de la educación media del sistema educativo nacional, matriculados a inicio de año por sexo, según grado, 1996-1997.

Grado	Masculino	Femenino	Total
Primero	40,250	46,355	86,605
Segundo	29,240	36,778	66,018
Tercero	21,342	28,098	49,440
Cuarto	15,858	21,624	37,482
Total	**106,690**	**132,855**	**239,545**

Fuente: Secretaría de Estado de Educación .

a) Construya una gráfica de barras compuestas.

b) Construya una gráfica de barras componentes.

4.29 Las edades (en años) de las personas que solicitaron préstamos en un banco comercial se presentan a continuación:

Edad (años)	Número de personas
25-30	10
30-35	15
35-40	25
40-45	40
45-50	35
50-55	20
55-60	5

a) Construya un histograma.

b) Construya un polígono de frecuencias.

c) Construya una ojiva.

4.30 La población estimada al 1 de julio del 2006 de las quince provincias mas pobladas de la Republica Dominicana y el resto del país se presenta en la siguiente tabla.

Provincia	Población	%
Santo Domingo	1,979,101	21.4
Distrito Nacional	994,627	10.8
Santiago	937,086	10.1
San Cristóbal	583,062	6.3
La vega	408,778	4.4
Duarte	332,384	3.6
Puerto Plata	322,701	3.5
San Pedro de Macorís	302,271	3.3
Azua	269,577	2.9
San Juan de la Maguana	262,780	2.8
La Romana	246,884	2.7
Espaillat	238,185	2.6
Sánchez Ramírez	211,617	2.3
Barahona	183,242	2.0
Valverde	182,150	2.0
Resto País	1,775,411	19.3
Total	9,229,856	100.0

Fuente: Consejo Nacional de Población Familiar CONAPOFA

a) Con los valores porcentuales, construya una grafica de barras.

4.31 El tiempo (en minutos) requerido por un cajero bancario para despachar los clientes que se presentan a su ventanilla se muestra en la siguiente tabla.

Tiempo (minutos)	Número de clientes
0 a menos de 1.0	3
1.0 a menos de 2.0	8
2.0 a menos de 3.0	12
3.0 a menos de 4.0	5
4.0 a menos de 5.0	2
5.0 a menos de 6.0	1
6.0 a menos de 7.0	1
Total	**32**

a) Construya un polígono de frecuencia.

b) Construya una ojiva.

4.32 Los ingresos anuales (en miles de RD$) de los empleados de una empresa de producción industrial aparecen en la siguiente tabla.

Ingresos anuales (miles de RD$)	Número de empleados
10-20	50
20-30	75
30-40	110
40-50	125
50-60	45
60-70	30
70-80	23
80-90	12
90-100	5
Total	**475**

a) Construya un histograma.

b) Construya un polígono de frecuencia.

c) Construya una ojiva.

4.33 El consumo de energía eléctrica (en megavatios / hora) durante el período 2003-2005, según sectores, se presenta a seguidas:

Sector	2003	2004	2005
Residencial	2,795,298	2,312,358	2,390,256
Industrial	2,160,678	1,655,282	1,699,047
Comercial	556,838	408,969	442,470
Gobierno	709,485	611,781	690,209
Total	**6,222,299**	**4,988,390**	**5,221,982**

Fuente: Banco Central de la República Dominicana

a) Construya una gráfica de barra 100% por sector para el año 2003.

b) Construya una gráfica circular por sector para el año 2004.

c) Construya una gráfica de barras con los datos del año 2005.

4.34 Alumnos matriculados a inicio de año en el nivel básico en República Dominicana, por sexo y grado; 1996-1997.

Grado	Masculino	Femenino	Total
Primero	172,341	145,751	318,092
Segundo	117,765	106,435	224,200
Tercero	94,591	89,607	184,198
Cuarto	77,479	78,641	156,120
Quinto	69,739	71,076	140,815
Sexto	59,171	63,183	122,354
Séptimo	46,731	52,236	98,967
Octavo	40,083	46,369	86,452
Total	**677,900**	**653,298**	**1,331,198**

Fuente: Secretaría de Estado de Educación y Cultura.

a) Construya una gráfica de barras componentes.

b) Construya una gráfica circular de todos los alumnos del nivel básico por sexo.

4.35 Los gastos (en millones de RD$) del Poder Judicial para el período 1991-1999 se presentan en la siguiente tabla.

Año	Gastos (millones de RD$)
1991	35.8
1992	71.5
1993	96.3
1994	97.1
1995	110.6
1996	148.7
1997	303.5
1998	542.1
1999	635.4

Fuente: Contraloría General de la República.

a) Construya una gráfica lineal.

4.36 La siguiente distribución muestra los precios al detalle que se pagan por 168 pantalones Jeans para mujer que se distribuyen en una tienda de la ciudad.

Precios (RD$)	Cantidad de pantalones
150-200	10
200-250	15
250-300	25
300-350	30
350-400	40
400-450	24
450-500	16
500-550	8
Total	**168**

a) Construya un histograma de esta distribución.

b) Construya un polígono de frecuencia con la distribución anterior.

4.37 Las remesas familiares (en US$) que ingresaron a la República Dominicana durante el período, 1995-2005 se presentan a continuación.

Años	Valor remesas (millones de US$)
1995	794.5
1996	914.0
1997	1088.9
1998	1326.0
1999	1518.7
2000	1689.0
2001	1807.8
2002	1959.6
2003	2060.5
2004	2230.2
2005	2410.7

Fuente: Banco Central de la República Dominicana.
Banco Interamericano de Desarrollo.

a) Construya una gráfica de barras
b) Construya una gráfica lineal

4.38 Alumnos matriculados al inicio del año escolar por sexo, según nivel y/o modalidad, 1997-98.

Nivel y/o modalidad	Masculino	Femenino	Total
Inicial	96,252	94,289	190,541
Básica	757,249	735,523	1,492,772
Media	145,560	184,384	184,384
Técnico Profesional	12,768	17,176	29,944
Adultos	56,185	46,169	102,354
Otras modalidades*	3,739	18,037	21,776
Total	1,071,753	1,095,578	2,167,331

Fuente: Secretaría de Estado de Educación.
* Incluye Educación Laboral y Técnico Básica.

a) Construya una gráfica de barras componentes.
b) Construya una gráfica de barras compuestas del nivel de estudio de alumnos según el sexo.

4.39 La evolución del salario promedio semanal (RD$) pagado en las Zonas Francas de República Dominicana a técnicos y obreros, durante el período 1993-2001 se presenta a continuación:

Años	Técnicos	Obreros
1993	648.00	297.00
1994	808.00	497.00
1995	1,096.00	540.00
1996	1,257.00	577.00
1997	1,441.00	634.00
1998	1,557.00	638.00
1999	1,587.00	701.00
2000	1,705.00	725.00
2001	1,833.00	778.00

Fuente: Consejo de Zonas Francas de la República Dominicana.

4.40 Recaudaciones del gobierno central de la República Dominicana a través del cobro del Impuesto a la Transferencia de Bienes Industrializados y Servicios.

Años	Recaudaciones (RD$)
1990	968.8
1991	1,467.6
1992	2,403.9
1993	3,201.5
1994	3,524.5
1995	4,106.2
1996	4,640.3
1997	6,237.6
1998	7,178.4
1999	8,562.7
2000	6,427.92
2001	9,047.42
2002	10,063.22
2003	12,185.66

Fuente: Banco Central de la República Dominicana.

a) Construya una gráfica lineal.
b) Construya una gráfica de barras.

4.41 La tabla que se presenta a continuación representa los indicadores de eficiencia interna del sector público del Sistema Educativo Nacional durante el período 1990-2001.

Año Escolar	Promoción (%)	Repitencia (%)	Deserción
1990-91	62.4	13.8	23.8
1991-92	65.2	14.5	20.3
1992-93	65.7	16.2	18.1
1993-94	64.8	16.0	19.2
1994-95	66.8	15.5	17.7
1995-96	70.6	11.9	17.5
1996-97	78.8	5.5	15.7
1997-98	79.9	5.2	14.9
1998-99	85.5	5.6	8.9
1999-00	86.7	5.7	7.6
2000-01	85.1	6.3	8.6

Fuente: Secretaría de Estado de Educación.

a) Construya una gráfica lineal que le permita comparar los tres indicadores.

4.42 Población de 15 años y más de la República Dominicana según estado civil, Censo Nacional de población y viviendas, 2002.

Estado civil	Cantidad de personas
Soltero	1,798,029
Casado	1,307,678
Divorciado	121,203
Viudo	248,114
Unión Libre	1,843,859
Separado legal	52,777
Separado de Unión Libre	285,765
Total	5,657,425

Fuente: Oficina Nacional de Estadistica.

a) Represente mediante una gráfica circular.

b) Represente mediante una gráfica de barras.

4.43 El estado civil y el sexo de un grupo de estudiantes universitarios consultados se presentan a continuación.

Estado Civil	Masculino	Femenino	Total
Soltero	14	16	30
Casado	4	11	15
Divorciado	1	2	3
Unión libre	1	3	4
Total	20	32	52

a) Construya una grafica de barras compuestas de estado civil de los estudiantes según el sexo.

4.44 Las carreras que cursan los estudiantes universitarios encuestados se presentan en la siguiente tabla.

Carrera	No. de Estudiantes
Pedagogía	13
Contabilidad	10
Mercadeo	9
Administración	8
Ingeniería	7
Derecho	5
Total	52

4.45 La siguiente tabla recoge los datos acerca de las carreras que cursan un grupo de estudiantes universitarios, según su condición laboral.

Carreras	Trabaja	No trabaja	Total
Pedagogía	9	4	13
Contabilidad	4	6	10
Mercadeo	3	6	9
Administración	2	6	8
Ingeniería	1	6	7
Derecho	4	1	5
Total	23	29	52

a) Construya una gráfica de barras componentes de la carrera que cursa según su condición laboral.

CAPITULO

V

Medidas de tendencia central o de posición

5.0 Introducción

Una medida de tendencia central sirve para describir de manera resumida una gran masa de datos cuantitativos. Asimismo, las medidas de tendencia central sirven para comparar un conjunto de datos con otro (con igual unidad de medida).

A las medidas de tendencia central también se les llama promedios, y se definen como aquel valor que se utiliza para representar un conjunto de datos.

Las medidas de tendencia central reciben este nombre porque expresan el valor central de un conjunto de datos, o sea, se encuentran localizadas en el centro o cerca del centro de la distribución de los datos.

Se expresan en las mismas unidades que los valores de la variable.

Las medidas de tendencia central o promedios más usados son:

1.- Media aritmética
2.- Mediana
3.- Moda o modo
4.- Media geométrica
5.- Media armónica

Cada uno de los anteriores promedios tiene sus características particulares, las cuales podrán ser estudiadas en el presente capítulo.

La media aritmética, la mediana y la moda son las tres medidas de tendencia central de mayor uso, mientras que la media geométrica y la media armónica son usadas solamente en casos especiales.

Antes de iniciar el desarrollo de cada una de las medidas de tendencia central señaladas anteriormente, haremos una breve exposición acerca de la notación sumatoria, por considerarlo de alto interés para la comprensión de éste y los demás capítulos.

5.1 Notación sumatoria

Para el cálculo de una gran cantidad de indicadores estadísticos casi siempre es necesario determinar la suma de un conjunto de valores.

Para denotar una suma se utiliza la letra griega $\sum$, que es la letra griega mayúscula S o sigma, como es llamada, e indica suma total (de ahí la expresión sumatoria).

Si la variable x toma los valores 2, 5, 8, 10, 4; entonces la

$$\sum x = 2+5+8+10+4=29$$

Si tenemos un conjunto de valores individuales como $x_1, x_2, x_3, x_4, \ldots, x_n$; entonces la media aritmética es:

$$\bar{x} = \frac{1}{n}(x_1+x_2+x_3+x_4+\ldots+x_n); \text{ donde}$$

La expresión $(x_1+x_2+x_3+x_4+\ldots+x_n)$ se representa matemáticamente como:

$$\sum_{i=1}^{n} x_i$$

La expresión i=1 bajo la notación sumatoria o sigma indica el elemento que va a sumarse (i) y el primer término de la suma (1).

La **n** sobre la notación sumatoria o sigma indica el término final de la suma.

La expresión $\sum_{i=1}^{n} x_i$ se lee como,

"Sumatoria de x_i, desde i = 1 hasta n" e indica el rango de los elementos de x_i a sumarse. Cuando la variable que va a sumarse y el rango de los elementos es claro, la expresión por lo regular se escribe simplemente como $\sum x_i$.

Para mayor comprensión de la notación sumatoria se presentan a continuación sus principales propiedades:

1.- La sumatoria de una constante es igual al producto de la constante por el número de veces que se presenta.

$$\sum_{i=1}^{n} K = nK \qquad 5.1$$

sea, n = 4 y K = 8

$$\sum_{i=1}^{4} K = 8+8+8+8 = 32$$

es lo mismo que:

4(8) =32; ya que n = 4 y K = 8

2.- La sumatoria del producto de una constante por una variable es igual al valor de la constante por la sumatoria de la variable.

$$\sum_{i=1}^{n} Kx_i = K\sum_{i=1}^{n} x_i \qquad 5.2$$

así,

$$\sum_{i=1}^{3} 3x_i = 3x_1 + 3x_2 + 3x_3$$

$$\sum_{i=1}^{3} 3\,x_i = 3(x_1+x_2+x_3)$$

$$\sum_{i=1}^{3} 3x_i = 3\sum_{i=1}^{3} x_i$$

3.- La sumatoria de la suma (o diferencia) de dos variables es igual a la suma (o diferencia) de las sumatorias individuales de las dos variables.

$$\sum_{i=1}^{n}(x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i \qquad 5.3$$

Sean los valores de (x_i) y (y_i)

X_i	Y_i	(x_i+y_i)
4	6	10
3	4	7
5	5	10
8	7	15
20	22	42

$$\sum_{i=1}^{4}(x_i + y_i) = 42$$

$$\sum_{i=1}^{4} x_i + \sum_{i=1}^{4} y_i = 20 + 22 = 42$$

5.2 La media aritmética

Es la medida de tendencia central o promedio más conocida y más ampliamente usada de las que hemos mencionado.

Se define como el cociente que se obtiene al dividir la suma de los valores de la variable entre el total de observaciones.

La media aritmética de un conjunto de datos se designa mediante el símbolo x (que se lee "x barra").

Por definición,

$$\text{la } \bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + \ldots + x_n}{n}$$

donde,

$(x_1, x_2, x_3, x_4, \ldots x_n)$ representan los valores de la variable y, **n** es igual al total de datos u observaciones.

Si aplicamos la notación sumatoria, $\sum$, a los valores de la variable que aparecen entre paréntesis, tendremos, que la media aritmética, x, será:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad 5.4$$

La media aritmética es la medida de tendencia central más confiable, además, presenta más facilidad para posteriores tratamientos estadísticos que la mediana y la moda, ya que está definida algebraicamente.

Asimismo, la media aritmética es la medida de tendencia central que satisface el criterio de los mínimos cuadrados, o sea, la suma de las desviaciones cuadradas de los valores de la variable con respecto de la media aritmética será menor que la suma de las desviaciones cuadradas respecto de un valor cualquiera diferente de la media aritmética.

Esto es, $\sum (x_i-\bar{x})^2$ es menor que $\sum (x_i-Me)^2$, menor que $\sum (x_i-A)^2$, siendo A un valor cualquiera diferente de la media aritmética.

El criterio de los mínimos cuadrados es uno de los más utilizados en procedimientos estadísticos inferenciales.

Los indicadores que satisfacen el criterio de los mínimos cuadrados, como la media aritmética, generalmente son excelentes para fines de inferencia estadística en comparación con aquellos que no lo hacen, como la mediana y la moda.

5.2.1 La media aritmética para datos no agrupados

La forma más sencilla para calcular la media aritmética es la usada cuando disponemos de pocos valores, que no ameritan ser organizados en una distribución de frecuencias.

A estos datos, por su forma de presentación le llamamos datos no agrupados y usaremos para el cálculo de la media aritmética la fórmula 5.4.

Ejemplo 5.1

Los pesos (en libras) de cinco niños recién nacidos en una hora en un centro de maternidad fueron los siguientes:

$x_i = 6, 8, 7, 10, 4$

Determine la media aritmética del peso de los recién nacidos.

Solución

Como $\bar{x} = \dfrac{\sum x_i}{n}$ 5.5

$\sum x_i = 6+8+7+10+4 = 35$ (suma de los pesos)

y $n = 5$ (cantidad de niños)

entonces

$$\bar{x} = \frac{35}{5} = 7 \text{ libras}$$

Esto significa que la media aritmética del peso de los cinco niños es de siete libras.

Ejemplo 5.2

Calcule la media aritmética de los salarios semanales (en RD$) devengados por ocho obreros de zona franca.

Los salarios devengados fueron los siguientes:

$x_i = 900, 1075, 1100, 1200, 850, 875, 925, 800$

Solución

Como sabemos,

$$\bar{x} = \frac{\sum x_i}{n}$$

de donde,

$$\sum x_i = 900+1075+1100+1200+850+875+925+800 = 7725$$

y $n = 8$

entonces,

$$\bar{x} = \frac{7725}{8} = \text{RD\$ } 965.62$$

La media aritmética de los salarios semanales de los ocho obreros es de RD $965.62.

5.2.2 La media aritmética para datos agrupados

Cuando la media aritmética se calcula para una distribución de frecuencias o datos agrupados utilizo el punto medio de clase como forma de aproximar a un solo valor los datos contenidos en cada clase.

El punto medio se representa mediante la simbología xi, donde "x" representa la variable considerada en cada clase y el subindice "i" indica el orden de cada una de las clases de la distribución de frecuencias.

La media aritmética para datos agrupados se presenta así,

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 + \ldots + x_n f_n}{f_1 + f_2 + f_3 + \ldots + f_n}$$

Si aplicamos la notación sumatoria,

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i}$$

donde $\sum f_i = n$

así,

$$\bar{x} = \frac{\sum x_i f_i}{n}$$

Como se puede observar en la fórmula 5.5 el procedimiento para el cálculo de la media aritmética para datos agrupados consiste:

1.- Obtener cada punto medio de clase x_i

2.- Multiplicar cada punto medio (x_i) por su correspondiente frecuencia (f_i), ($x_i f_i$)

3.- Sumar los productos anteriores

$\sum(x_i f_i)$

4.- Obtener la media aritmética, al dividir la suma del producto de los puntos medios por sus correspondientes frecuencias, entre el número total de observaciones ($\sum f_i = n$)

$$\bar{x} = \frac{\sum x_i f_i}{n}$$

Ejemplo 5.3

Calcule la media aritmética de las estaturas (pulgadas) de sesenta estudiantes universitarios.

Tabla 5.1 Estatura de sesenta estudiantes universitarios

Estaturas (pulgadas)	N° de Est. (f_i)	X_i	$x_i f_i$
60-62	4	61	244
62-64	8	63	504
64-66	10	65	650
66-68	12	67	804
68-70	15	69	1035
70-72	8	71	568
72-74	3	73	219
Total	60		4024

Solución:

como $\sum x_i f_i = 4024$

y $n = 60$

entonces, $\bar{x} = \frac{\sum xifi}{n} = \frac{4024}{60} = 67.06$ pulgadas

$\bar{x} = 67.06$ pulgadas

Ejemplo 5.4

El tiempo de espera (minutos) de los clientes que se presentaron a realizar alguna transacción a un banco comercial de Santo Domingo se presenta en la siguiente tabla.

Calcule la media aritmética del tiempo de espera.

Tabla 5.2 Tiempo de espera de clientes bancarios (minutos)

Tiempo (minutos)	N° de clientes (f_i)	x_i	$x_i f_i$
0-5	8	2.5	20
5-10	12	7.5	90
10-15	25	12.5	312.5
15-20	20	17.5	350
20-25	10	22.5	225
25-30	5	27.5	137.5
Total	80		1135

Solución:

Como sabemos:

$$\bar{x} = \frac{\sum x_i f_i}{n}$$

Tenemos que:

$$\sum x_i f_i = 1135$$

y $n = 80$

entonces,

$$\bar{x} = \frac{1135}{80} = 14.18 \text{ minutos}$$

Como se puede observar, el tiempo promedio (media aritmética) de espera, para los clientes que visitaron el banco a hacer sus transacciones fue de 14.18 minutos.

5.2.3 La media aritmética ponderada

Cuando al calcular la media aritmética se considera a cada uno de los valores de acuerdo con su importancia en el grupo se le llama media aritmética ponderada.

La media aritmética ponderada la identificamos como, xw.

$$\bar{x}_w = \frac{x_1w_1 + x_2w_2 + x_3w_3 + \ldots + x_nw_n}{w_1 + w_2 + w_3 + \ldots + w_n}$$

Si aplicamos la notación sumatoria, tendremos:

$$\bar{x}_w = \frac{\sum w_i x_i}{\sum w_i}$$

donde; w = factor de ponderación
y x_i = cada uno de los valores de la variable

Ejemplo 5.5

Un estudiante universitario obtuvo las siguientes calificaciones durante un semestre académico. Calcule la media aritmética de las calificaciones para ese semestre.

Tabla 5.3 Calificaciones de un estudiante por asignaturas.

Asignaturas	Calificaciones (x_i)	Créditos (w_i)	$w_i x_i$
Matemáticas	75	4	300
Física	80	4	320
Letras	85	3	255
Biología	75	4	300
Historia	80	3	240
Total		**18**	**1415**

Solución:

Sabemos que,

$$\bar{x}_w = \frac{\sum w_i x_i}{\sum w_i}$$

donde,

$$\sum w_i x_i = 1415$$

y $$\sum w_i = 18$$

entonces,

$$\bar{x}_w = \frac{1415}{18} = 78.61 \text{ puntos.}$$

El resultado anterior indica que ese estudiante universitario obtuvo un índice académico de 78.61 puntos durante ese semestre.

Ejemplo 5.6

La siguiente tabla muestra el porcentaje de cuentas incobrables manejadas durante un año por una compañía de cobros, según empresas acreedoras.

Calcule la media aritmética de las cuentas incobrables.

Tabla 5.4 Porcentaje de cuentas incobrables manejadas durante un año por una compañía.

Empresas acreedoras	Porcentaje Ctas. incobrables (x_i)	N° de Ctas. manejadas (w_i)	$w_i x_i$
A	2.5	6,000	15,000
B	1.8	7,500	13,500
C	3.2	7,000	22,400
Total		**20,500**	**50,900**

Solución:

Como sabemos,

$$\bar{X}_w = \frac{\sum w_i x_i}{\sum w_i}$$

donde,

$$\sum w_i x_i = 50,900$$

y $\sum w_i = 20,500$

entonces,

$$\bar{X}_w = \frac{50,900}{20,500} = 2.483\%$$

En promedio el 2.483% de las cuentas manejadas por la compañía de cobros durante un año fueron dejadas de pagar por los deudores.

5.2.4 Características de la media aritmética

Tal y como hemos observado, la media aritmética es una medida simple y muy conocida, por eso su gran uso en la descripción de un conjunto de datos.

La media aritmética de una distribución de frecuencias o de una serie de datos no agrupados tiene las siguientes características:

1. Es una medida totalmente numérica, o sea, sólo puede ser calculada en datos de características cuantitativas.
2. En su cálculo se toman en cuenta todos los valores de la variable.
3. Es lógica desde el punto de vista algebraico.

$$\bar{x} = \frac{\sum x_i}{n}$$

si despejamos $\sum x_i$ de la ecuación anterior

tendremos que $\sum x_i = n\bar{x}$

Esta condición siempre va a ser real, si tomamos en cuenta que en su cálculo intervienen todos los valores de la variable.

Ejemplo 5.7

Sea $x_i = 2, 4, 5, 6, 3$

$$\bar{x} = \frac{20}{5} = 4$$

así, $n\bar{x} = \sum xi$

$5(4) = 20$

4. La media aritmética es altamente afectada por valores extremos.

Ejemplo 5.8

Sea $x_i = 5, 4, 700, 10, 7, 3$

valor extremo

La $\bar{x} = \frac{729}{6} = 121.5$

Como se puede observar, la media aritmética ($\bar{x}$=121.5) es un valor que no se corresponde con la realidad, ya que esta serie de datos ha sido afectada sensiblemente por un valor que no es representativo de los demás (700), o sea, tenemos un valor muy grande dentro de varios valores pequeños.

