UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CONTADURÍA Y ADMINISTRACIÓN

AUTOR: ANTONIO CAMARGO MARTÍNEZ

Matemáticas III (Estadística Descriptiva)		Clave: 1368
Plan: 2005		Créditos: 8
Licenciatura: Informática		Semestre: 3º
Área: Matemáticas		Hrs. asesoría: 4
Requisitos: Ninguno		Hrs. por semana: 4
Tipo de asignatura:	Obligatoria (x)	Optativa ( )

Objetivo general de la asignatura

Al final del curso, el alumno aplicará y evaluará los principios estadísticos para resolver problemas en las ciencias contables administrativas.

Temario oficial (horas sugeridas 64 hrs.)

1. Introducción (4 hrs.)
2. Estadística descriptiva (16 hrs.)
3. Análisis combinatorio (8 hrs.)
4. Teoría de la probabilidad (16 hrs.)
5. Distribuciones de probabilidad (16 hrs.)
6. Números índice (4 hrs.)

Introducción general

En esta asignatura el estudiante investigará lo relativo a la estadística descriptiva.

En el **tema 1**, se describirán las generalidades de la estadística en general y ejemplos de aplicación en diversos aspectos de la administración. Se señalarán las principales características de muestras y poblaciones, las diferencias entre los estadísticos y los parámetros poblacionales y la diversificación de la estadística en descriptiva e inferencial.

En el **tema 2**, se estudiarán las diversas características de un conjunto de datos, desde los diferentes tipos de variables y sus escalas de medición. Se estudiará la metodología para la organización y procesamiento de datos, sus distribuciones de frecuencias absolutas y relativas, así como su presentación gráfica en histogramas, polígonos de frecuencias y ojivas. Por otra parte, se conocerán las más importantes medidas de tendencia central y de dispersión. Por último se analizarán los teoremas de Tchebysheff y de la regla empírica.

En el **tema 3**, se expondrán los principios básicos de conteo a partir de los cuales se deducen las fórmulas y técnicas del análisis combinatorio. Se especificarán las principales diferencias entre las ordenaciones, permutaciones y combinaciones. Estos métodos de conteo constituyen la base de la teoría de la probabilidad.

En el **tema 4**, se revisarán las diversas clases de probabilidad, así como los conceptos de espacio muestral y eventos. También se analizarán las reglas fundamentales de la adición y de la multiplicación. Se elaborarán e interpretarán las tablas de probabilidad conjunta, la probabilidad condicional y se conocerá y aplicará el teorema de Bayes.

El **tema 5** abarca el conocimiento de las características y diferencias de las variables discretas y continuas, así como de la distribución general de una variable discreta. Además, se analizarán las principales particularidades y fórmulas de una distribución binomial, de una distribución de Poisson, de una distribución normal y de una distribución exponencial. Por último, se enunciará la ley de los grandes números y su interpretación.

El **tema 6** está relacionado con los diversos tipos de números índice, de los índices ponderados y el de los precios al consumidor.

Tema 1. Introducción

Objetivo particular

Al terminar la unidad el estudiante deberá conocer los principios y generalidades de la estadística, así como las principales características de muestras y poblaciones.

Temario detallado

1. Introducción.
    - 1.1 Generalidades.
    - 1.2 Poblaciones y Muestras.

1. Introducción

El mundo de los negocios se manifiesta fundamentalmente a través de datos de diferentes tipos, los cuales requieren, de acuerdo con su naturaleza, un tratamiento particular. Del correcto manejo de la información depende en gran medida el éxito de una negociación; ahí radica la importancia de conocer los diferentes tipos y las escalas con las cuales pueden medirse las diferentes maneras de extraer una muestra que conserve las características del conjunto del cual proviene, la forma de agruparlos correctamente y a partir de ello aplicar los métodos estadísticos para representarlos gráficamente y determinar las medidas de tendencia central y de dispersión, para inferir un comportamiento y características de la población donde se originan. La descripción de la información desde el punto de vista de la estadística constituye la parte fundamental del proceso de análisis de un conjunto de datos.

1.1. Generalidades

La **estadística** agrupa un conjunto de técnicas mediante las que se recopilan, agrupan, estructuran y, posteriormente, se analizan conjuntos de datos.

El propósito de la estadística es darles **sentido o “carácter”** a los datos recolectados, es decir, nos puedan dar una idea de la situación que reflejan para, con base en ella, tomar decisiones. Algunos ejemplos nos pueden aclarar este concepto.

A un administrador le entregan en una caja un listado de computadora de 3000 hojas que contiene el detalle (departamento, cliente, productos vendidos e importe de cada transacción) de las ventas de un mes de una gran tienda departamental. La presentación de los datos del listado difícilmente le será útil para la toma de decisiones, por lo que el administrador tendrá que ordenarlos, clasificarlos y concentrarlos para que le sean útiles. Las técnicas que permiten ese ordenamiento, clasificación y concentración son, precisamente, técnicas estadísticas.

En una situación similar, a un auditor le muestran el archivo en el que se encuentran las copias fiscales de las 46,000 facturas que una empresa emitió durante el ejercicio fiscal. Desde luego, los datos contenidos en las copias son valiosos para su trabajo de auditoria y tal vez sean indispensables para fundamentar su opinión de la empresa con el fin de emitir su dictamen. Sin embargo, la información tal como se encuentra en las copias difícilmente le será útil. Otra vez, como en el caso anterior, será necesario ordenar, clasificar y procesar los datos para obtener conclusiones sobre ellos.

En el caso de los licenciados en Informática, dado que su profesión se dedica precisamente a buscar los mejores medios de procesar la información, es de suyo evidente que las técnicas (estadísticas) que hacen más eficiente ese trabajo deben interesarle.

1.2. Poblaciones y muestras

En nuestro estudio de la realidad, frecuentemente debemos hacer frente a conjuntos muy grandes de hechos, situaciones, mediciones, etc.

A continuación se dan algunos ejemplos:

Si deseamos instalar una cafetería en nuestra Facultad, debemos saber con claridad quiénes serán nuestros clientes: pueden ser los estudiantes, los maestros y el personal administrativo de la propia Facultad y tal vez algunos visitantes. Todas estas personas conformarán la población cuyos hábitos de consumo de alimentos y bebidas deseamos conocer.

Cuando un auditor desea investigar los egresos de una entidad económica deberá estudiar todos los cheques emitidos por ésta. La población que desea estudiar es, por tanto, la de todos los cheques emitidos por el organismo en el periodo que desea investigar.

Un administrador desea estudiar la duración o vida útil de todos los focos producidos por una pequeña fábrica durante un mes. La población de estudio será la de todos los focos producidos durante ese mes.

De los ejemplos anteriores podemos ver que el concepto de *población* se parece, en algunos casos, a la idea que tenemos de un conjunto de personas (como en la población de un país). Tal es el caso del primer ejemplo; en los otros dos, las poblaciones mencionadas no son de personas, sino de cheques y de focos. Podemos decir, que una *población* es el conjunto de todas las mediciones u observaciones de interés para el investigador que realiza un trabajo con un objetivo concreto de conocimiento de la realidad.

Existen diversas circunstancias por las cuales un investigador no desea o no puede físicamente verificar observaciones en toda la población y se tiene que conformar con estudiar un subconjunto de ellas. Entre estas circunstancias se encuentran las siguientes:

Limitaciones de tiempo. Si deseamos instalar la cafetería del ejemplo ya citado dentro de seis meses y la investigación de los hábitos de consumo de todos los clientes potenciales nos lleva ocho meses, es claro que deberemos resolver nuestra necesidad de información de otra manera.

Limitaciones de recursos. El auditor de nuestro segundo ejemplo podría desear estudiar todos los cheques emitidos, pero la empresa auditada no puede pagar el costo de una revisión tan exhaustiva. Por ello, el auditor debe basar su opinión en una investigación más limitada.

Imposibilidad física. Si el administrador de la fábrica de focos desea saber la duración o vida útil de un foco, lo único que puede hacer es dejarlo prendido constantemente hasta que se funda y registrar el tiempo en el que eso ocurre. Desde luego que si se sigue este procedimiento al final la fábrica no contará con ningún foco para vender.

Cuando por los motivos antes citados no es conveniente, o incluso posible, obtener la información que se necesita de toda la población, los investigadores recurren a estudiar una parte de esa población a la que se llama *muestra*.

Una *muestra* es, entonces, cualquier subconjunto de una población.

A las características de las poblaciones las denominamos *parámetros* y a las características similares de las muestras las denominamos *estadísticos*. Así, la *media* de la población (a la que conocemos con la letra griega μ) es un parámetro y la media de la muestra (a la que conocemos como $\bar{x}$) es un estadístico.

Recordemos que la *media* o *promedio aritmético* es una medida de tendencia central que toma en cuenta el valor específico de cada uno de los datos que se estudian y se obtiene sumando esos valores entre el número de datos.

La *desviación estándar* de la población (a la que conocemos como σ). es un parámetro y la desviación estándar de la muestra (a la que denominamos s) es un estadístico.

Recordemos que la desviación estándar es una *medida de dispersión* o *desviación* de cada uno de los datos que forma una muestra o una población respecto de la media o promedio aritmético. Es la raíz cuadrada del resultado de dividir la sumatoria de las desviaciones cuadráticas de cada uno de los datos con respecto a la media, entre el número de datos en estudio.

Normalmente cuando hacemos estudios con base en muestras, conocemos los estadísticos (los datos de la muestra) y éstos nos sirven para estimar los datos reales de la población a los que conocemos como parámetros.

En resumen, los *parámetros* son datos de las poblaciones, en tanto que los *estadísticos* son datos de las muestras. Los estadísticos nos sirven para tratar de estimar o inferir los parámetros cuando no podemos conocerlos estudiando directamente toda la población.

La estadística es una rama de la ciencia matemática encargada del diseño de experimentos (procedimientos de muestreo), del análisis de datos, y los procedimientos para inferir acerca de las características de una población con base en la información obtenida de una muestra. La estadística entonces, se divide en 2 tipos: la **estadística descriptiva** y la **estadística inferencial** o inferencia estadística.