En este caso es recomendable no usar la media aritmética como medida de tendencia central, ya que presenta una gran distorsión en la representatividad, sino, hacer uso de otras medidas que veremos más adelante.

5. La media aritmética no puede ser calculada en distribuciones de frecuencias que tengan clases abiertas.

Ejemplo 5.9

La siguiente tabla muestra las edades (años) de un grupo de personas.

Tabla 5.5 Edades de un grupo de personas.

Edad (años)	N° de personas
10-15	2
15-20	5
20-25	7
25-30	10
30-35	8
35-40	6
40 y más	5
Total	**43**

La clase "40 y más" es abierta, ya que carece del límite superior; esto imposibilita el cálculo del punto medio de esta clase y por consiguiente el cálculo de la media aritmética.

Una distribución puede ser abierta en la primera clase, "menos de", cuando carece del límite inferior de la clase.

Existen otras medidas de tendencia central que pueden ser usadas en casos como éste para hacer una razonable descripción de los datos.

6. La media aritmética es única, o sea, un conjunto de datos numéricos tiene una y sólo una media aritmética.

5.2.5 Propiedades matemáticas de la media aritmética

Primera propiedad

La suma de las desviaciones individuales de los valores de la variable con relación a la media aritmética es igual a cero.

$$\sum(x_i - \bar{x}) = 0 \qquad 5.7$$

Demostración

Tenemos que,

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

Si aplicamos la notación sumatoria a la expresión 5.8, tenemos:

$\sum x_i - \sum \bar{x} = 0$

Según vimos en la página 103, la $\sum k = nk$,

o sea, que $\sum \bar{x} = n\bar{x}$

sustituyendo por su igual tenemos:

$$\sum x_i - n\bar{x} = 0 \qquad 5.8$$

Recordando que, $\bar{x} = \frac{\sum x_i}{n}$

donde, $n\bar{x} = \sum x_i$

Al sustituir por su igual en 5.8 tenemos que,

$\sum x_i - \sum x_i = 0$

La suma de dos valores iguales con signos diferentes es igual a cero, lo que comprueba la primera propiedad.

Ejemplo 5.10

Sea $x_i = 2, 3, 5, 6$

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

Solución.

Tabla 5.6 Cálculos

x_i	$(x_i - \bar{x})$
2	(2-4) = -2
3	(3-4) = -1
5	(5-4) = 1
6	(6-4) = 2
16	**-3+3 = 0**

$$\bar{x} = \frac{\sum x_i}{4} = \frac{16}{4} = 4$$

Al sumar (-3+3) se comprueba que $\sum(x_i - \bar{x}) = 0$

Segunda propiedad

Si una constante, k, se le suma o se le resta a cada valor de un conjunto de datos cuya media aritmética es $\bar{x}$, los valores resultantes tendrían una media aritmética igual a $\bar{x} \pm k$.

Demostración

Sabemos que,

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

$$\frac{1}{n}\sum_{i=1}^{n}(x_i \pm k = \frac{1}{n}\sum_{i=1}^{n} x_i \pm \frac{1}{n}\sum_{i=1}^{n} k$$

según las propiedades de la notación sumatoria, tenemos que:

$$\frac{1}{n}\sum_{i=1}^{n} x_i \pm \left(\frac{1}{n}\right) nk = \bar{x} \pm k \qquad 5.9$$

Sustituyendo por su igual y simplificando términos iguales, tenemos que, al sumar o restar una constante a cada valor de un conjunto de datos cuya media es $\bar{x}$, los valores resultantes serían igual a la media aritmética más o menos la constante.

Ejemplo 5.11

Sea $x_i = 2,3,5,6$

Compruebe que,

$$\frac{1}{n}\sum_{i=1}^{n}(xi+k) = \bar{x} + k$$

Solución

Como sabemos, la media aritmética de los datos anteriores es igual a $\bar{x} = 4$

Sea, $k = 3$

Así, los nuevos valores del conjunto de datos serán:

$(x_i + 3) = 5, 6, 8, 9$

De donde,

$$\bar{x} = \frac{28}{4} = 7$$

Así, queda comprobado que la media aritmética de los valores a los cuales le hemos sumado la constante, k=3, es igual a, $\bar{x} = 7$, o sea, la media de los datos originales, $\bar{x} = 4$, más la constante, $k = 3$.

Tercera propiedad

Si una constante, k, se multiplica o divide por cada valor de un conjunto de datos cuya media aritmética es $\bar{x}$, la nueva media del conjunto de datos resultará multiplicada o dividida por dicha constante.

Demostración

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

entonces,

$$\frac{1}{n}\sum_{i=1}^{n}(x_i k) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} k\right)$$

de donde,

$$\frac{1}{n}\sum_{i=1}^{n}(x_i k) = \bar{x}\left(\frac{1}{n}nk\right)$$

Simplificando en el segundo término,

$$\frac{1}{n}\sum_{i=1}^{n}(x_i k) = k\bar{x} \qquad 5.10$$

Ejemplo 5.12

Sea $x_i = 2, 3, 5, 6$

Compruebe que:

$$\frac{1}{n}\sum_{i=1}^{n}(kx_i) = k\bar{x}$$

Solución

La media aritmética de los datos anteriores es, $\bar{x} = 4$.

Sea $k = 5$

Así, los nuevos valores del conjunto de datos serán,

$(kx_i) = 10, 15, 25, 30$

de donde,

$$\frac{1}{n}\sum_{i=1}^{n}(kx_i) = k\bar{x} \quad ; \quad \sum_{i=1}^{n}(kx_i) = 80$$

Sustituyendo en 5.10 e igualando, tenemos que:

$$\frac{1}{4}(80) = 5(4)$$

$$20 = 20$$

Así, se comprueba la propiedad que establece que si una constante se multiplica por los valores de un conjunto de datos cuya media aritmética es $\bar{x}$, la media aritmética del nuevo conjunto de datos resultará multiplicada por dicha constante.

Cuarta propiedad

La suma de las desviaciones cuadráticas de los valores de la variable con relación a la media aritmética es menor que la suma de las desviaciones cuadráticas respecto de un valor cualquiera, diferente de la media aritmética.

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 < \sum_{i=1}^{n}(x_i-A)^2 \;\forall A \neq \bar{x}$$

La expresión anterior se lee así:

La sumatoria desde i=1 hasta n de las desviaciones cuadráticas de los valores de la variable, xi, con relación a la media aritmética, $\bar{x}$, es menor que la sumatoria de las

desviaciones cuadráticas de los valores de la variable, x_i, con relación a un valor cualquiera, A, para todo valor de A diferente de la media aritmética.

Demostración

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 < \sum_{i=1}^{n}(x_i-A)^2 \; \forall A \neq \bar{x}$$

Sea $A = (\bar{x} \pm k)$

Sustituyendo el valor de A por su igual,

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 < \sum_{i=1}^{n}(x_i-(\bar{x}+k))^2$$

reagrupando términos, tenemos que,

$$\sum_{i=1}^{n}(x_i-k)^2 < \sum_{i=1}^{n}[(x_i-\bar{x})-k]^2$$

desarrollando el binomio de la derecha,

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 < \sum_{i=1}^{n}[(x_i-\bar{x})^2 - 2k(x_i-\bar{x}) + k^2]$$

aplicando notación sumatoria a la expresión de la derecha,

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 < \sum_{i=1}^{n}(x_i-\bar{x})^2 - 2k\sum(x_i-\bar{x}) + nk^2$$

como sabemos $\sum(x_i-\bar{x}) = 0$, entonces,

$$2k\sum(x_i-\bar{x}) = 0$$

Por tanto,

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 < \sum_{i=1}^{n}(x_i-\bar{x})^2 + nk^2$$

La expresión de la derecha siempre será mayor que la de la izquierda porque nk^2 siempre serán positivos.

Ejemplo 5.13

Sea $x_i = 2, 3, 5, 6$

Comprobar que:

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 < \sum_{i=1}^{n}(x_i-A)^2 \; \forall A \neq \bar{x}$$

Solución

Tabla 5.7 Cálculos

x_i	$(x_i-\bar{x})$	$(x_i-\bar{x})^2$	(x_i-A)	$(x_i-A)^2$
2	(2-4)=-2	4	(2-3) =-1	1
3	(3-4)=-1	1	(3-3) = 0	0
5	(5-4)= 1	1	(5-3) = 2	4
6	(6-4)= 2	4	(6-3) = 3	9
Total **16**	**0**	**10**		**14**

Sea A = 3

Como se puede observar en la tabla,

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 = 10 \qquad y \qquad \sum_{i=1}^{n}(x_i-A)^2 = 14$$

así queda comprobado que,

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 < \sum_{i=1}^{n}(x_i-A)^2, \text{ o sea, } 10 < 14$$

5.3 La mediana

Como señalamos anteriormente, la media aritmética es la medida de tendencia central de un conjunto numérico de datos más utilizada y más conocida, no siendo así, la única medida que se puede utilizar para indicar la tendencia central de un conjunto de datos. Otra de las medidas de tendencia central de gran uso es la mediana, de la cual expondremos a continuación.

La mediana es una medida de tendencia central que divide la distribución en dos partes iguales, de forma tal que la mitad de los datos son mayores que la mediana y la otra mitad es menor que ella.

La mediana es igual al percentil de orden cincuenta y al cuartil de orden dos.

La voy a identificar como, Me (mediana)

$$Me = \frac{n+1}{2} \quad \text{(posición de la mediana)}$$

5.3.1 Mediana para datos no agrupados

Para calcular la mediana cuando disponemos de un conjunto de datos no organizados en una distribución de frecuencias debemos tener presente el siguiente procedimiento.

1. Organizar los valores en orden ascendente o descendente.
2. Encontrar la posición de la mediana mediante la siguiente fórmula,

$$Me = \frac{n+1}{2} \qquad 5.12$$

3. Si el número total de valores n es impar, la posición de la mediana será un número entero, y por consiguiente, la mediana la encontramos de manera directa.
4. Si el número total de valores n es par, la posición de la mediana será un número fraccionario, y por consiguiente, la mediana la encontramos mediante interpolación.

Ejemplo 5.14

Calcule la mediana de la siguiente serie de valores.

$x_i = 5, 8, 10, 7, 4, 13, 16$

Solución:

Procederemos a organizar los valores en orden ascendente.

Sea, $x_i = 4, 5, 7, 8, 10, 13, 16$

Por tanto,

$$Me = \frac{n+1}{2} = \frac{7+1}{2} = 4 \text{ (posición de la mediana)}$$

Así, revisando el orden de los valores, se observa que la posición número cuatro corresponde al valor igual a ocho,

$Me = 8$

Este valor, 8, divide el conjunto de datos en dos partes iguales, o sea, la mitad de los datos son mayores que la mediana y la otra mitad son menores que ella.

$$x_i = \underbrace{4,\ 5,\ 7,}_{\text{valores menores que la mediana}}\ \underbrace{8,}_{\text{mediana}}\ \underbrace{10,\ 13,\ 16}_{\text{valores mayores que la mediana}}$$

Cuando el número de valores es par, veamos lo que ocurre,

Ejemplo 5.15

$x_i = 50, 95, 89, 78, 65, 58, 75, 80, 100, 110$

Solución

A continuación organizamos los valores dados en forma ascendente.

Sea $x_i = 50, 58, 65, 75, 78, 80, 89, 95, 100, 110$

Por lo tanto,

$$Me = \frac{n+1}{2} = \frac{10+1}{2} = 5.5 \text{ (posición de la mediana)}$$

Así, la mediana se encuentra entre las posiciones número cinco y número seis. A la posición número cinco corresponde el valor setenta y ocho y a la posición número seis corresponde el valor ochenta. De esta manera, a la posición 5.5 le corresponde el valor comprendido entre 78 y 80.

Así, la mediana,

$$Me = \frac{78+80}{2} = 79$$

Veamos gráficamente el cálculo anterior,

$x_i =$ 50, 58, 65, 75, 78 ↑ 80, 89, 95, 100, 110

Valores menores que la Me. — Mediana = 79 (posición 5.5) — Valores mayores que la Me.

5.3.2 La mediana para datos agrupados

Para calcular la mediana cuando los datos son agrupados en una distribución de frecuencias es necesario, en primer lugar, determinar la clase que contiene el valor de la mediana.

El procedimiento para establecer cuál es la clase que contiene el valor de la mediana se efectúa de la siguiente forma:

1. Se toma la mitad del total de las observaciones (n/2).
2. Localizar aquella clase que contiene la frecuencia acumulada que iguala o excede la mitad del total de observaciones.
3. Al identificar esta clase, determinar el valor de la mediana, a través de la siguiente fórmula:

$$Me = Li + \left[\frac{n/2 - F_{i-1}}{fi}\right] Ci \qquad 5.13$$

donde,

Li = Límite inferior de la clase que contiene la mediana.

n = Número total de observaciones en la distribución de frecuencias.

F_{i-1} = Frecuencia acumulada anterior a la frecuencia acumulada que contiene la clase mediana.

fi = Frecuencia simple de la clase que contiene la mediana.

Ci = Intervalo de la clase que contiene la mediana.

Ejemplo 5.16

Calcule la mediana de las edades (años) de los estudiantes de un curso de estadística que se presentan en la siguiente tabla.

Tabla 5.8 Edades de los estudiantes de un curso de estadística.

Edad (años)	N° de estudiantes (fi)	Fi=Frec. Acumulada
18-20	10	10
20-22	12	10+12 = 22
22-24	18	22+18 = 40
24-26	13	40+13 = 53
26-28	9	53+ 9 = 62
28-30	8	62+ 8 = 70
Total	**70**	

Solución

$$\text{Sea la } Me = Li + \left[\frac{n/2 - Fi\text{-}1}{fi}\right] Ci$$

Al calcular la mitad del total de las observaciones tendremos,

$n/2 = 70/2 = 35$

Con el resultado obtenido (n/2=35) localizamos aquella clase que contiene la frecuencia acumulada que iguala o excede a dicho valor.

De esta manera, la clase que contiene la mediana es aquella donde la frecuencia acumulada es igual a cuarenta (Fi=40), o sea, la clase de 22 a menos de 24 años.

donde:

$n/2 = 35$ años
Li = 22 años
Fi--1 = 22 estudiantes
fi = 18 estudiantes
Ci = 2 años

Sustituyendo en 5.13 tenemos,

$$Me = 22 + [\frac{35 - 22}{18}] 2$$

así,

$$Me = 22 + [\frac{13}{18}] 2$$

de donde,

Me = 23.44 años

Como se puede observar, la mitad de los estudiantes del curso de estadística tienen 23.44 años o menos de edad y la otra mitad tienen 23.44 años o más de edad.

Ejemplo 5.17

Calcule la mediana de las ventas diarias (en miles de RD$) de un supermercado durante un mes.

Tabla 5.9 Ventas diarias (en miles de RD$) de un supermercado durante un mes.

Ventas (miles de RD$)	N° de días (fi)	Frecuencia Acumulada (Fi)
150-200	1	1
200-250	5	1+5= 6
250-300	6	6+6=12
300-350	9	12+9=21
350-400	4	21+4=25
400-450	3	25+3=28
450-500	2	28+2=30
Total	**30**	

Solución

Sea la $Me = Li + [\frac{n/2 - Fi-1}{fi}] Ci$

Al calcular la mitad de las observaciones tendremos,

n/2 = 30/2 = 15

Con el resultado obtenido (n/2=15) localizamos aquella clase que contiene la frecuencia acumulada que es igual o excede a dicho valor.

Así, la clase que contiene la mediana es aquella cuya frecuencia acumulada es igual a veintiuno (Fi=21), o sea, la clase de 300 mil a menos de 350 mil pesos.

donde,

n/2 = 15 días

Li = 300 mil (RD$)

Fi-1 = 12 días

fi = 9 días

Ci = 50 mil (RD$)

Sustituyendo en 5.13 tenemos que,

$$Me = 300 + [\frac{15-12}{9}] 50$$

Así,

$$Me = 300 + [\frac{3}{9}] 50$$

de donde,

Me = 316.67 (miles de RD$).

Este supermercado registró ventas diarias, promedio mediano, de RD$ 316,670.00.

5.3.3 Características de la mediana

Las principales características de la mediana son:

1. En su cálculo no se incluyen todos los valores de la variable.
2. La mediana no es afectada por valores extremos.

   Por ejemplo, la mediana de los valores, 10, 12, 15, 18, 20 es igual a 15.

   Asimismo, la mediana de los valores, 10, 12, 15, 18, 2000, sigue siendo igual a 15.

3. La mediana puede ser calculada en distribuciones de frecuencias con clases abiertas.

Por ejemplo, en la siguiente tabla se puede calcular la mediana.

Tabla 5.10. Estaturas (en pulgadas) de varias personas.

Estaturas (pulgadas)	N° de personas (fi)
Menos de 60	2
60 - 62	5
62 - 64	8
64 - 66	12
66 - 68	9
68 - 70	7
70 - 72	4
Total	**47**

4. No es lógica desde el punto de vista algebraico, como lo es la media aritmética.

5.4 La moda

Entre las medidas de tendencia central que hemos visto figuran la media aritmética y la mediana, que describen y resumen con alta confiabilidad una distribución de frecuencias.

Asimismo, la moda es otra medida de tendencia central de gran importancia en la descripción y resumen de una distribución de frecuencias.

En la distribución de un conjunto de datos, aquel valor que más se repite se le define como el valor modal; o sea que, la moda es el valor que más se repite o que ocurre con mayor frecuencia en una distribución o conjunto numérico de datos. Es el valor con más alta probabilidad de ser seleccionado de un conjunto de datos, cuando el valor es seleccionado al azar.

La moda presenta dos grandes ventajas, y consisten en que puede ser calculada tanto para datos de características cuantitativas como para datos de características cualitativas; asimismo, cuando se tiene un conjunto de datos poco

numerosos la moda se obtiene por observación, no se requiere de ningún tipo de cálculo.

La moda la vamos a identificar como:

Mo = moda

Cuando una distribución de datos tiene un solo valor que es el que más se repite, a esta distribución le vamos a llamar unimodal o monomodal. Cuando en la distribución existen dos modas, le llamaremos bimodal. Si la distribución tiene más de dos modas, ésta recibe el nombre de plurimodal.

Asimismo, cuando una distribución no tiene ningún valor que se repita, a esa distribución la vamos a llamar amodal.

5.4.1 La moda para datos no agrupados

Cuando se dispone de una cantidad reducida de valores, la moda puede ser obtenida por simple observación; escogiendo aquel valor que más se repita o que ocurra con mayor frecuencia.

Ejemplo 5.18

Determine la moda del peso (libras) de ocho niños al momento de nacer.

$x_i = 5, 8, 7, 6, 7, 5, 10, 7.$

Solución

Como podemos notar por simple observación la moda del peso de estos niños al momento de nacer es igual a 7 libras.

O sea, Mo = 7 libras.

Ejemplo 5.19

Calcule la moda de los siguientes valores

$x_i = 10, 12, 15, 10, 11, 15, 13, 14, 16$

Solución:

En este caso, existen dos modas, que son, 10 y 15.

O sea, Mo = 10 y 15.

5.4.2 La moda para datos agrupados

El procedimiento para el cálculo de la moda para un conjunto de datos agrupados en una distribución de frecuencias consiste en:

1. Identificar la clase que contiene la moda.
   Este proceso es sencillo, ya que se obtiene por simple observación la clase que tiene la frecuencia más elevada o donde el número de observaciones es mayor.
2. Identificada esta clase se determina el valor de la moda mediante la fórmula que se ilustra a continuación.

$$Mo = Li + \left[\frac{\Delta_1}{\Delta_1 + \Delta_2}\right] Ci \qquad 5.14$$

donde:

Li = Límite inferior de la clase que contiene la moda.

$\Delta_1 = fm - fm\text{-}1$

$\Delta_2 = fm - fm+1$

fm = frecuencia simple de la clase que contiene la moda.

fm-1 = frecuencia simple, anterior a la de la clase que contiene la moda.

fm+1 = frecuencia simple, posterior a la de la clase que contiene la moda.

Ci = intervalo de la clase que contiene la moda.

Ejemplo 5.20

Calcule la moda de las estaturas (pulgadas) de un grupo de personas.

Tabla 5.11. Estaturas (pulgadas) de varias personas

Estaturas (pulgadas)	N° de personas (fi)
56-58	2
58-60	5
60-62	8
62-64	10
64-66	16
66-68	14
68-70	7
70-72	3
Total	**65**

Solución

La clase que contiene la moda es aquella que comprende las estaturas de 64-66 pulgadas, cuya frecuencia simple es 16.

Sea,

$$Mo = Li + \left[\frac{\Delta_1}{\Delta_1 + \Delta_2}\right] Ci \qquad 5.16$$

donde,

$Li = 64 ; \Delta_1 = 16\text{-}10 = 6 ; \Delta_2 = 16\text{-}14 = 2 ; Ci = 2$

Sustituyendo en 5.16, tendremos que:

$$Mo = 64 + \left[\frac{6}{6+2}\right] 2 = 65.5 \text{ pulgadas}$$

Mo = 65.5 pulgadas es la estatura que mas se repite entre las personas.

Ejemplo 5.21

Las calificaciones de los estudiantes de historia se presentan en la tabla 5.12. Calcule la moda de las calificaciones.

Tabla 5.12 Calificaciones de los estudiantes de historia

Calificaciones	No. de estudiantes (fi)
40-50	6
50-60	10
60-70	20
70-80	25
80-90	12
90-100	7
Total	**80**

Solución

La clase que contiene la moda es aquella donde 25 estudiantes obtuvieron calificaciones entre 70 y 80 puntos.

Así,

$Li = 70$; $\Delta_1 = 25-20 = 5$; $\Delta_2 = 25-12 = 13$; $Ci = 10$

Sustituyendo en 5.16 tendremos que:

$$Mo = 70 + [\frac{5}{5+13}]10$$

Mo = 72.77 puntos fue la calificación mas frecuente entre los estudiantes de historia.

5.4.3 Características de la moda

Las principales características de la moda son:

1. En su cálculo no se incluyen todos los valores de la variable.
2. El valor de la moda puede ser afectado grandemente por el método de designación de los intervalos de clase. Cuando un mismo conjunto de valores son agrupados con diferentes tipos de intervalos, el valor de la moda va a variar para cada distribución.
3. No está definida algebraicamente como está definida la media aritmética.
4. La moda puede ser calculada en distribuciones con clases abiertas.
5. No es afectada por valores extremos.

5.5 La media geométrica

Entre las medidas de tendencia central o medidas de posición hemos señalado como las tres principales la media aritmética, la mediana y la moda; no obstante existen otras medidas que permiten describir un conjunto de datos, aunque con un uso más limitado.

Una de estas medidas es la media geométrica, que se define como la raíz n-ésima del producto de los valores de la variable.

Si hay tres valores, la media geométrica será la raíz cúbica del producto de esos valores.

La media geométrica es frecuentemente usada para promediar tasas de cambio, razones y valores que muestren una progresión geométrica.

Se identifica con la letra G (que se lee como media geométrica).

5.5.1 La media geométrica para datos no agrupados.

$$G = \sqrt[n]{\text{Producto de los valores de la variable}}$$

Sean los valores de la variable representados simbólicamente por $x_1, x_2, x_3, x_4, \ldots, x_n$.

entonces,

$$G = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot x_4 \ldots x_n} \qquad 5.15$$

Ejemplo 5.22

Calcule la media geométrica para los siguientes valores,

$x_i = 2, 5, 7, 4, 9$

Solución

Sea $n = 5$

si aplicamos la fórmula 5.15 tendremos:

de donde,

G = Antilogaritmo de 0.0688.

Así,

G = 1.171

Este resultado indica que la media geométrica de las razones o del crecimiento del gasto del gobierno central durante el período 1992-1999 fue de 1.171 ó 117.1%.

5.5.2 La media geométrica para datos agrupados

Cuando se trata de un conjunto de datos agrupados en una distribución de frecuencias la fórmula para calcular la media geométrica puede ser escrita de la siguiente forma:

$$\text{Log } G = \frac{f_1 \log x_1 + f_2 \log x_2 + f_3 \log x_3 + \ldots\ldots + f_n \log x_n}{f_1 + f_2 + f_3 + \ldots\ldots + f_n}$$

Sabemos que,

$$f_1 + f_2 + f_3 + \ldots\ldots + f_n = \sum_{i=1}^{n} f_i = n$$

Entonces, la fórmula anterior puede resumirse como,

$$\text{Log } G = \frac{1}{n} \sum_{i=1}^{n} (f_i \log x_i) \qquad 5.17$$

Ejemplo 5.25

Las temperaturas (grados celsius) registradas en diferentes localidades del país durante un mes fueron las siguientes.

Calcule la media geométrica de las temperaturas.

Tabla 5.14. Temperaturas (grados celsius) de localidades del país.

Temperaturas	No. de localidades (f_i)	x_i	Log x_i	$f_i \log x_i$
18-20	2	19	1.278	2.556
20-22	5	21	1.322	6.610
22-24	8	23	1.361	10.888
24-26	9	25	1.398	12.582
26-28	15	27	1.431	21.465
28-30	13	29	1.462	19.006
30-32	9	31	1.491	13.419
Total	**61**			**86.526**

Solución

Aplicando los resultados parciales de la tabla 5.14 a la fórmula 5.17 tendremos que,

$$\text{Log } G = \frac{1}{n} \sum_{i=1}^{n} (fi \log xi)$$

Así,

$$\text{Log } G = \frac{1}{61} (86.526) = 1.4184$$

Entonces,

G = Antilogaritmo de 1.4184

de donde,

G = 26.21 grados celsius

Ejemplo 5.26

Calcule la media geométrica de las edades de 30 personas tomadas al azar.

Tabla 5.15 Edades (en años) de 30 personas.

Edad	N° de personas (f_i)	x_i	$\log x_i$	$f_i \log x_i$
20-24	4	22	1.342	5.368
24-28	6	26	1.415	8.490
28-32	10	30	1.477	14.770
32-36	7	34	1.531	10.717
36-40	3	38	1.579	4.737
Total	**30**			**44.082**

Solución:

Según los resultados de la tabla de cálculo 5.15 la media geométrica será,

$$\text{Log } G = \frac{44.082}{30} = 1.4694,$$

de donde,

G = Antilogaritmo de 1.4694,

así,

G = 29.47 años

5.5.3 Características de la media geométrica

1. En el cálculo de la media geométrica se toman en cuenta todos los valores de la variable. Por esta razón, si uno de los valores de la variable es igual a cero, el valor de la media geométrica es igual a cero.

   Si $x_i = 5, 4, 7, 0, 10$

   $G = 0$
2. La media geométrica es afectada por valores extremos, aunque en menor medida que la media aritmética.
3. La media geométrica de un número y su recíproco será siempre igual a uno.

   Sea $x_i = 5, 1/5$

   $G = \sqrt{(5)\,(1/5)} = 1$

4. No puede ser calculada en distribuciones con clases abiertas.
5. Es mayormente usada para promediar tasas de cambio, razones y valores que muestren una progresión geométrica.

5.5.4 Tasa promedio de crecimiento

Cuando en una secuencia de valores, son conocidos el primero de éstos, el último y el número de éstos, entonces podemos aplicar la siguiente fórmula para determinar la tasa promedio de crecimiento.

Así,

$$r = \sqrt[n]{\left(\frac{P_n}{P_o}\right)} - 1 \qquad 5.18$$

donde,

r = tasa promedio de crecimiento.

P_o = valor al inicio del período o año base.

P_n = valor al final del período o año dado.

n = número de valores en el período.

El año base siempre será igual a cero.

Ejemplo 5.27

Según la Oficina Nacional de Estadística, la población de la República Dominicana en 1981 fue de 5,545,741 habitantes y en 1993 fue de 7,293,390 habitantes.

Determine la tasa de crecimiento promedio anual durante el período.

Solución

Sean $P_o = P1981 = 5,545,741$

$P_n = P1993 = 7,293,390$

n = 12 años (1993 menos 1981)

Sustituyendo en la fórmula 5.18,

$$r = \sqrt[12]{\frac{7,293,390}{5,545,741}} - 1$$

r = 2.30% (tasa promedio de crecimiento anual).

Ejemplo 5.28

Determine la tasa promedio de crecimiento del gasto total del gobierno central de 1992 a 1997, según los datos de la tabla 5.13 de la página 147.

Solución

Sean, $P_o = P_{1992} = RD\$\ 16{,}056{,}000{,}000$

$P_n = P_{1997} = RD\$\ 33{,}627{,}000{,}000$

n = 5 años (1997-1992 = 5 años)

Sustituyendo en la fórmula 5.18.

$$r = \sqrt[5]{\frac{33{,}627{,}000{,}000}{16{,}056{,}000{,}000}} - 1$$

Tenemos que,

r = 15.93%

Según el resultado anterior, los gastos totales del gobierno central presentaron un crecimiento promedio anual de 15.93%.

De la fórmula 5.18 podemos deducir lo siguiente,

$$\text{Sea, } r = \sqrt[n]{\frac{P_n}{P_o}} - 1$$

$$(r + 1) = \sqrt[n]{\frac{P_n}{P_o}}$$

elevando ambos términos a n, tendremos que:

$$(r+1)^n = \frac{P_n}{P_o}$$

Así,

$$P_n = P_o(r+1)^n \qquad 5.19$$

La fórmula 5.19 permite estimar un valor futuro siempre que se conozca la tasa promedio de crecimiento para el último período.

Ejemplo 5.29

La población de la República Dominicana en 1993 fue de 7,293,390 personas. La tasa promedio de crecimiento intercensal fue de 2.3%. Si asumimos que la población crecerá al mismo ritmo; ¿cuál será la población para el año 2,000?

Solución

Sean $P_o = P_{1993} = 7{,}293{,}390$ habitantes.

n = 7 años (2000-1993 = 7).

r = 0.023 = 2.3%

Como sabemos,

$$P_n = P_o(r+1)^n$$

entonces

$P_{2000} = 7{,}293{,}390\,(0.023+1)7 = 8{,}551{,}827$ habitantes.

$P_{2000} = 8{,}551{,}827$ habitantes.

Nota:

Debemos aclarar que existen otros métodos más eficientes para hacer proyecciones demográficas y económicas ya que toman en cuenta diferentes variables al momento de proyectar. Sin embargo, cuando no se exige tanta rigurosidad científica puede ser usada la fórmula 5.19.

5.6 Media armónica

Es una medida de tendencia central que se define como el recíproco de la media aritmética del recíproco de los valores de la variable.

El recíproco de un número entero es igual al cociente de 1 dividido entre el número entero (el recíproco de 5 es igual a 1/5).

El recíproco de una fracción es igual a la misma fracción con los términos invertidos (el recíproco de 3/8 es 8/3).

La media armónica la vamos a identificar como H (que se lee, media armónica de un conjunto de datos).

Es útil como medida de tendencia central cuando en el problema estudiado intervienen unidades de medidas compuestas por un cociente o por una relación inversa. Por ejemplo problemas de velocidad, de rendimiento de eficiencia, etc. pueden ser resueltos adecuadamente usando la media armónica.

5.6.1 La media armónica para datos no agrupados

Cuando queremos calcular la media armónica para datos no agrupados usamos la siguiente fórmula.

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \ldots + \frac{1}{x_n}} \qquad 5.20$$

que se puede escribir como,

$$H = \frac{n}{\sum_{i=1}^{n} (1/x_i)}$$

Ejemplo 5.30

Calcular la media armónica para los siguientes valores:

x_i: 5, 4, 8, 10, 2

$$H = \frac{5}{\frac{1}{5} + \frac{1}{4} + \frac{1}{8} + \frac{1}{10} + \frac{1}{2}} = \frac{47}{40} = 4.25$$

5.6.2 La media armónica para datos agrupados

Cuando los datos son agrupados en una distribución de frecuencias el procedimiento de cálculo es similar al anterior, con la diferencia de que ahora vamos a ponderar cada recíproco con su correspondiente frecuencia.

Esto es,

$$H = \frac{n}{f_1(1/x_1) + f_2(1/x_2) + \ldots + f_n(1/x_n)} \qquad 5.21$$

que puede escribirse como,

$$H = \frac{n}{\sum_{i=1}^{n} (f_i . 1/x_i)}$$

Ejemplo 5.31

Calcule la media armónica para la siguiente distribución:

Tabla 5.16

Clases	Frecuencia f_i	x_i	$f_i.1/x_i$
2-4	2	3	2/3
4-6	4	5	4/5
6-8	8	7	8/7
8-10	5	9	5/9
10-12	1	11	1/11
Total	**20**		

de donde,

$$H = \frac{20}{3.256} = 6.14$$

5.6.3 Características de la media armónica

1. En su cálculo se toman en consideración todos los valores de la variable al igual que la media aritmética y la media geométrica.
2. Es menos afectada por valores extremos que la media aritmética y que la media geométrica.
3. La media armónica no es usada con tanta frecuencia como un promedio como lo es la media aritmética, sin embargo es de gran uso cuando se trata de casos especiales relacionados con espacio y tiempo y cuando se trata de promediar razones.
4. La media armónica no puede ser calculada en distribuciones con clases abiertas.

5.7 Medidas de posición no central

Recordamos que la mediana divide la distribución en dos partes iguales, o sea, el 50% de los valores de la distribución será menor que la mediana y el otro 50% será mayor que la mediana. De igual manera, una distribución puede dividirse en diez partes iguales (deciles), en cuatro partes iguales (cuartiles) o en cien partes iguales (percentiles).

Así, el segundo cuartil es igual a la mediana; el quinto decil y el percentil cincuenta son también iguales a la mediana.

5.7.1 Deciles

Son medidas de posición no central que dividen la distribución en diez partes iguales.

Los deciles los vamos a identificar como DK (donde D identifica al decil y k el orden del decil) que se lee decil de orden k.

Entonces,

$$D_K = \frac{k(n+1)}{10} \qquad 5.22$$

Si k = 4

$$D_4 = \frac{4(n+1)}{10}$$

La fórmula anterior responde al cálculo del decil de orden cuatro, o sea, el 40% de la distribución.

Ejemplo 5.32

Las edades (años) de 12 estudiantes universitarios tomados al azar fueron: 18, 22, 19, 20, 21, 22, 23, 20, 19, 24, 22, 21. Calcular el decil de orden cuatro (D_4).

Solución:

Lo primero que hacemos es organizar las informaciones de menor a mayor ,

$x_i = 18, 19, 19, 20, 20, 21, 21, 22, 22, 22, 23, 24,$

luego aplicamos la fórmula 5.24.

Sea,

$$D_4 = \frac{4(n+1)}{10}$$

de donde,

$$D_4 = \frac{4(12+1)}{10} = 5.2$$ (indica la posición del decil que buscamos).

Según el resultado anterior, el decil de orden cuatro se encuentra en la posición 5.2, lo que indica que debo interpolar el verdadero valor del D_4 entre los valores 20 y 21; que representan las posiciones 5 y 6. Veamos,

$D_4 = 20 + 0.2\,(21-20)$

$D_4 = 20.2$

Este resultado significa que el 40% de las personas tienen 20.2 años o menos de edad.

5.7.2 Cuartiles

Son medidas de posición no central que dividen la distribución en cuatro partes iguales.

Los cuartiles los vamos a identificar como Q_K (donde Q identifica al cuartil y k el orden del cuartil) que se lee cuartil de orden k.

Entonces,

$$Q_K = \frac{k(n+1)}{4} \qquad 5.23$$

Si k = 1

$$Q_1 = \frac{1(n+1)}{4} \text{ (primer cuartil)}$$

$Q_2 = Me$

Ejemplo 5.33

Calcular el primer cuartil con los datos del ejemplo 5.32.

Solución:

El ejemplo 5.32 se refiere a las edades (años) de 12 estudiantes universitarios.

$x_i = 18, 19, 19, 20, 20, 21, 21, 22, 22, 22, 23, 24$

$$Q_1 = \frac{1(n+1)}{4}, \text{ sustituyendo,}$$

$$Q_1 = \frac{1(12+1)}{4} = 3.25 \text{ (indica la posición del primer cuartil).}$$

El resultado anterior indica que el primer cuartil (Q_1) se encuentra en el término o lugar 3.25; lo que significa que debo interpolar su verdadero valor que se encuentra entre las edades 19 y 20; que representan las posiciones o lugares 3 y 4.

Veamos,

$Q_1 = 19 + 0.25\,(20-19) = 19.25$ años

El resultado obtenido significa que el 25% (primer cuartil) de los estudiantes tienen 19.25 años o menos.

5.7.3 Percentiles

Son medidas de posición no central que dividen la distribución en cien partes iguales.

Los percentiles los vamos a identificar como PK (donde P identifica al percentil y k el orden del percentil) que se lee percentil de orden k.

$$P_K = \frac{k(n+1)}{100} \qquad 5.24$$

El $P_{50} = Me$

Si k = 60 entonces,

$$P_{60} = \frac{60(n+1)}{100}$$

Ejemplo 5.34

Con los datos del ejemplo 5.32, calcular el percentil sesenta. (P_{60})

Solución:

Sean las edades (años) expresadas en el ejemplo 5.32.

$x_i = 18, 19, 19, 20, 20, 21, 21, 22, 22, 22, 23, 24$

$$P_{60} = \frac{60(n+1)}{100}, \text{ sustituyendo,}$$

$$P_{60} = \frac{60(12+1)}{100} = 7.8 \text{ (indica la posición del percentil sesenta, } P_{60}).$$

Este resultado, 7.8, indica que en esta posición de los términos se encuentra el P_{60}, por lo tanto debo interpolar el verdadero valor que se encuentra entre las edades 21 y 22; que representan las posiciones 7 y 8.

$P_{60} = 21 + 0.8\,(22-21)$

$P_{60} = 21.8$ años

El 60% (percentil sesenta) de los estudiantes tienen 21.8 años o menos.

5.7.4 Deciles, Cuartiles y Percentiles para datos agrupados en una distribución de frecuencias.

El procedimiento para calcular estas medidas para datos agrupados en una distribución de frecuencias es el mismo al usado en el cálculo de la mediana, con una ligera diferencia en la determinación de la clase que contiene el valor de la mediana, los deciles, los cuartiles y los percentiles; siendo diferentes en cada caso en particular.

Las fórmulas, en sentido general, para determinar el K-ésimo decil, cuartil o percentil serán las siguientes:

$$D_k = Li + \left[\frac{\frac{kn}{10} - Fi\text{-}1}{fi}\right] Ci \qquad 5.25$$

$$Q_k = Li + \left[\frac{\frac{kn}{4} - Fi\text{-}1}{fi}\right] Ci \qquad 5.26$$

$$P_k = Li + \left[\frac{\frac{kn}{100} - Fi\text{-}1}{fi}\right] Ci \qquad 5.27$$

Ejemplo 5.35

Con los salarios mensuales de las 50 personas que se presentan a continuación calcule el D_3, Q_3 y P_{85}.

Tabla 5.17 Salarios (en RD$) de 50 personas

Salarios (RD$)	N° de personas (fi)	Fi
1400-1500	2	2
1500-1600	5	7
1600-1700	7	14
1700-1800	9	23
1800-1900	12	35
1900-2000	8	43
2000-2100	4	47
2100-2200	3	50
Total	**50**	

Solucion:

Deciles

Como sabemos,

$$D_3 = Li + \left[\frac{\frac{3n}{10} - Fi\text{-}1}{fi}\right] Ci$$

donde,

$$\frac{3n}{10} = \frac{3(50)}{10} = 15 \text{ (posición de la clase decil).}$$

Con el resultado obtenido , 3 (50) / 10 = 15, localizamos aquella clase que contiene la frecuencia acumulada igual o inmediatamente mayor a él; que será la clase que contiene los salarios de 1700-1800 y cuya frecuencia acumulada es igual a 23 (Fi=23).

Sustituyendo en la fórmula 5.25 tendremos,

$$D_3 = 1700 + \left[\frac{15 - 14}{9}\right] 100$$

$D_3 = \$\,1711.00$

El 30% de las personas ganan mensualmente RD$ 1711.00 ó menos.

Solución:

Cuartiles

$$Q_3 = Li + \left[\frac{\frac{3n}{4} - Fi\text{-}1}{fi}\right] Ci$$

donde,

$$(3n)/4 = \frac{3(50)}{4} = 37.5$$

Con dicho resultado, 37.5, localizamos aquella clase que contiene la frecuencia acumulada igual o inmediatamente superior a él, que será la clase que contiene los salarios de 1900-2000 y cuya frecuencia acumulada es igual a 43 (Fi=43).

Sustituyendo en la fórmula 5.26 tendremos,

$$Q_3 = 1900 + \left[\frac{37.5-35}{8}\right] 100$$

$$Q_3 = \text{RD\$ } 1{,}931.25$$

Este resultado significa que el 75% (tercer cuartil) de las personas ganan RD$ 1931.25 o menos mensualmente.

Solución:

Percentiles

$$P_{85} = Li + \left[\frac{\frac{85n}{100}-Fi-1}{fi}\right] Ci$$

donde,

$$(85n)/100 = \frac{85(50)}{100} = 42.5$$

Con el resultado obtenido, 42.5, localizamos la clase que contiene la frecuencia acumulada igual o inmediatamente superior a él. Esta clase es la que contiene la frecuencia acumulada igual a 43 y cuyos salarios son los comprendidos entre 1900.00 y 2000.00.

Si sustituimos en la fórmula 5.27 tendremos:

$$P_{85} = 1900 + \frac{42.5-35}{8} 100$$

de donde,

$$P_{85} = \text{RD\$ } 1{,}993.75$$

El 85% (percentil ochenta y cinco) de las personas ganan mensualmente RD$ 1,993.75 o menos.

EJERCICIOS DE APLICACIÓN

5.1 La tasa de cambio promedio /días laborables del dólar estadounidense con relación al peso dominicano durante el mes de marzo del año 2006 se presenta a continuación:

(US$ 1.00 x tasa vigente) = RD$ (peso dominicano)

32.33	32.75	32.77	32.34	32.43
32.71	32.77	32.85	32.06	32.41
32.77	32.82	32.80	32.05	32.29
32.56	32.82	32.63	32.22	
32.62	32.81	32.53	32.56	

Fuente : Banco Central de la Republica Dominicana.

a) Calcula la media aritmética de la tasa de cambio durante el mes de marzo del 2006.

5.2 Con los datos del ejercicio 3.3, de la página 69, que se refieren al salario promedio semanal que devengaban los operarios de parques de zonas francas de la República Dominicana durante el año 2001, calcule la media aritmética.

5.3 Con las edades de los 52 estudiantes universitarios encuestados en una Universidad, que aparecen en el ejercicio 3.1de la pág. 67.

a) Determine la media aritmética de las edades de esos estudiantes.

5.4 Las temperaturas (en grados celsius) registradas en una muestra de 10 sectores de la ciudad de Santo Domingo tomados al azar a las 15 horas de un día de verano fueron las siguientes:

29, 31, 30.5, 28, 27, 32, 30, 29.5, 26, 29.

a) Cuál fue la temperatura media registrada en Santo Domingo ese día y a la hora establecida?

5.5 Durante los primeros seis meses del año 2002 las importaciones de gasolina regular a la República Dominicana (millones de galones) fueron las siguientes:

X_i: importaciones (millones de galones): 25.48; 23.92; 26.98; 25.58; 26.73; 25.97;

a) Determine la media aritmética de las importaciones de gasolina regular durante el primer semestre del 2002

5.6 Según la sección 5.2.5 de la página 127 se establece que una propiedad importante de la media aritmética es:

$$\sum_{i=1}^{n}(x_i-\bar{x}) = 0$$

a) Usando la cantidad de galones de gasolina regular importados durante los primeros seis meses del año 2002 a la República Dominicana, del ejercicio 5.5, compruebe que se cumple dicha propiedad.

b) Usando las temperaturas registradas en una muestra de 10 sectores de la ciudad de Santo Domingo, del ejercicio 5.4 compruebe que se cumple dicha propiedad.

5.7 Los préstamos efectuados por una entidad crediticia durante una semana fueron:

RD$ 75,000; 60,000, 100,000; 85,000; 90,000, 88,000, 87,000; 82,000; 93,000; 68,000; 73,000, 80,000, 78,000; 85,000; 98,000; 76,000; 94,000, 84,000.

a) Determine la media aritmética de los préstamos efectuados por la entidad crediticia.

5.8 La llegada mensual de pasajeros, vía aérea, a la República Dominicana, durante el año 2005 se presenta a continuación:

Mes	No. de pasajeros
Enero	375,194
Febrero	354,168
Marzo	400,792
Abril	323,551
Mayo	280,468
Junio	331,878
Julio	450,952
Agosto	368,794
Septiembre	226,503
Octubre	251,012
Noviembre	299,194
Diciembre	418,789

a) Calcule la media aritmética de la llegada mensual de pasajeros a la República Dominicana por vía aérea.

5.9 Si a las tasas de cambio promedio [illegible] días [illegible] dol[illegible] estadounidenses con relación al peso dominicano durante el mes de marzo del año 2006, que se presenta en el ejercicio 5.1 de la pág. 163 se le incrementa en RD$ 0.50, cual será la nueva media aritmética de la tasa de cambio durante el mes.

5.10 ¿Cuál será la nueva media aritmética de la llegada mensual de pasajeros por vía aérea a la República Dominicana, del ejercicio 5.8, si se produce un incremento mensual de un 12%?

5.11 Los salarios semanales, RD$, de los veinte técnicos de la empresa CP, S.A., de Zona Franca son los siguientes:

RD$ 1,800; 1,500; 1,000; 1,100; 1,400, 1,350; 1,200; 1,300; 1,275; 1,700; 1,600; 1,500; 1,350, 1,650; 1,500; 1,600; 1,300; 1,150; 1,500; 1,600..

Los técnicos de CP, S.A. han acordado exigir un aumento de salario semanal.

Algunos obreros establecen que deben solicitar un incremento de RD$150.00 semanales.

Otros establecen que deben exigir un incremento de un 10%.

Cuál de las dos alternativas de solicitud de incremento salarial es que más le conviene a los obreros?

5.12 Con los datos del ejercicio 5.11, que muestra los salarios semanales de veinte técnicos de la empresa CP, S.A.

a) Calcule la media aritmética de los salarios.

b) Si los salarios se incrementan en un 10%, ¿cuál será la nueva media aritmética de estos?

5.13 Según la sección 5.2.5 de la página 131 se establece que una propiedad de la media aritmética es:

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 < \sum_{i=1}^{n}(x_i-A)^2 \; \forall A \neq \bar{x}$$

a) Usando las temperaturas registradas en una muestra de 10 sectores de la ciudad de Santo Domingo, del ejercicio 5.4 compruebe que se cumple dicha propiedad.

b) Usando la cantidad de galones de gasolina regular importados durante los primeros seis meses del año 2002 a la República Dominicana, del ejercicio 5.5, compruebe que se cumple dicha propiedad.

5.14 En una competencia de velocidad de Fórmula Primera, los primeros siete pilotos en completar la competencia lograron las siguientes velocidades en la última vuelta:

198 millas / hora
195 millas / hora
194 millas / hora
196 millas / hora
193 millas / hora
197 millas / hora
190 millas / hora

a) Determine la media aritmética de las velocidades de los siete pilotos de Fórmula Primera logrados en la última vuelta de la competencia.

5.15 Durante el desarrollo de un semestre, un estudiante de estadística básica obtuvo las siguientes calificaciones:

Examen	Calificación	Ponderación (%)
Primer Parcial	80	10.0
Segundo Parcial	75	15.0
Final	82	45.0
Prácticas	84	30.0
Total		**100.0**

a) Calcule la media aritmética ponderada de las calificaciones de estadística básica.

5.16 El número de hijos por familia se presenta en la siguiente tabla.

N° de hijos (x_i)	Cantidad de familias (f_i)
0	5
1	12
2	16
3	18
4	17
5	10
6	3
7	2
Total	**83**

a) Calcule la media aritmética del número de hijos por familias.

5.17 Una fábrica de camisas ha determinado mediante la inspección de su producción por turnos, durante una semana, los siguientes porcentajes de camisas con defectos.