La **estadística descriptiva** incluye aquellas técnicas que nos permiten resumir y describir datos. La preparación de tablas, la elaboración de gráficos y las técnicas para el cálculo de los diferentes parámetros de las poblaciones forman parte de las técnicas de la estadística descriptiva. Los administradores, contadores e informáticos, que se han mencionado en los ejemplos arriba citados, harán bien en allegarse de técnicas de estadística descriptiva para *resumir y caracterizar* sus *datos* con el objeto de *tomar decisiones* correctas.

En México, una vez cada diez años se hace un estudio general de la población del país que recibe el nombre de *Censo general de población y vivienda*. Éste es un estudio muy amplio de estadística descriptiva para conocer diversas características demográficas de los mexicanos. A todos los estudios que se realizan estudiando a todos los elementos de una población se les conoce como *estudios censales o censos.*

La **estadística inferencial** comprende un conjunto de técnicas que nos permiten *estimar* (o inferir y de allí su nombre) las *características* (frecuentemente los parámetros) de una *población* con base en una muestra obtenida de ella y, una vez estimados, tomar decisiones sobre esa población. Estas decisiones incluyen un factor de riesgo, dado que las características de la población se infieren aproximadamente, pero *no se conocen con certeza*. Por ello, para la estadística inferencial ***se utilizan*** conceptos de *probabilidad*.

La estadística se utiliza frecuentemente durante las elecciones federales y locales que se pronostican los resultados con base en lo que se ha dado en llamar *conteos rápidos*. Estos conteos se realizan registrando los datos de un pequeño conjunto de casillas electorales cuidadosamente seleccionadas. Estos conteos rápidos son un ejemplo de un estudio muestral, es decir, hecho mediante muestras con el objeto de inferir características de toda la población.

El crecimiento de la población y con ella el surgimiento de nuevos problemas que resolver, hicieron posible la ampliación de las aplicaciones de la matemática de las ciencias físicas a otras como: las ciencias del comportamiento, las ciencias biológicas y las ciencias sociales entre otras.

Teniendo esto en consideración no es de extrañar que los usos y aplicaciones de la Probabilidad y la Estadística pasaran de los juegos de azar y los fines militares y posteriormente recaudatorios a las *entrañas* de las organizaciones convirtiéndose en un bastión de las empresas exitosas.

Históricamente, el crecimiento y desarrollo de la estadística moderna puede trazarse desde dos fenómenos separados:

- La necesidad del Gobierno de recabar datos sobre sus ciudadanos y
- El desarrollo en las matemáticas, de la teoría de probabilidades.

Así por ejemplo, durante las civilizaciones egipcia, griega y romana, los datos se obtenían principalmente con propósitos de impuestos y reclutamiento militar.
En la edad media, las instituciones eclesiásticas a menudo mantenían registros de nacimientos, muertes y matrimonios. En nuestro país, organismos tales como el INEGI realizan levantamientos de censos.

Por otra parte, la mayoría de los autores coinciden en que la estadística proporciona los elementos básicos para fundamentar una investigación, como son:

1. La Planeación para la obtención de los datos con los cuales se puedan extraer conclusiones confiables.
2. El Análisis de estos datos.
3. El tipo de conclusiones que pueden obtenerse con los datos disponibles.
4. La confianza que nos merecen los datos.

Como puede notarse, la estadística nos permite realizar estudios de tipo descriptivo y explicativo por medio de sus dos ramas, prácticamente en todas las áreas del conocimiento humano; Claro esta siempre y cuando apliquemos un método.

Bibliografía del tema 1

BERENSON L., Mark, David Levine M. y Timothy Krehbiel C., *Estadística para administración*, 2ª edición, México, Prentice Hall, 2001. LEVIN, Richard I. y David S. Rubin, *Estadística para administradores*, 6a. Edición, México, Prentice Hall, 1996.

MASON D., Robert, Douglas Lind A. y William Marchal G, *Estadística para administración y economía*, 11ª edición, Colombia, Alfaomega, 2004.

Actividades de aprendizaje

1.1 Con el estudio de la bibliografía específica sugerida para el tema 1, escribe en un cuadro los conceptos principales y sus aplicaciones.

1.2 Realiza un compendio de la historia de la estadística.

1.3 Proporciona por lo menos 5 situaciones profesionales o empresariales en que los datos son necesarios.

1.4 Redacta tres ejemplos prácticos relacionados con la estadística descriptiva en los libros de referencia para este tema.

1.5 Investiga tres ejemplos prácticos y en un cuadro relacionados con la estadística inferencial en los libros de referencia.

1.6 Investiga y describe las fuentes clave para la recopilación de datos.

1.7 Visita la página www.aulafacil.com, (Curso de estadística clase 1), compara los temas estudiados indicando tus conclusiones.

Cuestionario de autoevaluación

1. Explica brevemente el propósito de la estadística.
2. Describe en qué consiste una población.
3. Describe en qué consiste una muestra.
4. ¿Cuáles son las limitaciones más importantes para investigar las características de una población?
5. Define qué significa un parámetro e investiga por lo menos tres ejemplos.
6. Define qué significa un estadístico e investiga por lo menos tres ejemplos.
7. ¿Qué es la estadística descriptiva?
8. ¿Qué es la estadística inferencial?
9. ¿En qué consiste un censo poblacional?
10. ¿Qué es un conteo rápido, dónde y porqué se utiliza?
11. Explica porque un administrador necesita conocer la Estadística.
12. Indica a que se refiere el concepto de *pensamiento estadístico*
13. Dé tres ejemplos de muestras y poblaciones para el estudio de la estadística.

Examen de autoevaluación

1. Los elementos fundamentales de la estadística son:
   a. Números
   b. Letras
   c. Gráficas
   d. Datos
   e. Símbolos

2. Uno de los objetivos principales de la estadística es:
   a. Obtener conclusiones para la toma de decisiones.
   b. Obtener datos para procesarlos.
   c. Calcular medidas numéricas.
   d. Determinar bases de datos.
   e. Identificar las variables discretas y continuas.

3. La modalidad de la estadística que entre otras menciones incluye *procedimientos empleados para organizar*, *analizar* y *concluir sobre las características de un conjunto de datos numéricos* es:
    a. Numérica
    b. Descriptiva
    c. Cuántica
    d. Booleana
    e. Inferencial

4. El hecho de realizar un muestreo sobre la población *en condiciones de miseria* en la zona de las Lomas de Chapultepec puede conducirnos a obtener resultados:
    a. Equivocados pues ahí habitan mayoritariamente familias en condiciones de solvencia económica.
    b. Equivocados pues a la población que ahí habita no le interesa el tema en cuestión.
    c. Verdaderos pues ahí la gente es más culta y sabe demasiado del tema.
    d. Verdaderos en esa población encontramos pocas personas en condiciones de miseria
    e. Dudosos ya que esporádicamente se ve transitando gentes en condiciones de miseria.

5. Utilizar los resultados de un muestreo para estimar características de la población, es una aplicación de la estadística:
    a. Numérica
    b. Descriptiva
    c. Cuántica
    d. Booleana
    e. Inferencial

6. Es un área de las matemáticas que trata de los métodos y medios para recolectar, organizar, analizar e interpretar datos con el propósito de tomar decisiones:

   a. Análisis de mercado
   b. Censo
   c. Estadística
   d. Estudios socioeconómicos
   e. Muestreo

7. Los estadísticos nos sirven para estimar o inferir:

   a. Datos
   b. Números
   c. Imágenes
   d. Símbolos
   e. Parámetros

8. El estudiar parámetros se refiere a las características de:

   a. Datos
   b. Muestras
   c. Gráficas
   d. Poblaciones
   e. Números

9. El estudio general de las características de las familias de una población se le llama:

   a. Muestreo
   b. Investigación de mercados
   c. Auditoria
   d. Examen
   e. Censo

10. Cuando se efectúa un estudio para estimar la preferencia del electorado por cierto candidato, ¿qué parte de la estadística se usará?

   a. Descriptiva

   b. Cuantitativa

   c. Inferencial

   d. Numérica

   e. General

Tema 2. Estadística Descriptiva

Objetivo particular

Al terminar la unidad, el estudiante deberá conocer los procedimientos de recolección, ordenación y tabulación de datos para obtener, en su caso, sus frecuencias absolutas y relativas, y mostrar en distintos tipos de gráficas las distribuciones en estudio. Mediante el conocimiento de las principales medidas de posición y de dispersión podrá identificar estadísticos y parámetros que proporcionen información relevante para la toma de decisiones.

Temario detallado

2.1. Tabulación de datos

- 2.1.1. Escala para datos de tipo nominal
- 2.1.2. Escala para datos de tipo ordinal
- 2.1.3. Escalas numéricas
- 2.1.4. Escalas de intervalo
- 2.1.5. Escalas de razón
- 2.1.6. Uso de computadoras en estadística
- 2.1.7. Principales elementos de las tablas
- 2.1.8. Tablas simples
- 2.1.9. Tablas de frecuencias
- 2.1.10. Tablas de doble entrada
- 2.1.11. Tablas de contingencia

2.2. Distribuciones de frecuencia

- 2.2.1. Distribución de frecuencias para variables cuantitativas
- 2.2.2. Cuadros estadísticos
- 2.2.3. Frecuencias absolutas y relativas

2.3. Presentación gráfica de datos

- 2.3.1. Histogramas y polígonos de frecuencias
- 2.3.2. Pasos a seguir para la elaboración de un diagrama de frecuencias (o polígono de frecuencias) y un histograma.

Introducción

El mundo de los negocios se manifiesta fundamentalmente a través de datos de diferentes tipos, los cuales requieren de acuerdo con su naturaleza un tratamiento particular. Del correcto manejo de la información depende en gran medida el éxito de una negociación; ahí estriba la importancia de conocer los diferentes tipos y las escalas con las cuales pueden medirse las diferentes maneras de extraer una muestra que conserve las características del conjunto del cual proviene, la forma de agruparlos correctamente y a partir de ello aplicar los métodos estadísticos para representarlos gráficamente y determinar las medidas de tendencia central y de dispersión para inferir un comportamiento y características de la población de la cual provienen. La descripción de la información, desde el punto de vista de la estadística, constituye la parte fundamental del proceso de análisis de un conjunto de datos.