Turno	Porcentaje camisas con defectos (x_i)	Cantidad de camisas (w_i)
A	1.25	460
B	2.30	275
C	2.50	125

a) Calcule la media aritmética ponderada del porcentaje de camisas producidas con defectos.

5.18 Sueldos mensuales (RD$) de 80 empleados del sector público.

Sueldos (RD$)	Número de empleados
1800-2300	5
2300-2800	12
2800-3300	15
3300-3800	18
3800-4300	14
4300-4800	10
4800-5300	6
Total	**80**

a) Calcule la media aritmética del sueldo mensual de los 80 empleados.

5.19 Calcule la media aritmética de las calificaciones obtenidas por 75 estudiantes universitarios en un curso de matemáticas.

Calificación	Número de estudiantes
40-50	8
50-60	10
60-70	13
70-80	21
80-90	16
90-100	7
Total	**75**

5.20 La siguiente tabla presenta el número de accidentes de tránsito reportados durante un año, según estudio efectuado a una muestra de conductores en una ciudad de la República Dominicana.

Número de Accidentes (x_i)	Número de Conductores (f_i)
0	575
1	450
2	156
3	51
4	10
5	2
Total	**1244**

a) Calcular la media aritmética del número de accidentes por conductor.

5.21 Con los datos del ejercicio 4.32 de la página 105 que se refiere a los ingresos anuales de 475 empleados de una empresa de producción industrial; calcule la media aritmética de dichos ingresos.

5.22 La compañía de auditores C.P. requieren sus técnicos para auditar las cuentas de una empresa de servicios. La tabla que se presenta a continuación muestra el tiempo requerido y el número de cuentas.

Tiempo para auditar (minutos)	Número de cuentas
15-20	2
20-25	5
25-30	6
30-35	8
35-40	11
40-45	15
45-50	7
50-55	4
55-60	2
Total	**60**

a) Calcule el tiempo medio necesario para que los técnicos de la compañía de auditores CP, S.A. auditen cada cuenta.

Mediana

5.23 El número de pasajeros que abordaron 10 autobuses del transporte público al salir de su terminal, durante una hora, fue como sigue:

No. de Pasajeros (X_i): 44, 46, 33, 43, 53, 65, 55, 36, 54, 48.

a) Calcule la mediana del número de pasajeros que abordaron los 10 autobuses durante una hora.

5.24 Calcule la mediana con los datos del ejercicio 5.1 de la página 163 que se refieren a la tasa de cambio promedio/ días laborales del dólar estadounidense con relación al peso dominicano durante el mes de marzo del año 2006.

5.25 Calcule la mediana de la llegada mensual de pasajeros por vía aérea a la República Dominicana durante el año 2005, que aparecen en el ejercicio 5.8 de la página 164.

5.26 Durante una semana los obreros de una empresa energética trabajaron las siguientes horas extras.

Horas extras (X_i): 20, 15, 17, 21, 25, 10, 23, 28, 24, 19, 16, 19, 20, 24, 22, 25, 14, 16 ,18, 25, 24, 22, 19, 17, 26

a) Determine la mediana de las horas extras trabajadas por los obreros.

5.27 A cada persona de una muestra de 16 que participaron en un estudio se le asignó una serie de tareas y se anotó el tiempo de terminación (minutos) de cada una de ellas.

18.5	14.6	21.4	17.6	15.4	16.7	22.4	17.8
14.2	18.4	15.7	20.3	23.1	22.8	21.5	24.4

a) Calcule la mediana del tiempo empleado por cada persona en la terminación de la tarea asignada.

b) Comente este resultado.

5.28 Calcule la mediana de los sueldos mensuales (RD$) de los 80 empleados del sector público, del ejercicio 5.18 de la página 167.

5.29 Determine la calificación que obtuvo el 50% o menos de los estudiantes universitarios del curso de matemáticas, del ejercicio 5.19 de la página 168.

5.30 Las edades de cada uno de los visitantes extranjeros no residentes que arribaron por vía aérea a la República Dominicana durante el año 2005 se presentan en la siguiente tabla:

Edad (años)	Número de visitantes extranjeros
0-12	251,335
13-20	282,808
21-35	1,048,109
36-49	842,230
50 años y mas	663,735
Total	**3,088,247**

Fuente: Banco Central de la Republica Dominicana

a) Calcule la mediana de las edades de los visitantes extranjeros que llegaron al país durante el año 2005.

b) Comente el resultado obtenido.

5.31 El tiempo de espera (minutos) de los clientes de un restaurant para recibir sus órdenes son:

Tiempo (minutos)	Número de clientes
5-10	10
10-15	18
15-20	15
20-25	8
25-30	5
30 y más	2
Total	**58**

a) Calcule el tiempo mediana de espera.

b) Cómo usted considera la calidad del servicio en base al tiempo de espera de cada cliente?

5.32 Los montos de los 70 préstamos otorgados por una entidad financiera a la pequeña y mediana empresa son:

Monto de los préstamos (en miles de RD$)	Número de Préstamos
50-70	5
70-90	7
90-110	11
110-130	16
130-150	13
150-170	9
170-190	6
190-210	3
Total	**70**

a) Calcule la mediana del monto de los préstamos otorgados.

5.33 Calcule la mediana con los datos del ejercicio 5.16 de la página 167 que se refiere al número de hijos por familias.

5.34 La siguiente tabla muestra la distribución de las edades de los cabezas de familia en una localidad de la República Dominicana.

Edad (años)	Cabezas de familia
Menor de 25	175
25-34	208
35-44	300
45-54	568
55-64	425
65-74	302
75 y más	150
Total	**2128**

a) Calcule la edad mediana de los cabezas de familia.

b) Comente el resultado obtenido.

La moda

5.35 Los salarios semanales de 15 obreros de una empresa de zona franca fueron:

RD$ 1400, 1500, 1200, 1400, 1450,
1475, 1300, 1550, 1400, 1350,
1400, 1525, 1400, 1600, 1250.

a) Determine la moda de los salarios semanales de los 15 obreros.

5.36 Las calificaciones en estadística de 25 estudiantes se muestran a continuación.

80, 85, 72, 68, 65, 79, 81, 85, 79, 78, 79, 81, 83, 79, 81, 80,
81, 76, 70, 60, 63, 81, 78, 79, 67.

a) Determine la moda de las calificaciones.

5.37 Calcule la moda con los datos del ejercicio 5.18 de la página 167 que se refieren a los sueldos mensuales de 80 empleados del sector público.

5.39 Después de revisar un libro, un especialista en gramática encontró la siguiente cantidad de errores por página.

Número de Errores	Cantidad de Páginas
0-2	8
3-5	14
6-8	16
9-11	9
12-14	4
Total	**51**

a) Determine el número de errores modal por página.

5.40 Calcule el tiempo modal necesario para que los técnicos de la compañía de auditores CP, S.A., del ejercicio 5.22 de la página 169, auditen estas cuentas.

5.41 Las estaturas de 53 personas se presentan en la siguiente tabla.

Estatura (pulgadas)	Número de personas
60-62	2
62-64	5
64-66	7
66-68	10
68-70	18
70-72	8
72-74	3
Total	**53**

a) Calcule la moda de las estaturas de las 53 personas.

Media geométrica

5.42 Calcule la media geométrica para las siguientes series de datos.

a) x_i: 5, 8, 10, 12, 3, 16, 17.
b) x_i: 2, 5, 6, 4, 3.
c) x_i: 25, 28, 18, 0, 19, 24, 17, 23.
d) x_i: 4, 1/4.
e) x_i: 10, 11, 8, 7, 12, 16, 9.

5.43 Las tasas de interés activa que aplicaron los bancos comerciales de la República Dominicana a sus clientes a un plazo de seis meses a un año durante el periodo 1995-1999 fueron las siguientes:

Años	Tasa de interés activa (%)
1994	29.06
1995	30.31
1996	23.49
1997	20.78
1998	26.18
1999	25.17

Fuente: Banco Central de la República Dominicana.

a) Calcule la media geométrica de la tasa de interés activa para el período.

5.44 Las recaudaciones de la República Dominicana por impuestos al tabaco manufacturado durante el periodo 1990-1999 se presentan a seguidas.

Años	Recaudaciones (Mill. RD$)
1990	102.4
1991	93.7
1992	99.9
1993	113.4
1994	135.1
1995	262.1
1996	359.1
1997	358.2
1998	381.9
1999	403.3
2000	469.0
2001	996.3

Fuente: Banco Central de la República Dominicana y Periódico Hoy, 29-10-2000.

a) Calcule la media geométrica de las recaudaciones durante el periodo.

5.45 Las tasas de crecimiento del sector Hoteles, Bares y Restaurantes (Turismo) por años se presentan a continuación.

Años	Tasa de Crecimiento (%)
1994	15.0
1995	16.6
1996	12.8
1997	17.4
1998	4.7
1999	10.0

Fuente: Banco Central de la República Dominicana.

a) Calcule la media geométrica de las tasas de crecimiento del sector turismo.

5.46 Con los datos del ejercicio 5.19 página 168 que se refieren a las calificaciones obtenidas por 75 estudiantes universitarios en un curso de matemáticas, calcule la media geométrica.

5.47 Con los datos del ejercicio 5.22 de la página 169 referentes al tiempo necesario requerido por sus técnicos para auditar las cuentas de una empresa de servicios, calcule la media geométrica.

5.48 Los ingresos fiscales de la República Dominicana por concepto del cobro de ITBIS (Impuesto a la Transferencia de Bienes Industrializados y Servicios) durante el año 1991 fue de RD$ 1,467.6 millones y en el año 2001 los ingresos ascendieron a RD$ 14,225.3 millones

Fuente: Banco Central de la República Dominicana

a) Calcule la tasa de crecimiento promedio anual durante el período.

5.49 El gasto total del Gobierno Central de la República Dominicana fue en 1999 de RD$ 48,758.7 millones y en el 2001 fue de RD$ 58,863.4 millones

a) Calcule la tasa de crecimiento promedio anual del gasto durante el período.

5.50 El personal ocupado del gobierno dominicano (gobierno central, instituciones autónomas y descentralizadas y los municipios) al mes de septiembre de 1986 era de 282,920 personas al mes de septiembre de 1998 era de 305,861 personas y al mes de septiembre del 2002 habían 406,493 empleados en la administración pública.

Fuente: Banco Central de la República Dominicana.

a) Calcule la tasa de crecimiento promedio anual del personal ocupado del gobierno dominicano durante el período 1986-1998.

b) Calcule la tasa de crecimiento promedio anual del personal ocupado durante el periodo 1998-2002.

c) Calcule la tasa de crecimiento promedio anual del personal ocupado durante el periodo1986-2002.

5.51 El gasto total del gobierno central de la República Dominicana fue en 1993 de RD$ 20,286.6 millones y en el 2001 fue de RD$ 58,863.4 millones

a) Calcule la tasa de crecimiento promedio anual del gasto durante el período citado.

b) Si asumimos que el gasto, crecerá al mismo ritmo, cuál será el gasto estimado para el año 2005?

5.52 Las recaudaciones de la República Dominicana por impuestos al tabaco manufacturado en el año 1990 fueron de RD$ 102.4 millones y en el 1999 fueron de RD$ 403.3 millones.

a) Calcule la tasa de crecimiento promedio anual de las recaudaciones durante el período.

5.53 El gasto público en la educación de República Dominicana durante el año 1990 fue de RD$ 707.5 millones y durante el año 2001 fue de RD$ 10,470 millones.

a) Calcule la tasa de crecimiento promedio anual del gasto público en educación durante el período señalado.

5.54 Las recaudaciones fiscales de la República Dominicana a través de la Dirección General de Impuestos Internos* durante el período 1995 – 2005 se presentan a continuación:

Años	Recaudaciones (Millones de RD$)
1995	10,648.10
1996	12,068.50
1997	15,219.90
1998	16,795.4
1999	20,219.0
2000	23,392.9
2001	31,240.4
2002	34,346.4
2003**	42,978.9
2004**	61,259.8
2005**	82,846.2

Fuente: Banco Central de la República Dominicana y Dirección General de Impuestos Internos.

* Antes del año 1997 esta dirección la formaban la Dirección General de Impuestos sobre la Renta y la Dirección General de Rentas Internas.

** Cifras preliminares.

a) Calcule la media geométrica del incremento de las recaudaciones fiscales.

b) Calcule la tasa de crecimiento promedio anual durante el período.

Media armónica

5.56 Calcule la media armónica de las siguientes series de datos.

a) $x_i = 5, 4, 3, 8, 10.$

b) $x_i = 10, 12, 15, 18, 23, 19, 21.$

c) $x_i = 20, 21, 18, 17, 23, 24, 28, 16, 15.$

5.57 Un grupo de obreros realiza la construcción de los primeros 500 metros de una vía con un rendimiento de 40 metros por día y los segundos 500 metros con un rendimiento de

25 metros por día. ¿Cuál es el rendimiento promedio en los 1000 metros de construcción de la vía?

5.58 En un rally donde se recorrieron 1500 kilómetros, un equipo de cuatro profesores manejó 375 kilómetros cada uno. A continuación se presentan sus velocidades al manejar.

Profesor	Velocidad
A	80 km / hora
B	95 km / hora
C	90 km / hora
D	105 km / hora

a) Determine la media armónica de la velocidad por hora de los cuatro profesores durante el rally.

5.59 Con los datos del ejercicio 5.22 de la página 169, que se refieren al tiempo necesario que requieren los técnicos de la compañía de auditores CP, S.A. para auditar las cuentas de una empresa de servicios, calcule la media armónica.

5.60 Con los datos del ejercicio 5.31 de la página 171, que se refiere al tiempo de espera de los clientes de un restaurant para recibir sus órdenes, calcule la media armónica.

5.61 El barco A dura 15 días atravesando el Atlántico; el barco B dura 18 días y el barco C dura 22 días.

Determine la media armónica de los días que duran en atravesar.

Deciles - Cuartiles - Percentiles

5.62 Con los datos del ejercicio 5.1 de la página 163, que se refieren a la tasa de cambio promedio/días laborales del dolar estadounidense con relación al peso dominicano durante el mes de marzo del 2006, calcule:

a) El cuarto decil;
b) El tercer cuartil;
c) El percentil ochenta y cinco;
d) Comente cada uno de los resultados obtenidos.

5.63 Con los datos del ejercicio 5.26 de la página 170, referentes a las horas extras que trabajaron 25 obreros de una empresa energética durante una semana, calcule:

a) El sexto decil.
b) El primer cuartil.
c) El percentil cincuenta.
d) Comente cada uno de los resultados.

5.64 Con los datos del ejercicio 5.11 de la página 165, que se refieren a los salarios semanales RD$, de los veinte obreros de la empresa CP, S.A., de Zona Franca; calcule:

a) El cuarto decil.
b) El segundo cuartil.
c) El percentil ochenta y cinco.
d) Comente cada uno de los resultados.

5.65 Con los datos del ejercicio 5.18 de la página 167, referentes a los sueldos mensuales de 80 empleados del sector público; determine:

a) El valor del primer cuartil.
b) El valor del sexto decil.
c) El valor del percentil setenta de los sueldos.

5.66 Las áreas de construcción (mts2) de las 150 viviendas de un proyecto habitacional se presentan en la siguiente tabla:

Área de Construcción (Mts.^{2})	Número de Viviendas
70- 80	8
80- 90	12
90-100	17
100-110	25
110-120	40
120-130	26
130-140	16
140-150	6
Total	**150**

a) Determine el valor del sexto decil de las áreas de construcción de las viviendas.

b) Determine el primer cuartil de las áreas de construcción de las viviendas.

c) Determine el valor del percentil noventa de las áreas de construcción de las viviendas.

d) Comente cada uno de los resultados.

5.67 Los ingresos mensuales (en RD$) de 110 familias tomadas al azar en Santo Domingo aparecen en la siguiente tabla.

Ingresos (RD$)	Número de familias
4000-5000	14
5000-6000	18
6000-7000	25
7000-8000	23
8000-9000	17
9000 y más	13
Total	**110**

a) Determine el valor del percentil cincuenta de los ingresos familiares.

b) Determine el cuarto decil de los ingresos familiares.

c) Determine el tercer cuartil de los ingresos familiares.

d) Determine el percentil sesenticinco de los ingresos familiares.

e) Comente cada uno de los resultados.

CAPÍTULO

VI

Medidas de dispersión

6.0 Introducción

Dos conjuntos diferentes de datos pueden contener el mismo número de objetos y tener la misma media aritmética, pero un conjunto puede estar más disperso, alrededor del valor promedio que el otro.

Es necesario conocer dos tipos de medidas para describir en forma adecuada un conjunto de datos o informaciones numéricas. Además de tener un indicador que exprese la parte media del conjunto de datos es conveniente, también, disponer de un método que nos permita expresar la cantidad de dispersión existente entre los diferentes valores de la variable.

Las medidas de dispersión indican si los valores de la variable de un conjunto de datos están relativamente cercanos uno del otro o si por el contrario se encuentran dispersos o alejados uno de otro. Por ejemplo, si tenemos una medida de dispersión muy baja de los salarios quincenales de los obreros de una fábrica de prendas de vestir, dará la impresión o indicará que a los obreros se les paga salarios prácticamente iguales. En cambio, una alta dispersión de los salarios indicará al lector que los obreros de esa fábrica tienen salarios que varían de manera muy significativa. Visto lo anterior, podemos señalar que cuando una medida de dispersión es muy baja, el promedio se vuelve altamente significativo, en cambio, cuando la dispersión es muy alta, el promedio se considera de muy poca significación.

Las medidas de dispersión se clasifican en absolutas y relativas, las cuales estudiaremos a continuación.

6.1 Medidas de dispersión absolutas

Estas medidas de dispersión se caracterizan por expresarse en las mismas unidades que los valores de la variable y miden la variabilidad existente entre los valores de la variable y un determinado promedio. Por esta razón es imposible hacer comparaciones entre distribuciones con unidades diferentes. Por ejemplo, no es posible comparar la dispersión absoluta de una distribución cuya unidad de medida es año, con otra cuya unidad de medida es RD$.

Las principales medidas de dispersión absolutas son:

1.- Recorrido o Rango
2.- Desviación intercuartílica
3.- Desviación media
4.- Varianza
5.- Desviación típica o estándar

6.1.1 Recorrido o rango

El recorrido o rango de un conjunto de valores es la diferencia entre el valor máximo observado y el mínimo.

$$R = Xmáx - Xmín \qquad 6.1$$

donde Xmáx = valor más alto y Xmín = valor más pequeño.

En el cálculo del recorrido sólo se toman en cuenta dos valores, el más alto y el más pequeño. Por tanto, el recorrido es considerado como una burda estimación de la dispersión absoluta.

Ejemplo 6.1

Encontrar el recorrido de los siguientes valores;

x_i: = 5, 8, 4, 10, 12, 16, 13.

Solución:

R = 16 - 4 = 12

6.1.2 Desviación intercuartílica o cuartílica

La desviación entre cuartiles de un conjunto de datos, al igual que el recorrido, sólo está basada en dos valores; el primer y el tercer cuartil. Por esta razón también se le considera como una tosca medida de dispersión absoluta.

La desviación intercuartílica se identifica como,

DQ = Desviación intercuartílica.

donde,

$$DQ = \frac{Q_3 - Q_1}{2} \qquad 6.2$$

siendo

Q_3 = Cuartil de orden tres.

Q_1 = Cuartil de orden uno.

Ejemplo 6.2

Calcule la desviación intercuartílica de los salarios de las 50 personas que se presentan a continuación.

Tabla 6.1 Salarios (en RD$) de cincuenta personas

Salarios (RD$)	No. de personas (fi)	Fi
1400 - 1500	2	2
1500 - 1600	5	2+5 = 7
1600 - 1700	7	7+7 = 14
1700 - 1800	9	14+9 = 23
1800 – 1900	12	23+12 = 35
1900 – 2000	8	35+8 = 43
2000 – 2100	4	43+4 = 47
2100 – 2200	3	47+3 = 50
Total	**50**	

Solución:

La fórmula para calcular la desviación intercuartílica es,

$$DQ = \frac{\quad - \quad}{2}$$

donde,

$$Q_3 = Li + \left[\frac{\frac{3n}{4} - Fi\text{-}1}{fi}\right] Ci \qquad 6.3$$

$$Q_1 = Li + \left[\frac{\frac{n}{4} - Fi\text{-}1}{fi}\right] Ci \qquad 6.4$$

Sustituyendo en ambas fórmulas tendremos;

$\frac{3(50)}{4} = 37.5$ (posición de la clase del tercer cuartil)

$$Q_3 = 1900 + \left[\frac{37.5 - 35}{8}\right]100$$

$Q_3 = RD\$1931.25$

Asimismo,

$n/4 = 50/4 = 12.5$ (posición de la clase del primer cuartil).

$$Q_1 = 1600 + \left[\frac{12.5 - 7}{7}\right]100$$

$Q_1 = RD\$1678.57$

Sustituyendo en la fórmula de la desviación intercuartílica,

$$DQ = \frac{1931.25 - 1678.57}{2} = 126.34$$

$DQ = RD\$126.34$

Este resultado indica la variación entre los elementos centrales de la distribución.

6.1.3 Desviación media

Hasta ahora hemos estudiado el recorrido y la desviación intercuartílica, que son medidas de dispersión absolutas que se basan en las posiciones de ciertos valores en la distribución.

La desviación media, por el contrario, mide la dispersión del conjunto de valores de la variable con relación a un promedio.

La desviación media se define como la media aritmética del valor absoluto de las desviaciones individuales de los valores de la variable con relación a un promedio.

En el cálculo de la desviación media se toman en cuenta todos los valores de la variables, a diferencia del recorrido y la desviación intercuartílica.

Al calcular la desviación media, ignoramos el signo positivo o negativo de las desviaciones, al considerarse ésta como la media aritmética del valor absoluto de las desviaciones, razón por la cual la hace una débil medida de dispersión absoluta.

La desviación media la vamos a identificar como,

DM = desviación media.

cuya fórmula se ilustra a continuación,

$$DM = \frac{\sum | x_i - \bar{x} |}{n} \qquad 6.5$$

Cuando se trata de su cálculo para datos no agrupados, y,

$$DM = \frac{\sum | x_i - \bar{x} | fi}{n} \qquad 6.6$$

Cuando se trata de calcular la desviación media para un conjunto de datos agrupados en una distribución de frecuencias.

Calcule la desviación media para los siguientes valores,

$x_i = 5, 7, 10, 9, 4$

Solución:

Como se trata de un conjunto de datos no agrupados, vamos a aplicar la fórmula 6.5 cuyo procedimiento consiste en,

1- determinar la media aritmética ($\bar{x}$) del conjunto de datos,

$$\bar{x} = \frac{\sum x_i}{n}$$

$$\bar{x} = \frac{35}{5} = 7$$

2- Calcular el valor absoluto de las desviaciones individuales de los valores de la variable, con relación a la media aritmética,

$| x_i - \bar{x} |$ = valor absoluto de las desviaciones individuales.

3- Determinar la desviación media al dividir la suma del valor absoluto de las desviaciones individuales entre el total de casos.

Veamos a continuación:

Tabla 6.2 Cálculos

x_i	$\lvert x_i - \bar{x} \rvert$
5	$\lvert 5-7 \rvert = \lvert -2 \rvert = 2$
7	$\lvert 7-7 \rvert = \lvert 0 \rvert = 0$
10	$\lvert 10-7 \rvert = \lvert 3 \rvert = 3$
9	$\lvert 9-7 \rvert = \lvert 2 \rvert = 2$
4	$\lvert 4-7 \rvert = \lvert -3 \rvert = 3$
35	$\sum \lvert x_i - \bar{x} \rvert = 10$

$$\bar{x} = \frac{\sum x_i}{n}$$

$$\bar{x} = \frac{35}{5} = 7$$

Sustituyendo los resultados parciales en la fórmula 6.5, tendremos:

$$DM = \frac{\sum | x_i - \bar{x} |}{n} = \frac{10}{5} = 2$$

$$DM = 2$$

Ejemplo 6.4

Calcule la desviación media de los salarios de las 50 personas que se presentan a continuación.