2.1. Tabulación de datos

Los métodos estadísticos que se utilizan dependen, fundamentalmente, del tipo de trabajo que se desee hacer. Si lo que se desea es trabajar con los datos de las poblaciones para representarlos y caracterizarlos de una manera tal que sean útiles, estaremos hablando de métodos de la estadística descriptiva. Si lo que se desea es aproximar las características de una población con base en una muestra, se utilizarán las técnicas de la estadística inferencial. Estas últimas son tema de la materia de Estadística II, que el alumno estudiará posteriormente. En cuanto a las primeras, existe un sinnúmero de ellas. Las principales las podemos agrupar como técnicas de resumen de datos, técnicas de presentación de datos y técnicas de obtención de parámetros.

Tipos de técnica	Descripción
Las técnicas de resumen	Nos indican la mejor manera para ordenar y agrupar la información, de forma tal que ofrezca un sentido para el usuario de una manera que los datos en bruto no lo harían. Las técnicas de agrupación de datos y preparación de tablas se incluyen dentro de las técnicas de resumen.
Las técnicas de presentación de datos	Nos permiten obtener una serie de gráficas que, adecuadamente utilizadas, nos dan una idea visual e intuitiva de la información que manejamos. El alumno recuerda, sin duda, haber visto en algún periódico gráficas de barras o circulares (llamadas de (pastel) *pie* o "*pay*", por su pronunciación en inglés).
Las técnicas de obtención de parámetros	Nos llevan a calcular números que nos dan una idea de las principales características de la población. El conjunto de las 45 calificaciones que un alumno ha obtenido durante sus estudios profesionales nos pueden dar no mucha idea de su desempeño, pero si obtenemos su promedio (técnicamente llamada media aritmética) y éste es de 9.4, nos damos una clara idea de que es un buen estudiante. Los parámetros son números que nos sirven para representar (bosquejar una idea) de las principales características de las poblaciones.

Cuadro 2.1. Técnicas de la estadística descriptiva

En cualquier estudio estadístico, los datos pueden modificarse de sujeto en sujeto. Si, por ejemplo, estamos haciendo un estudio sobre las estaturas de los estudiantes de sexto de primaria en una escuela, la estatura de cada uno de los niños y niñas será distinta, variará. Por ello decimos que la estatura es una **variable**.

Los especialistas en estadística realizan experimentos o encuestas para manejar una amplia variedad de fenómenos o características llamadas variables aleatorias.
Los **datos variables** pueden registrarse de diversas maneras, de acuerdo con los objetivos de cada estudio en particular y, conforme con ello, podemos decir que podemos tener diferentes escalas de medición.

2.1.1. Escala para datos de tipo nominal

Son aquellas que no tienen un *orden* o *dimensión preferente* o *particular* y contienen observaciones que solamente pueden clasificarse o contarse. En un estudio de preferencias sobre los colores de automóviles que prefriere un determinado grupo de consumidores, se podrá decir que algunos prefieren el color rojo, otros el azul, algunos más el verde; pero no se puede decir que el magenta vaya *después* que el morado o que el azul sea *más grande* o más chico que el verde.

Para trabajar adecuadamente con escalas de **tipo nominal**, cada uno de los individuos, objetos o mediciones debe pertenecer *a una y solamente a una* de las categorías o clasificaciones que se tienen y el conjunto de esas categorías debe ser exhaustiva; es decir, tiene que contener a todos los casos posibles. También las categorías de datos no cuentan con un orden lógico.

2.1.2. Escala para datos de tipo ordinal

En esta escala, las variables sí tienen un **orden natural** (de allí su nombre) y cada uno de los datos puede localizarse dentro de alguna de las categorías disponibles. El estudiante habrá tenido oportunidad de evaluar a algún maestro; las preguntas implican frecuentemente categorías como "siempre, frecuentemente, algunas veces, nunca". Es fácil percatarse que "siempre" es más frecuente que "algunas veces" y "algunas veces" es más frecuente que "nunca". Es decir, en las escalas de tipo ordinal se puede *establecer una gradación* u orden natural para las categorías. No se puede, sin embargo, establecer comparaciones cuantitativas entre las categorías. No podemos decir, por ejemplo, que "frecuentemente" es el doble que "algunas veces" o que "nunca" es tres puntos más malo que "frecuentemente".

Para trabajar adecuadamente con escalas de tipo ordinal debemos recordar que las categorías son mutuamente excluyentes (cada dato puede pertenecer o una y sólo a una de las categorías) y deben ser exhaustivas (es decir, cubrir todos las posibles respuestas).

2.1.3. Escalas numéricas

Estas escalas, dependiendo del manejo de se le dé a las variables, pueden ser *discretas* o *continuas*.

Las *escalas discretas* son aquellas que solo pueden *aceptar determinados valores* dentro de un rango.

El número de hijos que tiene una pareja es, por ejemplo, un *dato discreto*. Una pareja puede tener 1, 2, 3 hijos, etc.; pero no tiene sentido decir que tienen 2.3657 hijos. Una persona puede tomar 1, 2, 3, 4, etc., baños por semana, pero tampoco tiene sentido decir que toma 4.31 baños por semana.