Tabla 6.3 Salarios (en RD$) de cincuenta personas

Salarios (RD$)	No. de peronas (fi)	x_i	$x_i fi$	$\lvert x_i - \bar{x} \rvert$	$\lvert x_i - \bar{x} \rvert fi$
1400 - 1500	2	1450	2900	358	716
1500 - 1600	5	1550	7750	258	1290
1600 - 1700	7	1650	11550	158	1106
1700 - 1800	9	1750	15750	58	522
1800 - 1900	12	1850	22200	42	504
1900 - 2000	8	1950	15600	142	1136
2000 - 2100	4	2050	8200	242	968
2100 - 2200	3	2150	6450	342	1026
Total	**50**		**90400**		**7268**

Solución:

Como estamos ante una distribución de frecuencias, aplicamos la fórmula 6.6 usada en los casos de datos agrupados.

Así, tendremos que,

$$DM = \frac{\sum | x_i - \bar{x} | fi}{n} \qquad 6.7$$

Procedimiento

Procedemos a calcular la media aritmética,

$\bar{x} = \frac{\sum x_i fi}{n}$, para datos agrupados.

La $\bar{x} = \frac{90400}{50} = \$\ 1808.00$

Ahora, determinaremos el valor absoluto de las desviaciones de los valores de la variable con relación a la media aritmética,

$| x_i - \bar{x} |$ = valor absoluto de las desviaciones

Luego de calcular para cada clase el valor absoluto de las desviaciones de los valores, multiplicamos por cada frecuencia simple de clase; $| x_i - \bar{x} | fi$.

Finalmente, determinamos la desviación media de la distribución, al dividir la sumatoria del producto de las desviaciones individuales de los valores por sus frecuencias correspondientes,

$\sum | x_i - \bar{x} | fi = 7268$, entre el número de observaciones (n).

$$DM = \frac{\sum | x_i - \bar{x} | fi}{n} = \frac{7268}{50} = RD\$\ 145.36$$

Este resultado nos indica la variabilidad o dispersión de los valores de la variable con relación a la media aritmética; es importante recordar que esta medida presenta la debilidad de que considera las desviaciones individuales como un valor absoluto.

6.1.4 La varianza

La varianza o variancia de un conjunto de datos expresa la variabilidad promedio de las desviaciones cuadráticas de los valores de la variable con respecto de la media aritmética.

La varianza la vamos a identificar como,

S^2 = (varianza, cuando se calcula en una muestra)

σ^2 = (varianza, cuando se calcula en una población)

En el cálculo de la varianza se toman en cuenta todos los valores de la variable. Además, no se ignora el signo de las desviaciones, pero su resultado se presenta como una expresión cuadrática, lo que la hace inmanejable.

El valor mínimo que puede tomar la varianza es el cero y esto ocurre cuando todos los valores de la variable son iguales.

Para una mayor ilustración, veamos la fórmula de la varianza,

Para datos no agrupados

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} \qquad 6.8$$

Para datos agrupados

$$S^2 = \frac{\sum (x_i - \bar{x})^2 fi}{n-1} \qquad 6.9$$

Antes de proceder a la realización de ejemplos para ilustrar esta medida de dispersión, estudiaremos la medida de dispersión absoluta ideal y más usada en estadística.

6.1.5 Desviación típica o estándar

La desviación típica o estándar de un conjunto de valores se define como la raíz cuadrada de la varianza.

Es la medida de dispersión absoluta más usada en estadística.

En su cálculo se toman en cuenta todos los valores de la variable, razón por la cual permite una mejor descripción de la dispersión que el recorrido y la desviación intercuartílica.

Mide la dispersión alrededor de la media, no la dispersión dentro de ciertos valores, como es medida por el recorrido y la desviación intercuartílica.

Al calcular la desviación estándar no se ignoran los signos de las desviaciones individuales de los valores, su resultado se presenta como una expresión lineal, a diferencia de la varianza, que se presenta como una expresión cuadrática.

Estas razones hacen que se considere la desviación estándar como la mejor medida de dispersión absoluta.

La desviación estándar se identifica con la siguiente simbología,

S = desviación estándar, cuando se calcula en una muestra

σ = desviación estándar cuando se calcula en la población.

La fórmula de la desviación estándar o típica es ilustrada a continuación,

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} \quad \Longrightarrow \text{para datos no agrupados} \quad 6.10$$

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2 f_i}{n-1}} \quad \Longrightarrow \text{para calcularla para datos agrupados en una distribución de frecuencias.} \quad 6.11$$

6.1.5.1 La desviación estándar para datos no agrupados

Cuando se calcula la desviación estándar para un conjunto de datos no agrupados vamos a proceder de la siguiente manera:

1- Calcular la media aritmética de los datos,

$$\bar{x} = \frac{\sum x_i}{n}$$

2- Determinar las desviaciones individuales de los valores de la variable con respecto a la media aritmética,

(n -1) = desviaciones de los valores de la variable con respecto a la media aritmética.

3- Elevar al cuadrado las desviaciones individuales de los valores de la variable con relación a la media y así obtener lo que hemos llamado, desviaciones cuadráticas.

$(x_i - \bar{x})$ = desviaciones cuadráticas

4- Determinar la suma de las desviaciones cuadráticas de los valores de la variable con relación a la media aritmética,

$\sum (x_i - \bar{x})^2$

5- Obtener la varianza, al dividir la suma de las desviaciones cuadráticas de los valores entre el número de datos (n ó $\sum f_i$).

$$S^2 = \frac{\sum x_i - \bar{x})^2}{n}$$

6- Obtener la desviación estándar, extrayendo la raíz cuadrada de la varianza.

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

Ejemplo 6.5

Calcule la desviación estándar de los pesos (libras) al nacer de 8 niños.

x_i: 7.0, 8.5, 6.5, 5.0, 4.5, 10.0, 7.5, 7.0

Solución

Tabla 6.4 Pesos (en libras) de ocho niños al nacer.

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
7.0	0	0
8.5	1.5	2.25
6.5	-0.5	0.25
5.0	-2.0	4.0
4.5	-2.5	6.2
10.0	3.0	9.0
7.5	0.5	0.25
7.0	0	0
56.0		$\sum (x_i - \bar{x})^2 = 22$

La media aritmética del peso al nacer de los ocho niños es,

$$\bar{x} = \frac{\sum x_i}{n} = \frac{56}{8} = 7 \text{ libras}$$

Sustituyendo los resultados parciales anteriores en la fórmula 6.8, tenemos que,

$$S^2 = \frac{\sum x_i - \bar{x})^2}{n} = \frac{22}{8} = 2.75 \text{ libras}$$

de donde,

$$S = \sqrt{3.14} = 1.77 \text{ libros}$$

6.1.5.2 La desviación estándar para datos agrupados

Cuando los datos se presentan agrupados en una distribución de frecuencias, el procedimiento para el cálculo de la desviación estándar es el siguiente:

1- Calcular la media aritmética de la distribución.

$$\bar{x} = \frac{\sum x_i fi}{n}$$

2- Determinar las desviaciones individuales de los valores de la variable con relación a la media aritmética.

$(xi - \bar{x})$ = desviaciones de los valores con relación a la media aritmética.

3- Elevar al cuadrado las desviaciones individuales de los valores de la variable con relación a la media aritmética y así obtener las desviaciones cuadráticas.

$(xi - \bar{x})^2$ = desviaciones cuadráticas.

4- Multiplicar las desviaciones cuadráticas de los valores con relación a la media aritmética por sus correspondientes frecuencias.

$(x_i - \bar{x})^2 fi$

5- Obtener la sumatoria del producto de las desviaciones cuadráticas de los valores con relación a la media aritmética por sus correspondientes frecuencias.

$\sum (x_i - \bar{x})^2 fi$

6- Determinar la varianza, al dividir la sumatoria del producto de las desviaciones cuadráticas de los valores con relación a la media aritmética entre el número total de datos ($\sum fi$ ó n)

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

7- Calcular la desviación estándar, extrayendo la raíz cuadrada de la varianza.

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

Ejemplo 6.6

El tiempo de espera (minutos) de los clientes que se presentaron a realizar alguna transacción a un banco comercial de Santo Domingo se presenta a seguidas. Calcule la desviación estándar del tiempo de espera.

Tabla 6.5 Tiempo de espera (en minutos) de los clientes que visitan un banco

Tiempo (minutos)	No. De clientes (fi)	x_i	$x_i fi$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 fi$
0-5	8	2.5	20	-11.69	136.65	1093.2
5-10	12	7.5	90	-6.69	44.75	537
10-15	25	12.5	312.5	-1.69	2.85	71.25
15-20	20	17.5	350	3.31	10.95	219
20-25	10	22.5	225	8.31	69.05	690.50
25-30	5	27.5	137.5	13.31	177.15	885.75
Total	**80**		**1135.0**			**3496.70**

La media aritmética del tiempo de espera de los clientes que realizaron alguna transacción en un banco comercial de Santo Domingo fue,

$$\bar{x} = \frac{135}{80} = 14.9 \text{ minutos}$$

Sustituyendo los resultados parciales anteriores en la fórmula 6.9, tenemos que,

$$S^2 = \frac{\sum (xi - \bar{x})^2 fi}{n-1} \Longrightarrow S^2 = \frac{3496.7}{80-1} = 44.25 \text{ minutos},$$

de donde,

$$S = \sqrt{44.25} = 6.65 \text{ minutos}$$

El resultado obtenido anteriormente indica la dispersión o variabilidad existente entre los valores de la variable y la media aritmética, o sea la "distancia" promedio entre los valores de la variable y su correspondiente media aritmética.

Como se trata de un valor absoluto, expresado en la misma unidad de medida (minutos) que los valores de la variable, no podemos hacer comparaciones con distribuciones que presenten unidades de medidas diferentes.

6.2 Medidas de dispersión relativas

Las medidas de dispersión estudiadas anteriormente son absolutas, por lo tanto, no son susceptibles de hacer comparaciones entre distribuciones con unidades diferentes. Así, una desviación típica de dos libras al medir el peso de una persona no se puede comparar con una desviación típica de dos pulgadas al medir la estatura de la misma persona.

Una medida de dispersión relativa que más frecuentemente se usa es el Coeficiente de Variación.

El coeficiente de variación es una medida de dispersión relativa que se define como el cociente que se obtiene al dividir la desviación típica ó estándar de un conjunto de datos entre su correspondiente media aritmética. Indica el grado de variabilidad o dispersión entre los valores de la variable y la media aritmética.

Esta es una medida adimensional, un valor abstracto, que no se expresa en ninguna unidad de medida, razón por la cual nos permite hacer comparaciones entre distribuciones con unidades diferentes.

Si tenemos dos distribuciones, A y B, y calculamos en ambas su coeficiente de variación; aquella que presente el coeficiente de variación más elevado presenta mayor dispersión de los valores con relación a la media aritmética y viceversa.

La simbología que identifica al coeficiente de variación de un conjunto de datos es,

CV= coeficiente de variación,

cuya fórmula es,

$$CV = \frac{S}{\bar{x}}(100) \qquad 6.12$$

Ejemplo 6.7

Con los datos del ejemplo 6.6, calcule el coeficiente de variación.

Solución:

Como se trata de un ejemplo donde ya calculamos su media aritmética y su desviación estándar, tendremos que,

$$CV = \frac{6.65}{14.19}(100) = 46.96\%$$

El resultado anterior (CV=46.96%) nos señala que existe una muy elevada dispersión de los valores de la variable alrededor de la media aritmética.

Ejemplo 6.8

Las edades de 54 niños de la educación primaria de una escuela de Santo Domingo se presenta a continuación. Calcule el coeficiente de variación de las edades de los niños.

Tabla 6.6 Edad (en años) de 54 niños de la educación básica

Edad (años)	No. de niños (fi)	x_i	$X_i fi$	$(x_i-\bar{X})$	$(x_i-\bar{X})^2$	$(x_i-\bar{X})^2 fi$
4-6	8	5	40	-3.93	15.44	123.52
6-8	10	7	70	-1.93	3.72	37.20
8-10	18	9	162	0.07	0.005	0.09
10-12	12	11	132	2.07	4.28	51.36
12-14	6	13	78	4.07	16.56	99.36
Total	**54**		**482**			**311.53**

Solución:

El coeficiente de variación es igual a,

$$CV = \frac{S}{\bar{x}} \quad (100)$$

donde,

$$\bar{x} = \frac{482}{54} = 8.93 \text{ años.}$$

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2 fi}{n-1}}, \text{ sustituyendo tenemos que,}$$

$$S = \sqrt{\frac{311.53}{54-1}} = 2.42 \text{ años.}$$

Sustituyendo en la fórmula 6.12 tendremos que,

$$C.V. = \frac{2.42}{8.93}(100) = 27.09\%.$$

$$CV = 27.09\%$$

Ejemplo 6.9

Calcule el coeficiente de variación de la distribución de la producción por día de leche (litros) de 20 vacas de una finca experimental.

Tabla 6.7 Producción por día de leche (litros) de 20 vacas.

Producción (litros/día)	No. de vacas	x_i	$x_i f_i$	$(x_i-\bar{x})$	$(x_i-\bar{x})^2$	$(x_i-\bar{x})^2 f_i$
20-22	1	21	21	-5.7	32.49	32.49
22-24	2	23	46	-3.7	13.69	27.38
24-26	4	25	100	-1.7	2.89	11.56
26-28	8	27	216	0.3	0.09	0.72
28-30	3	29	87	2.3	5.29	15.87
30-32	1	31	31	4.3	18.49	18.49
32-34	1	33	33	6.3	39.69	39.69
Total	**20**		**534**			**146.20**

Solución:

El coeficiente de variación es igual a,

$$CV = \frac{S}{\bar{x}} \quad (100),$$

donde,

$$\bar{x} = \frac{534}{20} = 26.7 \text{ litros}$$

$$S = \sqrt{\frac{\sum(x_i-\bar{x})^2 f_i}{n}}, \text{ sustituyendo tenemos que,}$$

$$S = \sqrt{\frac{146.20}{20}} = 2.77 \text{ litros}$$

sustituyendo en la fórmula 6.12, tendremos que:

$$CV = \frac{2.77}{26.7}(100) = 10.37\%.$$

$$CV = 10.37\%$$

EJERCICIOS DE APLICACIÓN

6.1 El salario promedio semanal pagado a los obreros en las zonas francas de la República Dominicana durante los últimos nueve años (1993-2001) se presenta a continuación:

X_i (RD$): 297.16; 496.74; 539.96; 576.92; 634.27; 638.32; 701.21; 725.25; 778.27

Fuente: Consejo Nacional de Zonas Francas

a) Calcule la desviación media de los salarios;

b) Calcule la varianza de los salarios;

c) Calcule la desviación típica o estándar de los salarios.

6.2 Con los datos del ejercicio 5.23 de la página 169, que se refieren al número de pasajeros que abordaron 10 autobuses del transporte público, al salir de su terminal, durante una hora, calcule:

a) La desviación típica o estándar del número de pasajeros que abordaron los autobuses.

6.3 Las ventas mensuales de una tienda de servicios fueron las siguientes:

RD$ 90,000; 92,000; 88,000; 87,500; 93,400; 98,000;

91,400; 85,000; 88,600; 90,000; 89,800; 92,500.

a) Calcule la desviación media de las ventas.

b) Calcule la desviación típica de las ventas mensuales.

6.4 Calcule la desviación media y la desviación típica de las calificaciones de biología de 15 estudiantes.

81, 78, 75, 64, 68, 73, 83, 88, 90, 76, 80, 81, 84, 91, 84.

6.5 Doce clientes se acercaron a las ventanillas de cobro de una empresa de servicio energético a pagar sus facturas.

El tiempo de espera (en minutos) en ser atendidos se presenta a seguidas:

Se pide:

a) Calcular la desviación típica o estándar del tiempo de espera;

b) Calcular el coeficiente de variación del tiempo de espera.

6.6 El valor de las facturas pagadas por los doce clientes del ejercicio 6.5 es:

X_i (RD$): 875; 1,250; 1,725; 975; 780; 1,550; 1,600; 1,320; 1,155; 1,000; 1,375; 1,425

a) Calcule la desviación media del valor de las facturas.

b) Calcule la desviación típica del valor de las facturas.

c) Calcule el coeficiente de variación.

6.7 La importación de gas licuado de petróleo (millones de galones) durante el primer semestre del año 2002 se presenta a continuación:

X_i (millones de galones): 21.16; 20.04; 21.48; 22.61; 20.98; 20.72;

Fuente: Secretaría de Estado de Industria y Comercio.

a) Encuentre la desviación media.

b) Encuentre la varianza.

c) Encuentre la desviación típica.

d) Encuentre el coeficiente de variación.

6.8 Determine la desviación típica del peso al nacer (en libras) de cinco niños.

7.5, 6.25, 8.0, 5.75, 8.5.

6.9 La siguiente distribución se refiere a los salarios quincenales (en RD$) de varios empleados de una fábrica

a) Calcule la desviación media de la distribución de los salarios.

b) Calcule la desviación intercuartílica de la distribución de los salarios.

Salarios (RD$)	N° de empleados
1400-1500	2
1500-1600	4
1600-1700	5
1700-1800	8
1800-1900	10
1900-2000	7
2000-2100	4
Total	**40**

6.10 El tiempo de servicio (en años) de los empleados de una oficina del sector público se muestra en la siguiente tabla.

Años en servicio	N° de empleados
0-5	45
5-10	36
10-15	24
15-20	11
20-25	5
25-30	3
30-35	1
Total	**125**

a) Calcule la varianza de los años en servicio.

b) Calcule la desviación típica o estándar.

c) Calcule el coeficiente de variación.

d) En base a los resultados obtenidos haga un breve comentario acerca de la distribución.

6.11 Se pide a un profesional especializado en administración financiera que analice la variabilidad de los precios de tres acciones diferentes para ayudar al gerente de un banco a hacer una inversión.

Las tres compañías están en la misma industria y se han visto afectadas por recientes normativas oficiales. Por tal razón, sólo

las 12 semanas anteriores indicaran la marcha futura. Las acciones han pagado dividendos parecidos hasta el momento, y ese es el criterio principal del gerente. Pero, para evitar especulaciones, el gerente prefiere además, acciones que no fluctúen mucho de precio. A continuación se dan los precios semanales al cierre de las tres acciones en las 12 semanas anteriores.

Acción A		Acción B		Acción C	
90	94	95	113	98	101
100	102	92	97	93	100
108	101	94	106	105	96
105	106	97	94	102	105
98	103	112	98	102	94
95	98	96	106	103	101

a) Calcular la varianza para cada acción.
b) Calcular la desviación típica para cada acción.
c) Calcular el coeficiente de variación para cada acción.
d) En base a los resultados obtenidos, cuál sería la recomendación al gerente?

6.12 Las edades de cien personas que obtuvieron licencia de conducir durante una semana se presentan en la siguiente tabla.

Edad (años)	N° de personas
20-24	12
25-29	16
30-34	32
35-39	22
40-44	12
45-49	5
50-54	1
Total	**100**

a) Calcule la varianza de las edades.
b) Calcule la desviación típica de las edades.
c) Calcule el coeficiente de variación de las edades.

6.13 La temperatura media (grados celsius) y la precipitación pluvial (milímetros) de algunas localidades de República Dominicana durante un año se muestran a continuación.

Localidades	Temperatura (grados celsius)	Precipitación pluvial (milímetro)
Azua	27.5	578.4
Polo	21.4	1300.3
Moca	25.3	1015.0
Constanza	18.1	759.4
La Vega	26.5	927.7
Mao	26.8	499.9

a) Diga cuál de las dos distribuciones (temperatura y precipitación) presenta mayor variabilidad; por qué?

6.14 Calcule el coeficiente de variación con los datos del ejercicio 5.18 de la página 167 que se refiere a los sueldos mensuales (RD$) de 80 empleados del sector público.

6.15 Calcule la varianza y el coeficiente de variación con los datos del ejercicio 5.19 de la página 168 referente a las calificaciones obtenidas por 75 estudiantes universitarios en un curso de matemáticas.

6.16 Calcule el coeficiente de variación con los datos del ejercicio 5.22 de la página 169 referentes al tiempo que requieren los técnicos de la compañía de auditores CP, S.A. para auditar las cuentas de una empresa de servicios.

6.17 El gasto mensual (RD$) de 75 familias de Santo Domingo en la adquisición de la canasta familiar se presenta a continuación:

Gasto Mensual (RD$)	No. de Familias
4,000 – 4,500	3
4,500 – 5,000	6
5,000 – 5,500	8
5,500 – 6,000	10
6,000 – 6,500	11
6,500 – 7,000	15
7,000 – 7,500	12
7,500 – 8,000	6
8,000 – 8,500	4
Total	**75**

a) Calcule la desviación típica o estándar del gasto mensual;
b) Calcule el coeficiente de variación.

6.18 Las estaturas (pulgadas) de una muestra de estudiantes universitarios tomados al azar fueron las siguientes:

Estaturas (Pulgadas)	No. de Estudiantes
56-59	13
59-62	15
62-68	19
65-68	21
68-71	30
17-74	12
74-77	6
Total	**116**

a) Calcule la desviación típica o estándar de la estaturas;
b) Calcule el coeficiente de variación;
c) Compare el resultado del coeficiente de variación obtenido con el coeficiente de variación del ejercicio 6.17 y diga ¿ cuál distribución es más dispersa y por qué?

CAPITULO

VII

Medidas de Asimetría y de Apuntamiento o Curtosis

7.0 Introducción

Como hemos observado en capítulos anteriores, las medidas de tendencia central y las medidas de dispersión son las más importantes medidas que resumen y describen un conjunto de datos numéricos. Sin embargo, pueden no ser suficientes para ciertos fines.

Es muy probable que en dos o más distribuciones de frecuencia se obtengan medias aritméticas iguales y además sus desviaciones estándar sean idénticas y sin embargo las distribuciones pueden ser diferentes.

En el presente capítulo vamos a estudiar las medidas que nos permitirán conocer la forma y dirección de una distribución de frecuencia. Estas son: las medidas de asimetría, que indican la dirección hacia la izquierda o hacia la derecha de una distribución cuando se representa gráficamente.

Asimismo, las medidas de apuntamiento o curtosis, que indican el grado de achatamiento de la curva que se obtiene de la representación gráfica de una distribución.

7.1 Asimetría

Un conjunto de datos obtenidos de la medición de una variable pueden tener variadas formas en su presentación, pero existe una forma que de manera muy precisa describe un conjunto de datos.

Esta es la distribución simétrica, que cuando se representa gráficamente tiene forma de campana, cuya ilustración se presenta en la Figura 7.1.

En la práctica, estas distribuciones no son muy comunes, pero existen razones técnicas para ajustar cualquier distribución a esta forma.

Otras distribuciones, que no se corresponden con las simétricas, aunque tengan forma acampanada, como las que se ilustran en las figuras 7.2 y 7.3, reciben el nombre de Asimétricas.

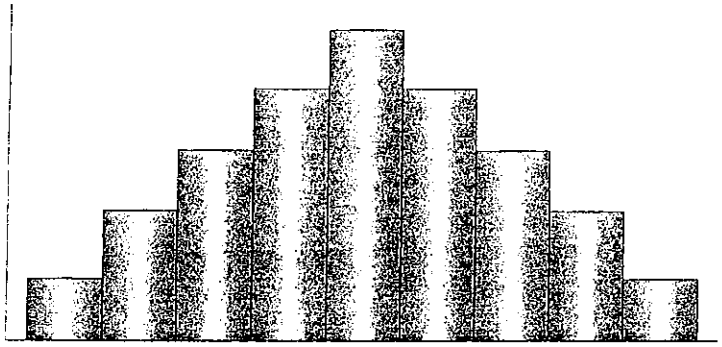
Las distribuciones asimétricas presentan una cola muy pronunciada a la derecha o a la izquierda, o sea indican la dirección de la distribución.

Las que tienen la cola hacia la izquierda son negativamente asimétricas y las que tienen la cola hacia la derecha son positivamente asimétricas.

Por lo regular, las distribuciones de los ingresos familiares de una comunidad son positivamente asimétricas, debido a que casi siempre existen ingresos muy altos entre pocas familias e ingresos bajos y muy bajos entre muchas familias.

Cuando una distribución es simétrica, como la de la figura 7.1, los valores de la media aritmética, la mediana y la moda coinciden.