Por el contrario, las **escalas continuas** son aquellas que pueden aceptar cualquier valor dentro de un rango y, frecuentemente, el número de decimales que se toman dependen más de la precisión del instrumento de medición que del valor del dato en sí.

Podemos decir, por ejemplo, que el peso de una persona es de 67 Kg.; pero si medimos con más precisión, tal vez informemos que el peso es en realidad de 67.453 Kg y si nuestra báscula es muy precisa podemos anotar un mayor número de decimales.

El objetivo del investigador condiciona fuertemente el tipo de escala que se utilizará para registrar los datos. Tomando el dato de la estatura, éste puede tener un valor puramente categórico. En algunos deportes, por ejemplo, el básquetbol, sólo de determinada estatura para arriba los candidatos a jugador se admiten en el equipo y de esa estatura para abajo no se admiten. La variable estatura tiene solo dos valores “aceptado” y “no aceptado” y es una **variable nominal**. Esta misma variable, para otro estudio, puede trabajarse con una escala de tipo ordinal: *bajos de estatura*, *de mediana estatura* y *altos*. Si tomamos la misma variable y la registramos por su valor en centímetros, la estaremos trabajando como una **variable numérica**.

Dependiendo de las intenciones del investigador, se le puede registrar como variable discreta o continua (variable discreta si a una persona se le registra, por ejemplo, como de 173 cm 174cm, y si mide unos milímetros más o menos se redondeará al centímetro más cercano; variable continua si el investigador registra la estatura reportada por el instrumento de medición hasta el límite de precisión de éste. Por ejemplo. 173.345 cm)

Las escalas de tipo numérico pueden tener una de dos características: las **escalas de intervalo** y las **escalas de razón**.

2.1.4. Escalas de intervalo

Son aquellas en las que el **cero es convencional o arbitrario** y pueden existir cantidades negativas.

Un ejemplo de este tipo de escalas es la de los grados Celsius o centígrados que se usan para medir la temperatura. En ella el cero es el punto de congelación del agua y, sin embargo, existen temperaturas más frías que se miden mediante números negativos. En estas escalas podemos hacer **comparaciones** por medio de diferencias o de sumas. Podemos decir, por ejemplo, que hoy la temperatura del agua de una alberca está cuatro grados más fría que ayer; pero no se pueden hacer comparaciones por medio de multiplicaciones, divisiones o porcentajes porque este tipo de comparaciones implica divisiones y éstas no tienen sentido en las escalas de intervalo. Si hoy la temperatura ambiente es, por ejemplo, de diez grados, no tiene sentido decir que hace el doble de frío que ayer, pues ayer estaba a veinte grados.

2.1.5. Escalas de razón

Son aquellas en las que el **cero absoluto sí existe**.

Tal es el caso de los grados Kelvin, para medir temperaturas, o muchas medidas que utilizamos en nuestra vida cotidiana. La estatura de las personas medida en centímetros, por ejemplo, está medida en una escala de razón pues sí existe el cero absoluto (nadie tiene estatura negativa). En el caso de las estaturas sí se puede decir que alguien mide el doble (o la mitad) que otra persona cualquiera.

La mayor parte de las herramientas que se aprenden en este curso son válidas para escalas numéricas, otras lo son para escalas ordinales y unas pocas (muchas de las que se ven en el tema de estadística no paramétrica) sirven para todo tipo de escalas.

2.1.6. Uso de computadoras en estadística

Algunas de las técnicas que se ven en este curso, y muchas que se ven en cursos más avanzados de estadística, requieren un conjunto de operaciones matemáticas que, con frecuencia, no son demasiado difíciles desde el punto de vista conceptual, pero sí son considerablemente laboriosas por el volumen de cálculos que conllevan. Por ello, **las computadoras**, con su gran capacidad para el manejo de grandes volúmenes de información son un **gran auxiliar**.

Existen herramientas de uso general como el **Excel o Lotus** que incluyen algunas funciones estadísticas y son útiles para muchas aplicaciones. Sin embargo, si se desea estudiar con mayor profundidad el uso de técnicas más avanzadas es importante contar con herramientas específicamente diseñadas para el trabajo estadístico.

Existen diversos paquetes de software en el mercado que están diseñados específicamente para ello. Entre otros se encuentran el SPSS y el SAS. Recomendamos al estudiante que ensaye el manejo de estas herramientas.

2.1.7. Principales elementos de las tablas

A continuación se presenta una tabla sencilla, tomada de un ejemplo hipotético. En ella se examinan sus principales elementos y se expresan algunos conceptos generales sobre ellos.

Título. Todas las tablas deben tener un título para que el lector sepa el asunto al que se refiere.

Encabezado. Se refiere a las categorías de datos que se manejan dentro de la propia tabla.

Cuerpo. En él se encuentran los datos propiamente dichos.

Fuente de información. Si los datos que se encuentran en la tabla no fueron obtenidos por el autor del documento en el que se encuentra la misma, es importante indicar de qué parte se obtuvo la información que allí se encuentra.

Estudiantes de la FCA que trabajan

Porcentajes por semestre de estudio* ← **Título**

Semestre que estudian	**Porcentaje**	
	Hombres	**Mujeres**
1	20	15
2	22	20
3	25	24
4	33	32
5	52	51
6	65	65
7	70	71
8	87	88
9	96	95

*Fuente: Perez José, "El trabajo en la escuela", Editorial Académica, México, 19XX

Encabezado

Cuerpo de la Tabla

Fuente de información

Cuadro 2.2. Tabla sencilla de datos

Independientemente de los principales elementos que puede tener una tabla, existen diversas maneras de presentar la información en ellas. No existe una clasificación absoluta de la presentación de las diferentes tablas, dado que, al ser una obra humana, se pueden inventar diversas maneras de presentar información estadística. No obstante lo anterior, se puede intentar una clasificación que nos permita entender las principales presentaciones.

2.1.8. Tablas simples

Relaciona una columna de categorías con una o más columnas de datos, sin más elaboración.

A continuación se presenta un ejemplo de una tabla simple.

FCA. Maestros de las distintas coordinaciones que han proporcionado su correo electrónico	
Coordinaciones	# de maestros
Administración Básica	23
Administración Avanzada	18
Matemáticas	34
Informática	24
Derecho	28
Economía	14

Cuadro 2.2. Tabla simple de datos

2.1.9. Tablas de frecuencias

Es un arreglo rectangular de información en el que las columnas representan diversos conceptos, dependiendo de las intenciones de la persona que la elabora, pero que tiene, siempre, en una de las columnas, información sobre el número de veces (frecuencia) que se presenta cierto fenómeno.

La siguiente tabla es un ejemplo de esta naturaleza. En ella, la primera columna representa las **categorías** o clases, la segunda las **frecuencias** llamadas **absolutas** y la tercera las **frecuencias relativas**. Esta última columna recibe esa denominación porque los datos están expresados en relación con el total de la segunda columna. Las frecuencias relativas pueden expresarse en porcentaje, tal como en nuestro ejemplo, o en absoluto (es decir, sin multiplicar los valores por 100). Algunos autores llaman al primer caso "frecuencia porcentual" en lugar de frecuencia relativa.

Deportes Batista, S.A. de C.V. Número de bicicletas vendidas por tienda		
Primer trimestre de 20XX		
Tienda	**Unidades**	**Porcientos (%)**
Centro	55	29.1
Polanco	45	23.8
Coapa	42	22.2
Tlalnepantla	47	24.9
Totales	**189**	**100.0**

Cuadro 2.3. Tabla de frecuencias relativas

2.1.10. Tablas de doble entrada

En algunos casos, se quiere presentar la información con un mayor detalle. Para ello se usan las tablas de doble entrada. Se llaman así porque la información se clasifica por medio de dos criterios en lugar de utilizar solamente uno. A continuación se presenta un ejemplo de tabla de doble entrada.

	Deportes Bicicletas Primer	Batista, S.A. vendidas por trimestre de	de C.V. modelo y tienda 20XX		
	Infantil	**Carrera**	**Montaña**	**Turismo**	**Total**
Centro	13	14	21	7	**55**
Polanco	10	14	11	10	**45**
Coapa	12	11	17	2	**42**
Tlalnepantla	9	8	13	17	**47**
Totales	**44**	**47**	**62**	**36**	**189**

Cuadro 2.4. Tabla de doble entrada

Podemos observar que esta tabla, en la columna de total presenta una información idéntica a la segunda columna de la tabla de frecuencias. Sin embargo, en el cuerpo de la tabla se desglosa una información más detallada, pues nos ofrece datos sobre los modelos de bicicletas, que la tabla de frecuencias no teníamos.

2.1.11. Tablas de contingencia

Un problema frecuente es el de definir la independencia de dos métodos para clasificar eventos.

Supongamos que una empresa que envasa leche desea clasificar los defectos encontrados en la producción tanto por tipo de defecto como en el turno (matutino, vespertino o nocturno) en el que se produjo el defecto. Lo que se desea estudiar es si se presenta el caso (la contingencia y de allí el nombre) de que exista una relación entre ambas clasificaciones. ¿Cómo se comporta la proporción de cada tipo de defecto de un turno a otro?

En el ejemplo de la empresa que quiere hacer este tipo de trabajo se encontró un total de 312 defectos en cuatro categorías distintas: volumen, empaque, impresión y sellado. La información encontrada se resume en la siguiente tabla.

	Lechería La Laguna, S,A. Tabla de contingencia en laque se clasifican los defectos del empaque de leche por tipo de defecto y por turno.									
Turno	**Volumen**		**Empaque**		**Impresión**		**Sellado**		**Totales**	
Matutino	16	5.13	22	7.05	46	14.74	13	4.17	**97**	**31.09**
Vespertino	26	8.33	17	5.45	34	10.90	5	1.60	**82**	**26.28**
Nocturno	33	10.58	31	9.94	49	15.71	20	6.41	**133**	**42.63**
Totales	**75**	**24.04**	**70**	**22.44**	**129**	**41.35**	**38**	**12.18**	**312**	**100.00**
	Los números en rojo representan los porcentajes									

Cuadro 2.5. Tabla de contingencia

De la información de la tabla antecedente, podemos apreciar que el mayor porcentaje de errores se comete en el turno nocturno y que el área en la que la mayor proporción de defectos se da es la de impresión. Como vemos, la clasificación cruzada de una tabla de contingencia puede llevarnos a obtener conclusiones interesantes que pueden servir para la toma de decisiones..