Figura 7.1

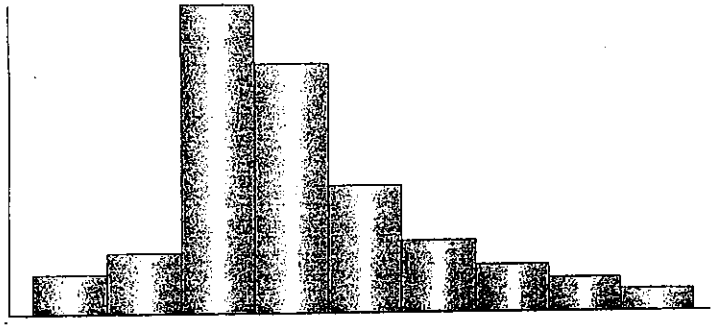


DISTRIBUCION SIMETRICA

Cuando se trata de una distribución asimétrica hacia la derecha o positivamente asimétrica, la media aritmética es mayor que la mediana (ver figura 7.2).

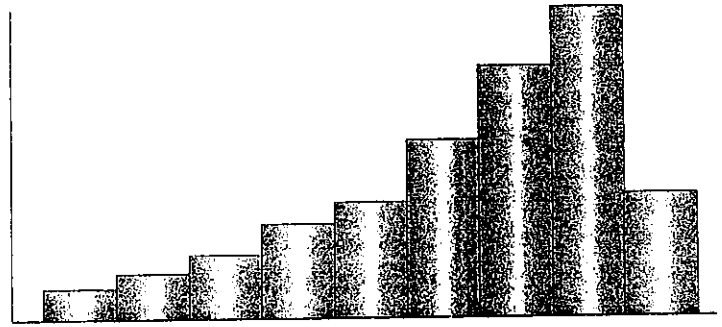
Asimismo, cuando se trata de una distribución asimétrica hacia la izquierda o negativamente asimétrica la mediana es mayor que la media aritmética (ver figura 7.3).

Figura 7.2



DISTRIBUCION POSITIVAMENTE ASIMETRICA

Figura 7.3



DISTRIBUCION NEGATIVAMENTE ASIMETRICA

Existe una medida del grado de asimetría de una distribución de datos, la cual vamos a presentar a continuación.

7.1.1 Coeficiente de asimetría de Pearson

Este coeficiente mide el grado en que una distribución es asimétrica. Su cálculo se basa en la relación existente entre la media aritmética y la mediana. El coeficiente de asimetría de Pearson se denota como, SK, y su fórmula es la siguiente,

$$SK = \frac{3\,(\text{media-mediana})}{\text{desviación estándar}}$$

Sustituyendo por la simbología correspondiente,

$$SK = \frac{3\,(\bar{x} - Me)}{S} \qquad 7.1$$

Esta ecuación expresa que la asimetría de una distribución se puede medir como la diferencia entre la media aritmética y la mediana, multiplicada por tres y dividiendo este resultado entre la desviación estándar.

Cuando una distribución es simétrica, el coeficiente de asimetría de Pearson, SK, es igual a cero.

Cuando SK es negativo, el valor de la mediana es mayor que el valor de la media aritmética y por consiguiente la asimetría es hacia la izquierda.

De manera que, si SK es positivo, el valor de la media aritmética es mayor que el valor de la mediana y por consiguiente, la asimetría es hacia la derecha.

Karl Pearson, también estableció que el coeficiente de asimetría se puede medir a través de la relación entre la media aritmética y la moda. Cuando la diferencia entre la media aritmética y la moda es dividida entre la desviación estándar, el resultado obtenido permite, según Pearson, medir el grado de asimetría en una distribución.

$$SK = \frac{3\,(\bar{x} - Mo)}{S} \qquad 7.2$$

Existe, del mismo modo, otra medida del grado de asimetría de una distribución, denominada, coeficiente de asimetría de momentos. El mismo se denota como $\alpha 3$, así,

$$\alpha_3 = \frac{m_3}{\sqrt{m^3_2}} = \frac{m_3}{S^3} \qquad 7.3$$

donde,

m_3 = momento de orden tres con relación a la media aritmética.

m^3_2 = momento de orden dos al cubo con relación a la media

Más adelante daremos mayor explicación y ejemplos acerca del coeficiente de asimetría de momentos. Ahora, nos limitamos a efectuar ejemplos del coeficiente de asimetría de Pearson, usando la mediana y la moda.

Ejemplo 7.1

Calcule el coeficiente de asimetría de Pearson del tiempo de espera (minutos), de los clientes que se presentaron a realizar alguna transacción a un banco comercial de Santo Domingo.

Tabla 7.1 Tiempo de espera (minutos) de los clientes que visitan un banco de Santo Domingo.

Tiempo (minutos)	No. de clientes (fi)	x_i	$x_i fi$	$(x_i - \bar{X})$	$(x_i - \bar{X})^2$	$(x_i - \bar{X})^2 fi$	Fi
0-5	8	2.5	20	-11.69	136.65	1093.2	8
5-10	12	7.5	90	-6.69	44.75	537	20
10-15	25	12.5	312.5	-1.69	2.85	71.25	45
15-20	20	17.5	350	3.31	10.95	219	65
20-25	10	22.5	225	8.31	69.05	690.50	75
25-30	5	27.5	137.5	13.31	177.15	885.75	80
Total	**80**		**1135**			**3496.70**	

La ecuación 7.1 establece que,

$$SK = \frac{3(\bar{x} - Me)}{S}$$

lo que implica el cálculo de la media aritmética, la mediana y la desviación estándar.

Sea $\bar{x} = \frac{\sum x_i f_i}{S}$

de donde $\bar{x} = \frac{1135}{80} = 14.19$ minutos

Asimismo,

$$Me = Li + \left[\frac{n/2 - Fi\text{-}1}{fi}\right] ci, \text{ donde,}$$

$n/2 = 80/2 = 40$, lo que implica que la clase que contiene la mediana es la clase de (10-15 minutos), sustituyendo en la fórmula,

$$Me = 10 + \left[\frac{40 - 20}{25}\right] 5$$

de donde,

Me = 14 minutos

La desviación estándar es igual a:

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

Sustituyendo por los resultados de la tabla 7.1, la desviación típica será,

$$S = \sqrt{\frac{3496.7}{80}} = 6.61 \text{ minutos}$$

al sustituir en la ecuación 7.1 tenemos que:

$$SK = \frac{3(14.19 - 14)}{6.61} = 0.086$$

SK = 0.086

Tomando como base este resultado, podemos afirmar que esta distribución es ligeramente asimétrica hacia la derecha. Podríamos afirmar también que la distribución es casi simétrica, ya que este valor, SK = 0.086, es muy pequeño.

Igualmente,

$$SK = \frac{\bar{x} - Mo}{S}$$

donde,

$$Mo = Li + \left[\frac{\Delta 1}{\Delta 1 + \Delta 2}\right] ci$$

así,

$$Mo = 10 + \left[\frac{13}{13+5}\right] 5 = 13.61 \text{ minutos}$$

sustituyendo en 7.4

$$SK = \left[\frac{14.19 - 13.61}{6.61}\right] = 0.087$$

La distribución es ligeramente asimétrica hacia la derecha.

Este resultado indica que la media aritmética es más grande que la moda en 8.7% del valor de la desviación estándar.

Ejemplo 7.2

Las inasistencias del personal de una empresa de Zona Franca durante un año se presentan en la siguiente tabla.

Calcula el coeficiente de asimetría de Pearson.

a) mediante la relación entre la media y la mediana.
b) mediante la relación entre la media y la moda.

Tabla 7.2 Inasistencias del personal de una empresa de Zona Franca durante un año.

Número de inasistencias	Nº de empleados (fi)	x_i	$x_i fi$	$(x_i-\bar{x})$	$(x_i-\bar{x})^2$	$(x_i-\bar{x})^2 fi$	Fi
0-4	4	2	8	-7.8	60.84	243.36	4
5-9	10	7	70	-2.8	7.84	78.4	14
10-14	7	12	84	2.2	4.84	33.88	21
15-19	2	17	34	7.2	51.84	103.68	23
20-24	1	22	22	12.2	148.84	148.84	24
25-29	1	27	27	17.2	295.84	295.84	25
Total	**25**		**245**			**904**	

Solución:

a) La ecuación 7.1 establece que,

$$SK = \frac{3(\bar{x} - Me)}{S}$$

lo que implica que el cálculo de la media aritmética, la mediana y la desviación estándar sea:

$$\bar{x} = \frac{\sum x_i fi}{n}$$

de donde $\bar{x} = \frac{245}{25} = 9.8$ inasistencias

(ver cálculos en la tabla 7.2)

De igual modo,

$$Me = Li + \left[\frac{n/2 - Fi\text{-}1}{fi}\right] Ci, \text{ donde,}$$

$n/2 = 25/2 = 12.5$, así, la clase que contiene la mediana es la clase de (5-9 inasistencias). Sustituyendo en la fórmula anterior tendremos,

$$Me = 5 + \left[\frac{12.5 - 4}{10}\right] 5 = 9.25$$

así,

Me = 9.25 inasistencias.

La desviación estándar será,

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

sustituyendo por los resultados de la tabla de cálculo 7.2 tendremos que:

$$S = \sqrt{\frac{904}{25}} = 6.01 \text{ inasistencias.}$$

S = 6.01 inasistencias.

Al sustituir en la ecuación 7.1 tendremos:

$$SK = \frac{3(9.8-9.25)}{6.01} = 0.274$$

SK = 0.274

Esta distribución es asimétrica hacia la derecha, e indica además, que la media aritmética es 27.4% más grande que la mediana con relación al valor de la desviación estándar.

tenemos que,

$$SK = \frac{\bar{x} - Mo}{S} \qquad 7.4$$

donde,

$$Mo = Li + [\frac{\Delta 1}{\Delta 1 + \Delta 2}] Ci$$

así,

$$Mo = 5 + [\frac{6}{6+3}] 5 = 8.33 \text{ inasistencias}$$

Sustituyendo en 7.4 tendremos:

$$SK = [\frac{9.8 - 8.33}{6.01}] = 0.244$$

$$SK = 0.244$$

Como se observa, la distribución es asimétrica hacia la derecha, pero en este caso la media aritmética es 24.4% más grande que la moda con relación al valor de la desviación estándar.

7.2 Apuntamiento o curtosis

En la descripción de una distribución de frecuencias hemos señalado las medidas de tendencia central y las medidas de dispersión, además en este capítulo, hemos tratado la asimetría como medida para indicar la dirección de la distribución, ya sea hacia los valores más altos (a la derecha) o hacia los valores más pequeños (a la izquierda).

Ahora iniciaremos el estudio de otra medida que nos permite describir una distribución de frecuencias; nos referimos a la medida de apuntamiento o curtosis; que permite conocer el grado de concentración de los valores de una distribución de

achatamiento y/o apuntamiento de la gráfica de una distribución.

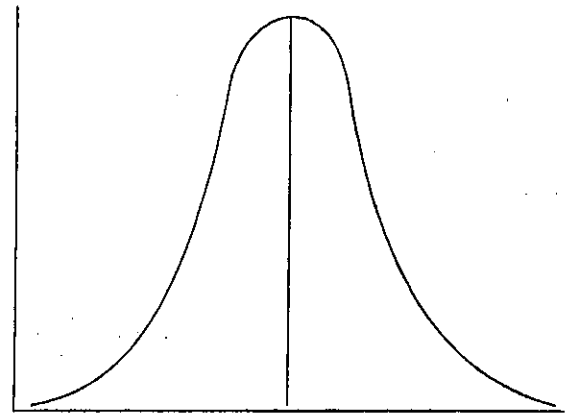
Cuando existe una alta concentración de los valores de la distribución alrededor de la media, gráficamente tendremos una curva con muy alto apuntamiento, la cual se reconoce con el nombre de Leptocúrtica. Véase la figura 7.4.

Cuando existe mucha dispersión de los valores de la distribución de frecuencias alrededor de la media aritmética, la curva resultante es achatada y recibe el nombre de Platicúrtica, como se ilustra en la figura 7.5.

Asimismo, cuando los valores de una distribución de frecuencias se concentran de forma normal, alrededor de la media aritmética, la curva resultante presenta un apuntamiento intermedio y recibe el nombre de Mesocúrtica, como se ilustra en la figura 7.6.

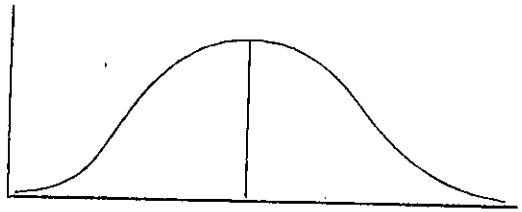
Es importante señalar que esta curva, la mesocúrtica, es la utilizada como referencia o patrón para comparar la curtosis de otras curvas, ya que esta es la curva normal, fruto de la representación gráfica de una distribución simétrica.

Figura 7.4



CURVA CON ALTO APUNTAMIENTO O LEPTOCURTICA

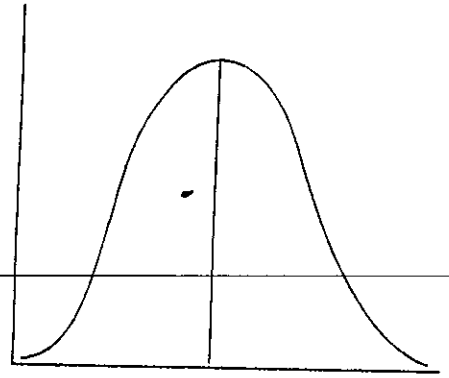
Figura 7.5



MEDIA

CURVA ACHATADA O PLATICURTICA

Figura 7.6



MEDIA

CURVA NORMAL O MESOCURTICA

Antes de explicar cómo medir estadísticamente la curtosis, vamos a señalar que el origen de los términos usados para nombrar los tipos de curvas son:

leptocúrtica (lepto, prefijo que significa empinado o estrecho),

platicúrtica (plati, prefijo que significa plano o ancho) y,

mesocúrtica (donde meso significa intermedio).

La mejor forma de medir la curtosis es a través del **coeficiente de curtosis**, que se denota como:

$$\alpha_4 = \frac{m_4}{m^2_2} \qquad 7.5$$

y mide el grado de apuntamiento de una curva, donde,

m_4 : es igual al momento de orden cuatro respecto de la media aritmética y,

m^2_2: es igual al momento de orden dos al cuadrado respecto de la media aritmética.

Nos referiremos brevemente a los momentos, en virtud de que necesitamos emplearlos para el cálculo de las medidas de curtosis y asimetría.

7.2.1 Momentos

Los momentos permiten obtener medidas resúmenes de conjunto de datos que por contener información variable resulta difícil de captar en sus características esenciales.

Los momentos pueden calcularse con respecto al origen de la distribución y se refieren de manera directa a la variable.

Si $x_1; x_2, x_3, \ldots x_n$ son los valores que toma la variable equis, el momento r de la variable se indica como:

$$x^r = \frac{x_1^r + x_2^r + x_3^r + \ldots + x^r}{n} = \frac{\sum x_i^r}{n}$$

El primer momento tiene como valor r = 1 y puede observarse que es igual a la media aritmética.

El segundo momento tiene como valor r=2 y de manera similar se pueden calcular otros momentos con sólo elevar el valor de x a la potencia correspondiente.

Cuando los momentos son calculados tomando como origen la media aritmética, obtenemos los momentos respecto a la media aritmética de la distribución, que son los que efectivamente vamos a usar en este capítulo.

$$m_r = \frac{\sum(x_i - \bar{x})^r}{n}$$, cuando se trata de datos no agrupados.

Si $r = 1$, entonces $m_1 = 0$

cuando $r = 2$, entonces,

$$m_2 = \frac{\sum(x_i - \bar{x})^2}{n} = S^2$$

Para datos agrupados en una distribución de frecuencias, los momentos respecto a la media aritmética se indican como:

$$m_r = \frac{\sum(x_i - \bar{x})^r fi}{n}$$

De esta manera;

$$m_1 = \frac{\sum(x_i - \bar{x})^1 fi}{n} = 0 \qquad 7.6$$

$$m_2 = \frac{\sum(x_i - \bar{x})^2 fi}{n} = S^2 \qquad 7.7$$

$$m_3 = \frac{\sum(x_i - \bar{x})^3 fi}{n} \qquad 7.8$$

$$m_4 = \frac{\sum(x_i - \bar{x})^4 fi}{n} \qquad 7.9$$

El grado de curtosis, como observamos en la fórmula 7.5, se obtiene al dividir el cuarto momento respecto a la media aritmética entre el momento de orden dos al cuadrado.

Pero, el $M_2 = (S)^2$; que es lo mismo que S^4 (desviación estándar a la cuarta potencia), entonces, el coeficiente de curtosis,

$$\alpha_4 = \frac{m_4}{S^4} \qquad 7.10$$

El valor del coeficiente de curtosis, a4, para una distribución simétrica es igual a tres. De esta manera, el coeficiente de curtosis indica el apuntamiento de la curva mediante las siguientes desigualdades.

Si $\alpha_4 - 3 = 0$ Indica que la distribución es simétrica o mesocúrtica

Si $\alpha_4 - 3 > 0$ En la medida en que la diferencia se hace mayor, por encima de cero, la curva tendrá más apuntamiento; o sea, la distribución es leptocúrtica.

Si $\alpha_4 - 3 < 0$ En la medida en que la diferencia se hace menor, por debajo de cero, más achatada será la curva; o sea, la distribución es platicúrtica.

Para ilustrar mejor a nuestros lectores, presentamos a continuación algunos ejemplos:

Ejemplo 7.3

Calcule el coeficiente de curtosis con los datos de la tabla 7.2 del ejemplo 7.2 de la página 212.

Tabla 7.3 Inasistencia del personal de una empresa de Zona Franca durante un año

N° de Inasistencias	N° de empleados	x_i	$x_i fi$	$(x_i-\bar{x})$	$(x_i-\bar{x})^2 fi$	$(xi-\bar{x})^3 fi$	$(xi-\bar{x})^4 fi$
0-4	4	2	8	-7.8	243.36	-1898.21	14806.02
5-9	10	7	70	-2.8	78.4	-219.52	614.66
10-14	7	12	84	2.2	33.88	74.54	163.98
15-19	2	17	34	7.2	103.68	746.50	5374.77
20-24	1	22	22	12.2	148.84	1815.85	22153.35
25-29	1	27	27	17.2	295.84	5088.45	87521.31
Total	**25**		**245**		**904**	**5607.61**	**130634.09**

Solución:

La ecuación 7.10 establece que,

$$\alpha_4 = \frac{m_4}{S_4}$$

donde,

$$m_4 = \frac{\sum(x_i-\bar{x})^4 fi}{n} = \frac{130634.09}{25} = 5225.36$$

y

$$S^4 = m^2_2 = \left[\frac{\sum (x_i-\bar{x})^2 fi}{n}\right] = \left(\frac{904}{25}\right) = 1307.55$$

así,

$$\alpha_4 = \frac{5225.36}{1307.55} = 3.99$$

Puesto que el valor de $\alpha_4 = 3.99$, la diferencia entre $(\alpha_4-3) = (3.99-3) = 0.99 > 0$, la distribución de las inasistencias del personal de la empresa de Zona Franca es leptocúrtica, o sea,

existe una alta concentración de los valores de la distribución alrededor de la media aritmética.

Ejemplo 7.4

Con los datos de la tabla 7.3 calcule el coeficiente de asimetría, por momentos.

Solución

Como sabemos, el coeficiente de asimetría a través de los momentos respecto de la media aritmética es, según la ecuación 7.3, igual a :

$$\alpha_3 = \frac{m_3}{S^3}$$

así,

$$m_3 = \frac{\sum(x_i-\bar{x})^3 fi}{n} = \frac{5607.61}{25} = 224.29$$

y,

$$S^3 = \sqrt{m^3_2} = \left[\sqrt{\frac{\Sigma (x_i-\bar{x})^2 fi}{n}}\right]^3 = \frac{904}{25} = (6.01)^3 = 217.4$$

entonces,

$$\alpha_3 = \frac{224.29}{217.08} = 1.03$$

El coeficiente de asimetría por momentos, $\alpha_3 = 1.03$, indica que la asimetría es hacia la derecha.

Cabe señalar que el coeficiente de asimetría que se obtiene mediante momentos es una medida absoluta de asimetría.

Ejemplo 7.5

Calcule el coeficiente de curtosis del peso (libras) de varios niños al nacer.

Tabla 7.4. Peso al nacer (en libras) de varios niños

Peso (libras)	N° de Niños	x_i	$x_i fi$	$(x_i-\bar{x})$	$(x_i-\bar{x})^2$ fi	$(x_i-\bar{x})^4$ fi
3.5-4.5	1	4	4	-2.95	8.70	75.73
4.5-5.5	2	5	10	-1.95	7.60	28.92
5.5-6.5	4	6	24	-0.95	3.61	3.26
6.5-7.5	7	7	49	0.05	0.017	0.00004
7.5-8.5	3	8	24	1.05	3.31	3.64
8.5-9.5	2	9	18	2.05	8.40	35.32
9.5-10.5	1	10	10	3.05	9.30	86.53
Total	**20**		**139**		**40.94**	**253.40**

Solución:

Sea $\alpha_4 = \dfrac{m_4}{S^4}$

donde

$$m_4 = \frac{\sum(x_i-\bar{x})^2 fi}{n} = \frac{253.40}{20} = 12.67$$

y,

$$S^4 = m^2_2 = \left[\frac{\Sigma (x_i-\bar{x})^2 fi}{n}\right]^2 = \left[\frac{40.94}{20}\right]^2 = (2.047)^2 = 4.19$$

así,

$$\alpha_4 = \frac{12.67}{4.19} = 3.02$$

Este resultado indica que la distribución del peso de los 20 niños al nacer es prácticamente normal, ya que,
$(\alpha_4 - 3) = (3.02-3) = 0.02$.

EJERCICIOS DE APLICACION

7.1 Calcular el coeficiente de asimetría de Pearson para las siguientes series de datos.

a) 2, 6, 19, 8, 14, 21, 15, 16, 15, 8, 17, 5

b) 110, 125, 115, 113, 111, 112, 100, 130, 108

c) 80, 85, 84, 80, 86, 90, 91, 80

d) 44, 52, 58, 60, 49, 52, 48, 57, 61, 55

7.2 Con los datos del ejercicio 6.5 de la página 199, que se refieren al tiempo de espera (en minutos) que permanecieron doce clientes que se acercaron a las ventanillas de cobro de una empresa de servicio energético a pagar sus facturas, calcule:

a) El coeficiente de asimetría de Pearson.

b) El coeficiente de curtosis.

c) Comente ambos resultados.

7.3 Con los datos del ejercicio 6.10 de la página 201, referentes al tiempo de servicio (en años) de los empleados de una oficina del sector público, calcule:

a) El coeficiente de asimetría.

b) El coeficiente de curtosis.

c) Haga un comentario de los resultados obtenidos.

7.4 Una compañía constructora de proyectos habitacionales ejecutó la construcción de un lote de 75 viviendas distribuidas según el tamaño de la siguiente forma.

Tamaño (Mts.²)	N° de viviendas
80-90	8
90-100	10
100-110	12
110-120	16
120-130	15
130-140	9
140-150	5
Total	**75**

a) Calcule el coeficiente de asimetría de Pearson.

b) Calcule el coeficiente de curtosis.

7.5 Las ventas (en miles de RD$) realizadas por una tienda de efectos electrodomésticos durante el mes de diciembre del 2005 se presentan en la siguiente tabla.

Ventas (miles de RD$)	Nº de días
100-125	2
125-150	4
150-175	5
175-200	9
200-225	5
225-250	3
250-275	2
275-300	1
Total	**31**

a) Calcule el coeficiente de asimetría de Pearson.

b) Calcule el coeficiente de curtosis.

CAPITULO

VIII

Números Índices

8.0 Introducción

En muchas circunstancias, tanto los gobiernos como los gerentes de empresas privadas, sean estas financieras, comerciales e industriales, requieren de indicadores estadísticos especializados para tomar decisiones acertadas.

Frecuentemente, los gerentes de estas empresas necesitan hacer comparaciones en sus sistemas de informaciones con situaciones del pasado para visualizar los cambios que se presentan y así poder tomar decisiones más confiables y oportunas.

Por lo general, estas comparaciones deben describir todos los aspectos de sus empresas, ya que cada aspecto operacional produce efectos sobre el funcionamiento general de la organización, y una variación en cualquiera de estos aspectos da como resultado una variación en la marcha general de la empresa.

Es probable que el gerente de una empresa industrial, como consecuencia de una variación en la tasa de cambio oficial del dólar, se interese en comparar el costo de la producción por unidad del presente mes, con el costo de la producción por unidad de los últimos ocho meses.

Una forma muy importante de resumir el grado de cambio que experimentan las variables económicas durante cierto período de tiempo lo constituyen los **Números índices.**

Los números índices indican el cambio relativo en el precio, cantidad o valor experimentado en algún momento

determinado anterior y el período actual. Al momento anterior le vamos a llamar período base o año base.

En sentido general, los números índices sirven para describir las variables económicas en un momento determinado.

Un número índice mide cuánto cambia una variable económica en función del tiempo. Lo calculamos determinando el cociente o razón del valor actual entre el valor del año base; luego multiplicamos este resultado por cien para expresar el número índice como un relativo porcentual.

El número índice para el año base siempre será igual al 100%.

Por ejemplo, supóngase que se nos informa, a través de un medio especializado de publicación oficial, que el índice del costo de la vida es ahora de 130% con relación al año 1994 (año base). Este resultado señala rápidamente que el costo de la vida ha aumentado en un 30% desde el año 1994.

En sentido general este índice describe la variación de los diferentes factores económicos que han incidido en el incremento del costo de la vida durante ese período de tiempo.

Cuando una ama de casa observa que una libra de carne le cuesta el doble de lo que le costaba hace ocho años, en efecto está utilizando un tipo de número índice.

Cuando el número índice está referido única y exclusivamente a un solo producto o mercancía, el índice se llama **Índice simple**; asimismo, cuando la comparación se refiere a un grupo de productos o mercancías, el índice se denominará **Índice compuesto**. Por ejemplo, si el ama de casa, además de observar el precio de una libra de carne, compara también los precios del arroz, el aceite, los huevos, la leche, las habichuelas está utilizando un índice compuesto.

Es posible que algunos de estos artículos hayan registrado aumentos de consideración en el precio; otros, a que hayan experimentado cambios moderados en el precio e incluso, que otros hayan reducido el precio. Es por esta razón que el objetivo de utilizar un índice compuesto es el de resumir los cambios

totales del precio, en relación a este conjunto de productos o artículos comestibles.

Por las razones antes expuestas, las compras de estos artículos pueden variar al paso de los años. Quizás el consumo de leche ha aumentado, a lo mejor por un incremento en el número de miembros de la familia. Por otra parte, el consumo de arroz y de aceite puede haber disminuido, como consecuencia de un estricto plan dietético en la familia.

De ahí, la necesidad de incluir cambios en la cantidad, así como en el precio, para de esa forma obtener una idea más clara y exacta del cambio total que se experimenta.

Es importante señalar que el empleo de números índices para resumir los cambios económicos es de vital importancia en cualquier economía, independientemente de su estructura social y política.

En economía y administración existen tres clasificaciones de números índices que son: Índices de precio, de cantidad y de valor.

En esta época de economía globalizada, revisten gran interés para todos los países del mundo, los diferentes índices que regularmente se producen y que sirven como fuente básica a sus inversionistas. El índice de producción industrial, el índice del mercado de valores y el índice de precios al consumidor; son valiosas informaciones que día a día son demandadas por estudiantes, profesionales, empresarios e inversionistas.

8.1. Números índices simples

Los números índices simples miden el cambio relativo de un solo artículo o variable económica que se presenta entre dos períodos de tiempo diferentes. El procedimiento para calcular un índice simple de precio, cantidad o valor de un período dado; al precio, cantidad o valor correspondientes de un período o año base; consiste en dividir cada precio, cantidad o valor de un período dado entre el precio, cantidad o valor correspondientes de un año base establecido.

A los números índices simples de precio, cantidad y valor también se les nombra como: precio relativo, cantidad relativa y valor relativo.

8.1.1 Indice de precios

El índice de precios es el de mayor uso y sirve para comparar los cambios en el precio entre dos períodos. El índice de precios al consumidor mide los cambios globales de una gran cantidad de bienes de consumo y de servicios y se utiliza para definir el costo de la vida.

Para los fines de identificar el índice de precios usamos la siguiente simbología: Ip (índice de precios), cuya fórmula es:

$$Ip = \frac{P_n}{P_o}(100) \qquad 8.1$$

donde,

P_n = precio de un artículo en un determinado año o momento.

P_o = precio de un artículo en el año base.

8.1.2 Indice de cantidad

Un índice de cantidad mide cuánto cambia con el tiempo el número o cantidad de una variable.

Vamos a identificar el índice de cantidad como: Iq (índice de cantidad), cuya fórmula es:

$$Iq = \frac{q_n}{q_o}(100) \qquad 8.2$$

donde,

q_n = cantidad consumida de un artículo en un determinado año o momento.

q_o = cantidad consumida de un artículo en el año base.

8.1.3 Índice de valor

Mide los cambios en el valor monetario de una variable. Este índice combina los cambios de precio y cantidad.

El índice de valor lo vamos a identificar con la siguiente simbología: IV (índice de valor) y para los fines de cálculo utilizaremos la siguiente fórmula:

$$Iv = \frac{P_n q_n}{P_o q_o}(100) \qquad 8.3$$

Ejemplo 8.1

Supongamos que los precios unitarios y las cantidades de queso producidas en una región de la República Dominicana en los años 1993 y 1997, se presentan en la siguiente tabla. Calcular los índices de precio, cantidad y valor para el año 1997, tomando como base el año 1993.

Tabla 8.1. Precios y cantidades de queso producidas en 1993 y 1997

Año	Precio por libra (P)	Cantidad vendida (q)	Valor (Pq)
1993	RD$ 18.00	500 libras	RD$ 9,000.00
1997	RD$ 23.00	600 libras	RD$ 3,800.00

Solución

Como ya hemos establecido, el año 1993 es el año base. Esto significa que consideraremos el precio de 1993 de RD$18.00 como equivalente al 100%. De igual forma ocurre con las cantidades vendidas en 1993 de 500 libras que serán consideradas como el 100%; asimismo ocurre con el valor o ingreso del mismo año, que será considerado como el 100%.

a) Precio relativo $= Ip = \frac{P_n}{P_o}(100)$

donde $P_n = P_{1997} = RD\$\,23.00$

y $P_o = P_{1993} = RD\$\,18.00$

$$Ip = \frac{23.00}{18.00}(100) = 128\%$$

Este resultado significa que el precio del queso se incrementó en un 28% entre 1993 y 1997.

b) Cantidad relativa $= Iq = \frac{q_n}{q_o}(100)$

donde, $q_n = q_{1997} = 600$ libras

y $q_o = q_{1993} = 500$ libras

Sustituyente en la fórmula 8.2, se tiene que,

$$Iq = \frac{600}{500}(100) = 120\%$$

Este resultado indica que la cantidad de queso vendida entre 1993 y 1997 se incrementó en un 20%.

c) Valor relativo $= Iv = \frac{P_n q_n}{P_o q_o}(100)$

donde $P_n q_n = (P_{1997})(q_{1997}) = (RD\$\ 23.00)(600\ libras)$

y $P_o q_o = (P_{1993})\, q_{1993}) = (RD\$\ 18.00)(500\ libras)$

sustituyendo en la fórmula 8.3 se tiene que:

$$Iv = \frac{(23)(600)}{(18)(500)}(100) = 153\%$$

Este resultado expresa que el valor de las ventas o ingresos entre 1993 y 1997 se incrementó en un 53%.

Ejemplo 8.2

Los datos que se presentan en la siguiente tabla muestran la cantidad y el precio de los autos nuevos vendidos por un distribuidor durante el período 1992-1997 en Santo Domingo.

Cantidad y precio de autos vendidos por un distribuidor

Año	Precio prom. de venta (Miles de RD$)	Autos vendidos	Valor (Miles de RD$)
1992	180	70	12,600
1993	195	78	15,210
1994	225	86	19,350
1995	280	88	24,640
1996	325	95	30,875
1997	450	99	44,550

Determina los números índices simples de: a) precio, b) cantidad y c) valor para cada año, tomando como base el año 1992; d) construya una nueva tabla con los resultados obtenidos.

a) Indice simple de precio o precio relativo

$$\textbf{Precio relativo} = Ip_{1993} = \frac{P_{1993}}{P_{1992}}(100)$$

donde, $P_{1993} = 195$

y, $P_{1992} = 180$

sustituyendo se tiene,

$$Ip_{1993} = \frac{195}{180}(100) = 108\%$$

$$\textbf{Precio relativo}_{1994} = \frac{P_{1994}}{P_{1992}}(100)$$

donde, $P_{1994} = 225$

y, $P_{1992} = 180$

sustituyendo se tiene,

$$Ip_{1994} = \frac{225}{180}(100) = 125\%$$

$$\text{Precio relativo}_{1995} = \frac{P_{1995}}{P_{1992}} (100)$$

donde, $P_{1995} = 280$

y, $P_{1992} = 180$

sustituyendo se tiene,

$$Ip_{1995} = \frac{280}{180} (100) = 155\%$$

$$\text{Precio relativo}_{1996} = \frac{P_{1996}}{P_{1992}} (100)$$

donde, $P_{1996} = 325$

y, $P_{1992} = 180$

sustituyendo se tiene,

$$Ip_{1996} = \frac{325}{180} (100) = 180\%$$

$$\text{Precio relativo}_{1997} = \frac{P_{1997}}{P_{1992}} (100)$$

donde, $P_{1997} = 450$

y, $P_{1992} = 180$

sustituyendo se tiene,

$$Ip_{1997} = \frac{450}{180} (100) = 250\%$$

b) Indice simple de cantidad o cantidad relativa

$$\text{Cantidad relativa}_{1993} = \frac{q_{1993}}{q_{1992}} (100)$$

donde, $q_{1993} = 78$

y, $q_{1992} = 70$

sustituyendo se tiene,

$$Iq_{1993} = \frac{78}{70} (100) = 111\%$$

$$\text{Cantidad relativa}_{1994} = \frac{q_{1994}}{q_{1992}} (100)$$

donde, $q_{1994} = 86$

y, $q_{1992} = 70$

sustituyendo se tiene,

$$Iq_{1994} = \frac{86}{70} (100) = 123\text{^}\%$$

$$\text{Cantidad relativa}_{1995} = \frac{q_{1995}}{q_{1992}} (100)$$

donde, $q_{1995} = 88$

y, $q_{1992} = 70$

sustituyendo se tiene,

$$Iq_{1995} = \frac{88}{70} (100) = 126\%$$

$$\text{Cantidad relativa}_{1996} = \frac{q_{1996}}{q_{1992}} (100)$$

donde, $q_{1996} = 95$

y, $q_{1992} = 70$

sustituyendo se tiene,

$$Iq_{1996} = \frac{95}{70} (100) = 136\%$$

$$\text{Cantidad relativa}_{1997} = \frac{q_{1997}}{q_{1992}} (100)$$

donde, $q_{1997} = 99$

y, $q_{1992} = 70$

sustituyendo se tiene,

$$Iq_{1997} = \frac{99}{70} (100) = 141\%$$

c) Indice simple de valor o valor relativo

$$\textbf{Valor relativo}_{1993} = \frac{P_{1993}\, q_{1993}}{P_{1992}\, q_{1992}} (100)$$

donde, $P_{1993}\, q_{1993} = (195)(78) = 15210$

y, $P_{1992}\, q_{1992} = (180)(70) = 12600$

sustituyendo se tiene,

$$Iv_{1992} = \frac{15210}{12600} (100) = 121\%$$

$$\textbf{Valor relativo}_{1994} = \frac{P_{1994}\, q_{1994}}{P_{1992}\, q_{1992}} (100)$$

donde, $P_{1994}\, q_{1994} = (225)(86) = 19350$

y, $P_{1992}\, q_{1992} = (180)(70) = 12600$

sustituyendo se tiene,

$$Iv_{1994} = \frac{19350}{12600} (100) = 154\%$$

$$\text{Valor relativo}_{1995} = \frac{P_{1995}\, q_{1995}}{P_{1992}\, q_{1992}} (100)$$

donde, $P_{1995}\, q_{1995} = (280)(88) = 24640$

y, $P_{1992}\, q_{1992} = (180)(70) = 12600$

sustituyendo se tiene,

$$Iv_{1995} = \frac{24640}{12600} (100) = 196\%$$

$$\textbf{Valor relativo}_{1996} = \frac{P_{1996}\, q_{1996}}{P_{1992}\, q_{1992}} (100)$$

donde, $P_{1996}\, q_{1996} = (325)(95) = 30875$

y, $P_{1992}\, q_{1992} = (180)(70) = 12600$

sustituyendo se tiene

$$Iv_{1996} = \frac{30875}{12600} (100) = 245\%$$

$$\textbf{Valor relativo}_{1997} = \frac{P_{1997}\, q_{1997}}{P_{1992}\, q_{1992}} (100)$$

donde, $P_{1997}\, q_{1997} = (450)(99) = 44550$

y, $P_{1992}\, q_{1992} = (180)(70) = 12600$

sustituyendo se tiene,

$$Iv_{1997} = \frac{44550}{12600} (100) = 353\%$$

d) A continuación la nueva tabla con los resultados obtenidos:

Tabla 8.3 Índices de precio, cantidad y valor para el ejemplo 8.2, sobre autos vendidos en Sto. Dgo. usando los datos del año 1992 como año base.

Año	Precio		Cantidad		Valor	
	Miles de RD$	Índice	Unidades	Índice	Miles de RD$	Índice
1992	180	100	70	100	12600	100
1993	195	108	78	111	15210	121
1994	225	125	86	123	19350	154
1995	280	155	88	126	24640	196
1996	325	180	95	136	30875	245
1997	450	250	99	141	44550	353

Los números índices simples que utilizan un año base común reciben el nombre de relativos de base fija.

Existe otro número índice simple, llamado relativo de enlace, que está basado en los cambios anuales. Cada precio, cantidad o valor anual se mide como una razón con relación al año anterior.

Los relativos de enlace se pueden calcular con los datos no procesados o también a partir de los relativos de base fija. Estos se expresan en forma de porcentajes. Por ejemplo, calcule el relativo de enlace para el precio, de los autos de 1995 de la tabla 8.3.

Usando los datos no procesados

$$\text{Relativo de enlace de precio}_{1995} = \frac{P_{1995}}{P_{1994}}(100)$$

$$\text{Relativo de enlace de precio}_{1995} = \frac{280}{225}(100) = 124\%$$

Usando los índices simples de 1995 y 1994 tendremos,

$$\text{Relativo de enlace de precio}_{1995} = \frac{Ip_{1995}}{Ip_{1994}}(100)$$

$$\text{Relativo de enlace de precio}_{1995} = \frac{155}{125}(100) = 124\%$$

Asimismo, pueden calcularse los relativos enlazados para la cantidad y el valor para otros años.

La principal limitación que presentan los números índices simples consiste en que sólo se refieren a un elemento, no obstante presentarse con frecuencia la necesidad de observar cambios respecto de un grupo de elementos.

8.2 Números índices compuestos

Los números índices compuestos se usan para señalar el cambio relativo en precio, cantidad o valor de un conjunto de

elementos o mercancías. Usted podría preguntarse, por ejemplo, si en sentido general, los precios de los artículos comestibles básicos para el consumo humano se han incrementado o reducido durante un período de tiempo establecido. Sin ninguna vacilación, usted aseguraría que muchos precios se han elevado, otros se han reducido y probablemente los precios de otros se han mantenido constantes.

Pero, para conocer en términos globales los cambios que se han originado es necesario observar una combinación de elementos o artículos, en vez de considerarlos de manera aislada.

Los números índices compuestos son los de mayor uso práctico, ya que si nos interesa conocer, por ejemplo, los cambios relativos del costo de la vida, no lo haríamos con el precio de un solo artículo, sino que incluiríamos los precios de un grupo de elementos o artículos que determinan el costo de la vida, como son: alimentación, vivienda, transporte, vestuario, etc.

Los gerentes empresariales se apoyan mucho en los números índices compuestos como forma de comprender mejor otras informaciones; es así como, el gran empleo del índice de precios al consumidor para determinar el real poder adquisitivo del dinero es un ejemplo más de cómo estos números nos permiten conocer mejor otros factores de interés.

A seguidas serán considerados dos clasificaciones de números índices compuestos, que son:

1.-Índice de agregados no ponderados

2.-Índice de agregados ponderados

8.2.1 Índice de agregados no ponderados

El índice de agregados no ponderados constituye el método de cálculo más sencillo de un índice compuesto.

Este método implica que todos los valores incluidos tienen igual importancia (no ponderados) y además, que todos esos valores han de ser sumados (agregados).

todos los elementos del compuesto durante cierto período dado y dividirlo entre la suma de los mismos elementos, durante el año base.

La siguiente fórmula expresa el método para calcular el índice de agregados de precios no ponderados.

$$\text{Agregados de precios no ponderados} = \frac{\sum P_n}{\sum P_o}(100) \quad 8.4$$

donde;

P_n = precio de cada elemento en el grupo durante el año dado.

P_o = precio de cada elemento en el grupo durante el año base.

En los casos del cálculo de agregados no ponderado de cantidad y valor, las respectivas fórmulas se presentan a continuación.

$$\text{Agregados de cantidad no ponderados} = \frac{\sum q_n}{\sum q_o}(100) \quad 8.5$$

donde;

q_n = cantidad de cada elemento en el grupo durante el año dado

q_o = cantidad de cada elemento en el grupo durante el año base

$$\text{Agregados de valor no ponderados} = \frac{\sum V_n}{\sum V_o}(100) \quad 8.6$$

donde:

V_n = valor de cada elemento en el grupo durante el año dado.

V_o = valor de cada elemento en el grupo durante el año base.

Ejemplo 8.3

Calcule el índice de agregados no ponderados de precios con los datos de la siguiente tabla.

Tabla 8.4. Precios de varios artículos

Artículos	Precios 1990 (Po)	1996 (Pn)
Huevos (1 docena)	$6.00	$10.00
Carne (1 libra)	15.00	20.00
Leche (1 litro)	8.00	10.00
Spaguettis (1 libra)	5.00	8.50
	$\sum P_o = 34.00$	$\sum P_n = 48.50$

Solución:

$$\text{Agregados no ponderado de precios} = \frac{\sum P_{1996}}{\sum P_{1990}}\ 100$$

sustituyendo la fórmula tenemos que:

$$\text{Agregados no ponderado de precios} = \frac{48.50}{34.00}\ 100 = 143\%$$

Los precios de estos artículos aumentaron en 43% durante el período 1990-1996. Es obvio reconocer que este método no es usado cuando se trata de análisis muy importantes, ya que su principal desventaja consiste en que no atribuye mayor importancia o peso al cambio de precios de un artículo de gran uso que a uno de poca demanda. Una familia comprará 200 libras de carne al año, pero sería inusual que comprara dos o más autos en un año. Un cambio importante en los precios de artículos de poco movimiento puede distorsionar por completo estos índices.

Otros ejemplos son igualmente aplicables para ilustrar los índices no ponderados de agregados de cantidad y de valor.

8.2.2 Índice de agregados ponderados

Como hemos señalado, los números índice de agregados no ponderados no incorporan esquemas de ponderación en sus cálculos, todos los valores de las variables son tratados con igual importancia, es decir, que cada variable se trata como si tuviera

igual efecto sobre las demás. Es importante reconocer que algunas variables tienen mayor efecto que otras o son más importantes en el cálculo de un índice.

El problema básico en la medición de cambios de precios en el caso de un conjunto de elementos o artículos es que por lo regular se registran cambios en las cantidades adquiridas, así como en los niveles de precios.

Este concepto puede incorporarse usando un número índice de agregados ponderados.

La ponderación nos permite incluir más información que el simple cambio de precios a través del tiempo. También nos permite mejorar los niveles de precisión de la estimación general del nivel de precios, basado en la muestra. El problema radica en establecer cuánto peso o importancia asignar a cada una de las variables. Este problema será superado al estudiar los dos métodos más usados para ponderar un índice, que son:

1.- Método de Laspeyres
2.- Método de Paasche

8.2.2.1 Método de Laspeyres

Este método consiste en usar las cantidades consumidas durante un período base, al calcular cada número índice, como factor de ponderación.

El método de Laspeyres es la técnica de mayor uso, por requerir medidas de cantidades y precios durante un solo período.

Las siguientes fórmulas, se usan para calcular los índices de precio, cantidad y valor, mediante el método de Laspeyres.

$$\text{Índice de precios de Laspeyres, } IPL = \frac{\sum P_n q_o'}{\sum P_o q_o} (100) \quad 8.7$$

donde;

P_n = precios en el año dado.

P_o = precios en el año base.

q_o = cantidades (vendidas o producidas) en el año base.

En este caso, las ponderaciones serán las cantidades del año base.

$$\text{Índice de cantidades de Laspeyres, } IQL = \frac{\sum q_n P_o}{\sum q_o P_o} (100) \quad 8.8$$

donde;

q_n = cantidades (vendidas o producidas) en el año dado.

q_o = cantidades (vendidas o producidas) en el año base.

P_o = precios en el año base.

En este caso, las ponderaciones serán los precios del año base.

$$\text{Indice de valores de Laspeyres, } IVL = \frac{\sum P_n q_n}{\sum P_0 q_0} (100) \quad 8.9$$

donde;

P_n = precios en el año dado.

P_o = precios en el año base.

q_n = cantidades (vendidas o producidas) en el año dado.

q_o = cantidades (vendidas o producidas) en el año base.

Procedimiento para calcular el índice de precios de Laspeyres

Para calcular el índice de precios de Laspeyres se procede de la siguiente manera:

1. Multiplicar el precio en el período actual por la cantidad vendida o producida del año base para cada elemento del grupo.
2. Obtener la suma del producto de los valores anteriores.

3. Multiplicar el precio del año base por la cantidad vendida o producida del año base para cada elemento del grupo.
4. Sumar los valores obtenidos por los productos de los precios y las cantidades del año base.
5. Dividir la primera suma entre la segunda y multiplicar el resultado por 100, para así convertir dicho resultado en un relativo porcentual.

Para mayor ilustración, vamos a resolver un ejemplo para comprobar la aplicación del método de Laspeyres.

Ejemplo 8.4

Los datos que se presentan en la tabla 8.5 constituyen los artículos consumidos por una familia y sus precios en los años 1990 y 1996. Calcule el índice de precios de Laspeyres.

Tabla 8.5. Artículos consumidos por una familia.

Elementos o artículos	Precio base P_o (1990)	Precio dado P_n (1996)	Cant. Promedio consumida en el año base q_o (1990)	$P_o q_o$	$P_n q_o$
Huevos (docena)	$6.00	$10.00	60 doc.	$600	$360
Carne (libra)	15.00	20.00	250 lib.	5,000	3750
Leche (litros)	8.00	10.00	200 lit.	2,000	1600
Spaghettis(libra)	5.00	8.50	75 lib.	637.5	375
Total				**$8237.5**	**$6085**

Solución

$$\text{Índice de precios de Laspeyres} = \frac{\sum P_n q_o}{\sum P_o q_o} (100)$$

La suma del producto de $P_n q_o = \$8237.5$, que fue el resultado de sumar los siguientes productos:

$[(10.00)(60) + (20.00)(250) + (10.00)(200) + (8.50)(75)] = \$8237.5$

La suma del producto de:

$P_o q_o = \$6085$; que se obtuvo así:

$[(6.00)(60) + (15.00)(250) + (8.00)(200) + (5.00)(75)] = \$6085$

Sustituyendo estos resultados en la fórmula, tenemos que:

$$\text{Índice de precios de Laspeyres} = \frac{8237.5}{6085} (100) = 135.4\%$$

Según el resultado obtenido, podemos concluir que el índice general de precios en 1996 es de 135.4%, o sea, que ha crecido en 35.4% desde 1990.

Una ventaja del método de Laspeyres es que podemos comparar un índice de precios de un año con otro índice de precios de otro año; ocurre lo mismo con la cantidad y el valor. Esta comparación directa se debe a que utilizamos, en el caso del precio, la misma cantidad del año base como factor de ponderación.

Otra ventaja para el cálculo del índice de precios por el método de Laspeyres consiste en que sólo usa una medida de cantidad, la del año base, por lo cual no es necesario hacer tabulaciones anuales para medir las cantidades consumidas.

Una desventaja del método de Laspeyres para el cálculo del índice de precios radica en que éste no toma en cuenta los cambios que se producen en los patrones de consumo.

Ejemplo 8.5

Con los datos de la tabla 8.6 calcule:

a) Índice de precio de Laspeyres.
b) Índice de cantidad de Laspeyres.
c) Índice de valor de Laspeyres.

Tabla 8.6. Datos acerca de las compras de un consumidor.

Artículos	1990		1997	
	Precio	cantidad	Precio	cantidad
Azúcar	3.00/lib.	2 libras	4.00/lib.	1.5 lb.
Naranjas	0.75 c/u	10 unidades	1.25 c/u	8 unid.
Refrescos	2.00 c/u	3 unidades	3.00 c/u	3 unid.
Periódico	3.00/ejemplar	1 ejemplar	7.00/ejemplar	1ejemplar

Indice de precio de Laspeyres = IPL

$$IPL = \frac{\sum(P_{1997}\, q_{1990})}{\sum(P_{1990}\, q_{1990})}\ (100)$$

de donde,

$$IPL = \frac{4(2) + 1.25(10) + 3(3) + 7(1)}{3(2) + 0.75(10) + 2(3) + 3(1)}\ (100) = \frac{36.5}{22.5}100 = 162\%$$

El índice de precios señala que, en conjunto, los precios se han incrementado en 62% en el año 1997 con relación a los precios de 1990.

Indice de cantidad de Laspeyres = IQL

$$IQL = \frac{\sum(q_{1997}\, P_{1990})}{\sum(q_{1990}\, P_{1990})}\ (100)$$

de donde:

$$IQL = \frac{1.5(3) + 8(0.75) + 3(2) + 1(3)}{2(3) + 10(0.75) + 3(2) + 1(3)}(100) = \frac{19.5}{22.5}(100) = 87\%$$

IQL = 87%

El resultado del índice de cantidad de Laspeyres indica que las cantidades totales de los artículos adquiridos por el consumidor disminuyeron en 1997 un 13% (o sea, 100% - 87% = 13%) con relación al 1990.

Para el cálculo del índice de cantidad de Laspeyres se usa el precio del año base como factor de ponderación.

Indice de valor de Laspeyres = IVL

$$IVL = \frac{\sum(P_{1997}\, q_{1997})}{\sum(P_{1990}\, q_{1990})}\ (100)$$

de donde:

$$IVL = \frac{4(1.5) + 1.25(8) + 3(3) + 7(1)}{3(2) + 0.75(10) + 2(3) + 3(1)}(100) = \frac{32}{22.5}(100) = 142\%$$

Para este comprador, el valor de los artículos consumidos en el año 1997 con relación al 1990 aumentaron en 42%.

8.2.2.3 Método de Paasche

El método de Paasche consiste en usar las cantidades consumidas en el período dado, y no en el año base, como factor de ponderación.

El método de Paasche es de gran utilidad, ya que combina los efectos de los cambios en los patrones de precio y de consumo. Así, este método es un mejor indicador de los cambios generales de la economía que el método de Laspeyres.

Los cambios de patrones de consumo de los diferentes elementos o conjunto de artículos manifiesta la diferencia existente entre los resultados que se obtengan por ambos métodos.

Una de las principales desventajas del método de Paasche consiste en la necesidad de disponer de las cantidades correspondientes del año en cuestión. Regularmente, las informaciones sobre las cantidades de cada período son muy difíciles de obtener o su disponibilidad no resulta precisamente asequible al interesado.

Las siguientes ecuaciones las vamos a usar para calcular los índices de precio, cantidad y valor mediante el método de Paasche.

Índice de Precios de Paasche = IPP

$$= \frac{\sum P_n\, q_n}{\sum P_o\, q_n}\ (100) \qquad 8.10$$

P_o = precios de los elementos en el año dado.
P_o = precios de los elementos en el año base.
q_n = cantidades (vendidas o producidas) en el año dado.

En este caso, los factores de ponderación serán las cantidades del año dado (q_n).

Índice de cantidades de Paasche = IQp

$$IQ_p = \frac{\sum q_n P_n}{\sum q_o P_n} (100) \qquad 8.11$$

Donde:
q_n = cantidades (vendidas o producidas) en el año dado.
q_o = cantidades (vendidas o producidas) en el año base.
P_n = precios de los elementos en el año dado.

Índice de valores de Paasche = IVP

$$IVP = \frac{\sum P_n q_n}{\sum P_o q_n} (100) \qquad 8.12$$

Donde:
P_n = precios de los elementos en el año dado.
P_o = precios de los elementos en el año base.
q_n = cantidades (vendidas o producidas) en el año dado.
q_o = cantidades (vendidas o producidas) en el año base.

Ejemplo 8.6

Con los datos de la tabla 8.6, calcule el índice de precios de Paasche.

Solución:

Índice de precios de Paasche = IPP

$$IPP = \frac{\sum (P_{1997} q_{1997})}{\sum (P_{1990} q_{1997})} (100)$$

Según los datos de la tabla 8.6 y sustituyendo en 8.10,

$$IPP = \frac{4(1.5) + 1.25(8) + 3(3) + 7(1)}{3(1.5) + 0.75(8) + 2(3) + 3(1)} (100) = \frac{32}{19.5} (100) = 164\%$$

El índice de precios, mediante el método de Paasche expresa que, en conjunto, los precios se incrementaron en 64% en 1997 con relación a los precios de 1990.

Como se puede observar en la página 244 el índice de precios mediante el método de Laspeyres, calculado a través de la tabla 8.6 es de 162%, para el año 1997 con relación a los precios del año 1990.

Estos resultados indican que el índice de precios de Paasche muestra una tendencia a los bienes y servicios más caros, pues señala un aumento de los precios de un 64% en vez de un incremento de 62%, obtenidos mediante el método de Laspeyres

Índice de Cantidad de Paasche = IQP

$$IQP = \frac{\sum (q_{1997} P_{1997})}{\sum (q_{1990} P_{1997})} (100)$$

Según los datos de la tabla 8.6 y sustituyendo en 8.11,

$$IQP = \frac{1.5(4) + 8(1.25) + 3(3) + 1(7)}{2(4) + 10(1.25) + 3(3) + 1(7)} (100) = \frac{32}{36.5} (100)$$

IQP = 87.7

Las cantidades totales adquiridas por el consumidor disminuyeron en 1997 en 12.3% (o sea, 100% - 87.7% = 12.3%) con relación al año 1990.

Índice de Valor de Paasche = IVP

$$IVP = \frac{\sum (P_{1997} q_{1997})}{\sum (P_{1990} q_{1997})} (100)$$

Según los datos de la tabla 8.6 y sustituyendo en 8.12,

$$IVP = \frac{4(1.5) + 1.25(8) + 3(3) + 7(1)}{3(2) + 0.75(10) + 2(3) + 3(1)}(100) = \frac{32}{22.5}(100) = 164\%$$

IVP = 164%

Este resultado indica que para este comprador el valor de los artículos consumidos en el año 1997 aumentó en 64% con relación al año 1990.

8.3 Índices más importantes en economía y administración

Los números índices son considerados como termómetros de la actividad económica y financiera en un momento determinado y sirven para indicar períodos de inflación, recesión, crecimiento económico, etc.

Los índices mundialmente más usados son: los índices de precios al consumidor, los de precios al por mayor, el del promedio industrial de Dow-Jones y el de la producción industrial.

8.3.1 Índice de precios al consumidor

El índice de precios al consumidor (IPC) mide el nivel general de precios de una canasta de los bienes y servicios adquiridos por la población en un período de referencia dado. Combina los precios individuales para formar un número compuesto de índice de precios. Mide los cambios experimentados en los artículos contenidos en la canasta familiar (conjunto de bienes y servicios de mayor uso por la población).

Este indicador, medidor por excelencia del nivel de inflación de un país, es utilizado también en la revisión de los contratos de alquiler; como punto de referencia en las negociaciones salariales; fijación de pensiones, actualización de seguros y en el cálculo del factor de ajuste de la base imponible del impuesto sobre la renta.

Unas de las razones principales por las que se debe realizar una medición lo más preciso posible de la inflación en un país es porque se trata de un fenómeno económico muy nocivo, por diversas razones, como:

a) Daña la estabilidad del poder adquisitivo de la moneda nacional.

b) Afecta el crecimiento económicos al hacer más riesgosos los proyectos de inversión.

c) Distorsiona las decisiones de consumo y ahorros.

d) Dificulta la intermediación financiera.

Los valores del índice de precios al consumidor se expresan como promedios anuales, cuatrimestrales, trimestrales y mensuales. En la tabla 8.7 se presenta una serie que contiene el índice de precios al consumidor de la República Dominicana desde el año 1984 hasta el año 2006, teniendo como período base al mes de enero de 1999. El período base del índice de precios es aquel en el cual los precios de los bienes y servicios son tomados para medir el cambio de los mismos durante el período de vigencia del índice de precios al consumidor. O sea, los precios del mes de enero de 1999 se tomaron como base, y por consiguiente el índice para este período se ha hecho igual a 100.

El nuevo índice de precios al consumidor de la República Dominicana, base Enero 1999, tiene como período de referencia aquel ciclo durante el cual se obtuvo la estructura de consumo; de octubre de 1997 a septiembre de 1998; período del levantamiento de la "Tercera Encuesta Nacional de Gastos e Ingresos de los Hogares", que sirvió de marco para el cambio de base del índice de precios al consumidor, cuyo período base anterior era noviembre de 1984.

TABLA 8.7 Indice de precios al consumidor 1984-2002
Series enlazadas, enero 1999 = 100

Años (a dic.)	Indice	Variación mensual	Var. porcentual con diciembre	Promedio 12 meses
1984	8.99	-	38.64	20.15
1985	11.76	-	30.85	45.34
1986	12.28	-	4.40	7.64
1987	15.07	-	22.67	13.55
1988	23.47	-	55.80	43.86
1989	31.59	-	34.59	40.66
1990	56.84	-	79.92	50.46
1991	61.33	-	7.90	47.08
1992	64.50	-	5.17	4.26
1993	66.30	-	2.79	5.25
1994	75.79	-	14.31	8.26
1995	82.77	-	9.22	12.54
1996	86.04	-	3.95	5.40
1997	93.24	-	8.37	8.30
1998	100.53	-	7.82	4.83
1999	105.66	-	5.10	6.47
2000	115.19	-	9.02	7.72
2001	120.24	-	4.38	8.88
2002	132.88	-	10.51	5.22
2003	189.56	-	42.66	27.45
2004	244.04	-	28.74	51.46
2005:				
Enero	245.96	0.79	0.79	48.09
Febrero	246.06	0.04	0.83	42.58
Marzo	245.88	-0.07	0.75	37.07
Abril	246.65	0.31	1.07	31.90
Mayo	245.91	-0.30	0.77	26.59
Junio	246.05	0.06	082	21.64
Julio	247.52	0.06	1.43	17.25
Agosto	250.83	1.34	2.78	13.38
Septiembre	259.01	3.26	6.13	10.38
Octubre	261.89	1.11	7.31	7.59
Noviembre	259.69	-0.84	6.41	5.55
Diciembre	262.19	0.96	7.44	4.19
2006:				
Enero	266.25	1.55	1.55	3.50
Febrero	265.93	-0.12	1.43	3.62

Fuente: Banco Central de la República Dominicana.

El índice de precios al consumidor se usa, además, para medir el "poder adquisitivo del consumidor o de la moneda,

determinándose así el valor real del dinero, que va a depender del nivel general de precios.

El poder adquisitivo del consumidor se expresa como el recíproco del índice de precios al consumidor en un momento determinado. De esta manera, (1/IPC) x 100 = poder adquisitivo.

Con los datos de la tabla 8.7 vamos a calcular el poder adquisitivo del peso dominicano para el período 1995-2006 (base, enero 1999).

Tabla 8.8 Poder adquisitivo del peso dominicano Base: enero 1999

Año (a dic.)	IPC	Poder adquisitivo (1/IPC) x 100
1995	82.77	1.208
1996	86.04	1.162
1997	93.24	1.072
1998	100.53	0.994
1999	105.66	0.945
2000	115.19	0.869
2001	120.24	0.831
2002	132.88	0.752
2003	189.56	0.527
2004	244.04	0.409
2005:		
Enero	245.96	0.406
Febrero	246.06	0.406
Marzo	245.88	0.407
Abril	246.65	0.405
Mayo	245.91	0.407
Junio	246.05	0.406
Julio	247.52	0.404
Agosto	250.83	0.398
Septiembre	259.01	0.386
Octubre	261.89	0.382
Noviembre	259.69	0.385
Diciembre	262.19	0.381
2006:		
Enero	266.25	0.375
Febrero	265.93	0.376

De acuerdo con los resultados de la tabla 8.8, el poder adquisitivo o de compra de un peso dominicano en febrero del

2006 es equivalente a 0.376 (treinta y ocho centavos) de enero del 1999. Con un peso dominicano de febrero del 2006 se compra lo que se compraba con treinta y ocho centavos en enero del 1999.

Asimismo, el IPC se usa para medir el Ingreso Real de los trabajadores, que es el ingreso ajustado para cambios en los precios. Al dividir el salario neto anual entre el valor del IPC para un momento determinado, multiplicado por cien, se obtendrá el ingreso real para ese momento.

Supongamos que un trabajador en el año 2001 obtuvo un salario neto de RD$ 150,000.00, ¿Cuál fue el ingreso real de ese trabajador en el año 2005?

Ingreso Real = (Salario neto / IPC) x 100

Ingreso Real = (RD$ 150,000.00 / 262.19) x 100 = RD$ 57,210.4

El ingreso real, ajustado para cambios en los precios, de ese trabajador en el 2005 fue de RD$ 57,210.4

8.3.2 Indice de precios al por mayor

Mide los cambios relativos de los precios que los productores o fabricantes pagan por la materia prima, e incluye los principales productos y materias que se utilizan en la industria. A este índice también se le denomina, índice de precios al productor.

8.3.3 Promedio industrial Dow-Jones

Este índice tiene como principal finalidad la descripción de la más importante entidad del mercado de valores, llamada Bolsa de Valores de Nueva York. La descripción se efectúa sobre la base de un solo precio promedio, que trata de reflejar la condición actual de todas las acciones, tomadas colectivamente, que están en lista de la Bolsa de Valores de Nueva York.

El promedio industrial Dow-Jones se calcula diariamente con los precios al cierre de las acciones de 30 firmas industriales

previamente elegidas. Se supone que 30 empresas son representativas de los precios accionarios en general.

8.3.4 Índice de producción industrial (IPI)

El índice de propocisión industrial (IPI) es un indicador estadístico que permite el seguimiento de la actividad productiva de las ramas industriales mineras y de servicios, con respecto, a un período de referencia denominado año base. Muestra la evolución conjunta de la cantidad y calidad producida, eliminando la influencia de los precios.

Este índice mide los cambios en el volumen físico de la producción de las empresas manufactureras, mineras y de servicios.

En los Estados Unidos, este índice está compuesto por alrededor de 100 industrias, que representan aproximadamente entre el 70 y el 80% de la producción de bienes y servicios en el sector privado. Este índice refleja el estado general de las empresas industriales.

8.6 Un consumidor ha realizado las siguientes compras de artículos comestibles en 1996 y 2006.

Artículos	1996		1998	
	Precio (RD$)	Cantidad	Precio (RD$)	Cantidad
Café (libra)	28.00	2	70.00	1.5
Arroz (libra)	4.10	15	20.00	14
Leche (litro)	10.00	4	35.00	3
Carne (libra)	18.00	6	60.00	6
Aceite (libra)	12.00	3	25.00	4

Tomando como base el año 1996, calcular:

a) Índice de precios de Laspeyres.
b) Índice de cantidad de Laspeyres.
c) Índice de valor de Laspeyres.
d) Comente los resultados anteriores.

8.7 Con los datos del ejercicio 8.6, calcular:

a) Índice de precios de Paasche.
b) Índice de cantidad de Paasche.
c) Índice de valor de Paasche.
d) Comente los resultados anteriores.

8.8 Establezca la diferencia entre los resultados de Laspeyres y de Paasche para el cálculo de índices compuestos de precio.

8.9 La tabla que se presenta a continuación contiene los precios y las cantidades vendidas por una empresa en los años 2003 y 2005.

Artículos	2003		2005	
	Precio (RD$)	Cantidad	Precio (RD$)	Cantidad
Computadoras (Unid.)	18,000	25	15,000	45
CD Players (Unid.)	5,500	150	4,000	275
Celulares (Unid.)	5,600	85	4,500	180

Calcular, tomando como año base el 2003:

a) El índice de precios de Laspeyres.
b) El índice de cantidad de Laspeyres.
c) El índice de valor de Laspeyres.

d) Haga un breve comentario de los resultados obtenidos.

8.10 Con los datos del ejercicio 8.9, calcular:

a) El índice de precios de Paasche.
b) El índice de cantidad de Paasche.
c) El índice de valor de Paasche.

8.11 Los salarios semanales promedio de los obreros de una industria se presentan para el período 1998–2005, así como el IPC para el mismo período:

Año	Salarios semanales (Promedio)	Índice de precio al consumidor (enero 1999 = 100)
1998	1,100	100.53
1999	1,400	105.66
2000	1,900	115.19
2001	2,100	120.24
2002	2,275	134.00
2003	3,500	189.56
2004	3,900	244.04
2005	4,100	262.19

Calcular:

a) El poder adquisitivo del dinero o poder de compra de los trabajadores usando como base Enero 1999;
b) Determine el salario real para cada año;
c) Elabore una nueva tabla con el poder adquisitivo del dinero y con el salario real de los trabajadores;
d) Comente los resultados obtenidos.

Bibliografía

Ary, Donald; Jacobs; Lucy; Razavieh, Asghar; 1982. **Introducción a la Investigación Pedagógica.** 2da. edición; Nueva editorial Interamericana, S.A. Mexico, D.F.

Berenson, Mark L. y Levine, David M.; 1998. **Estadística Básica en Administración, Conceptos y Aplicaciones;** 6ta. edición, Prentice Hall Hispanoamérica, S.A., México.

Cochram, William G.; 1982. **Técnicas de muestreo;** 3era. edición; Compañía Editorial Continental, S.A. (C.E.C.S.A.); México.

Estados Unidos, Organización de las Naciones Unidas; 1958. **Manual de métodos de censos de población: Aspectos generales de los censos de población;** Vol. I, New York.

Freund, John E.; Williams, Frank J. y Perles, Benjamín M.; 1996. **Estadística para la Administración, con enfoque moderno;** 5ta. edición; Prentice Hall Hispanoamérica, S.A., México.

Freund, John E.; Millar, Irwin; Miller Marylees; 2000. **Estadística Matemática con Aplicaciones;** 6ta. edición; Prentice Hall. México. S.A.

Glass, Gene V y Stanley, Julián C.; 1997. **Métodos Estadísticos Aplicados a las Ciencias Sociales;** 1era. edición en español; Prentice-Hall Hispanoamérica, S.A., México.

Bach, Antonio, 1961. **Estadística descriptiva y gráficos (con aplicación a las necesidades de las empresas comerciales, industriales y análogas);** 2da. edición; Editorial José Montero; Barcelona, España.

Guilford, J.P. y Fruchter, Benjamín; 1984. **Estadística Aplicada a la Psicología y la Educación;** 6ta. edición; Mc Graw Hill Latinoamericana, S.A.; Bogotá, Colombia.

Hoel, Paul G.; 1982. **Estadística Elemental;** 3era. edición. Compañía Editorial Continental, S.A. (C.E.C.S.A.), México.

Hophins, Kenneth; Hopking, B.R. y Glass, Gene V.; 1998. **Estadísticas Básica para las ciencias sociales y del comportamiento;** 3ra. edición; Prentice-Hall Hispanoamérica, S.A., México.

Johnson, Robert, y Kuby, Patricia; 2004. **Estadística Elemental: Lo Esencial;** 3era. edición, Internacional Thomson Editores, México.

Kazmier, Leonard y Mata Díaz, Alfredo; 1997. **Estadística aplicada a la Administración y a la Economía;** 2da. edición, Mc Graw-Hill Interamericana de México, S.A., México.

Levin, Richard I.; 1994. **Estadística para Administradores;** 2da. edición; Editorial Prentice-hall Hispanoamericana, S.A., México.

Lininger, Charles A. y Warwick, Donald P.; 1982. **La Encuesta por Muestreo; Teoría y Práctica;** 2da. edición; Compañía Editorial Continental, S.A. (C.E.C.S.A.); México.

Lind, Douglas A.; Marchal, Willam G.; Wathen, Samuel A.; 2005. **Estadística Aplicada a los Negocios y a la Economía.** 12va. Edición; Ma Graw Hill Interamericana; México.

Longley-Cook, L.H.; 1981. **Problemas de Estadística y cómo resolverlos;** 1era. edición en español; Compañía Editorial Continental, S.A. (C.E.C.S.A.); México.

Mc Collough, Celeste y Van Atta, Loche; 1971. **Estadística para sociólogos;** 1era. edición en español; Editorial Tecnos, S.A.; Madrid, España.

Neter, John y Wasserman, William; 1970. **Fundamentos de Estadística, aplicada a los negocios y la economía;** 7ma. impresión; Compañía Editorial Continental, S.A. (C.E.C.S.A.); México.

Perú, Ministerio de Transporte y Comunicaciones; 1977. **Normas sobre la presentación tabular de datos;** Lima.

Raj, Des; 1980. **Teoría del Muestreo;** 1era. edición en español; Fondo de Cultura Económica; México.

República Dominicana, Banco Central; 1998. **Boletín trimestral, Abril-Junio de 1998;** Volumen LIII, N° 5; 4, 5 y 6; Santo Domingo.

República Dominicana, Secretaría de Estado de Educación y Cultura; 1998. **Estadísticas e indicadores de educación, 1996-1997;** Santo Domingo.

República Dominicana, Oficina Nacional de Estadística; 2005. **Encuesta Nacional de Hogares de Propósitos Múltiples, ENHOGAR-2005;** Santo Domingo.

Richards, Larry E. y La Cava, Jerry J.; 1980. ***Estadística en los negocios: ¿Por qué y cuándo?*** *1era. edición en español; Editorial Mc Graw-Hill Latinoamericana, S.A., Bogotá, Colombia.*

Seijas E., Félix L., 1981. **Investigación por Muestreo;** 1era. edición; Editorial Tecni-Ciencia Libros, S.A.; Caracas, Venezuela.

Schwar, Johannes F. y Puy Huarte, José; 1975. **Métodos Estadísticos en Ingeniería de Tránsito;** 1era. edición en español; editores: Asociación Mexicana de Caminos, A.C. y Representaciones y Servicios de Ingeniería, S.A.; México.

Shao, Stephen P.; 1979. **Estadística para economistas y administradores de empresas;** 15ta. edición; Herrero Hermanos, Sucs. S.A.; México.

Snedecor, George W. y Cochram, William G.; 1982. **Métodos Estadísticos;** 9na. edición; Compañía Editorial Continental, S.A. (C.E.C.S.A.), México.

Spiegel, Murray R.; 1997. **Estadística;** 2da. edición; Mc-Graw Hill, México.

Stevenson, William J.; 1981. **Estadística para Administración y Economía;** 1era. edición en español; Editorial Harper and Row Latinoamericana, (Harla, S.A.); México.

Walpole, Ronaldo.; Myers, Raymond H.; Myers, Sharon L.; 1998. **Probabilidad y Estadística para Ingenieros;** 6ta. edición; Prentice Hall, Hispanoamericana, S.A.; México.

Yamane, Taro; 1976. **Problemas de Estadística,** 3era. edición; Editorial Harper and Row Latinoamericana, HARLA, S.A.; México.

COLOFÓN

Esta cuarta edición de ESTADÍSTICA BÁSICA, de Carlos Custodio, se terminó de imprimir en los talleres gráficos de Editora Búho, C. por A., calle Elvira de Mendoza No. 56, en Marzo del 2007.