2.2 Distribuciones de frecuencia

Una distribución de frecuencias o tabla de frecuencias no es más que la presentación tabular de las **frecuencias conque ocurre cada característica** (subclase) en las que ha sido divida una variable. Esta característica puede estar determinada por una cualidad o un intervalo; por lo tanto, la construcción de un cuadrado de frecuencia o tabla de frecuencias puede desarrollarse tanto para una variable cuantitativa como para una variable cualitativa.

2.2.1. Distribución de frecuencias para variables cuantitativas

Recordemos que las variables cuantitativas o métricas pueden ser de dos tipos: continuas o discretas.

Cuando la variable es **continua**, la construcción de una **tabla de frecuencia presenta** como su punto de mayor importancia la determinación del **número de intervalos o clases** que la formarán.

Una clase o intervalo de clase es el elemento en la tabla que permite condensar en mayor grado un conjunto de datos con el propósito de hacer un resumen de ellos. El número de casos o mediciones que quedan dentro de un intervalo reciben el nombre de frecuencia del intervalo, que se denota generalmente como *Fi*. La diferencia entre el extremo mayor y el menor del intervalo se llama longitud o ancho del intervalo.

La elaboración de una tabla de distribución de frecuencias se complementa, generalmente, con el cálculo de los siguientes elementos:

Elemento	Descripción
Marca de clase	Está constituida por el punto medio del intervalo de clase. Para calcularla es necesario sumar los dos límites del intervalo y dividirlos entre dos
Frecuencia acumulada de la clase	Se llama así al número resultante de sumar la frecuencia de la clase i con la frecuencia de las clases que la anteceden. Se denota generalmente como fi. La última clase o intervalo en la tabla contiene como frecuencia acumulada el total de los datos.
Frecuencia relativa de la clase	Es el cociente entre la frecuencia absoluta (fi) de la clase i y el número total de datos. Esta frecuencia muestra la proporción del número de casos que se han presentado en el intervalo "i" respecto al total de casos en la investigación.
Frecuencia acumulada relativa de la clase	Es el cociente entre la frecuencia acumulada de la clase i y el número total de datos. Esta frecuencia muestra la proporción del número de casos que se han acumulado hasta el intervalo i respecto al total de casos en la investigación

CUADRO. 2.6. Elementos de una tabla de frecuencias

En el caso de variables discretas, la construcción de una tabla de distribución de frecuencias sigue los lineamientos establecidos para una variable continua con la salvedad de que en este tipo de tablas no existen intervalos ni marcas de clase, lo cual simplifica la construcción de la tabla.

La construcción de tablas de frecuencia para variables cualitativas o no métricas requieren sólo del conteo del número de elementos o individuos que se encuentran dentro de cierta cualidad o bien dentro de determinada característica.

2.2.2. Cuadros estadísticos

El resultado del proceso de tabulación o condensación de datos se presenta en lo que en estadística se llaman **cuadros estadísticos**, también conocidos con el nombre incorrecto de tablas estadísticas, producto de la traducción inglesa.

Con base en el uso que el investigador le dé a un cuadro estadístico, éstos pueden ser clasificados en **dos tipos**: cuadros de trabajo y cuadros de referencia.

Los **cuadros de trabajo** son aquellos estadísticos que contienen datos producto de una tabulación. En otras palabras, son cuadros depositarios de datos que son utilizados por el investigador para obtener, a partir de ellos, las medidas estadísticas requeridas.

Los **cuadros de referencia** tienen como finalidad ayudar al investigador en el análisis formal de las interrelaciones que tienen las variables que están en estudio, es decir, contienen información ya procesada de cuadros de trabajo (proporciones, porcentajes, tasas, coeficientes, etc.)

La construcción de cuadros estadísticos de trabajo o de cuadros de referencia requiere prácticamente de los mismos **elementos en su elaboración**, pues ambos presentan las mismas características estructurales, por lo que los elementos que a continuación se describen deberán ser utilizados en la conformación de éstos indistintamente.

1. **Número del cuadro**. Es el primer elemento de todo cuadro estadístico. Tiene como objeto permitir una fácil y rápida referencia al mismo.

2. **Titulo**. Es el segundo elemento del cuadro estadístico. En él se deberá indicar el contenido del cuadro, su circunscripción espacial, el periodo o espacio temporal y las unidades en las que están expresados los datos.

3. **Nota en el título (encabezado).** Elemento complementario del título. Se emplea sólo en aquellos cuadros en los que se requiere proporcionar información relativa al cuadro como un todo o a la parte principal del mismo.

4. **Casillas cabecera**. Contienen la denominación de cada característica o variable que se clasifica.

5. **Columnas**. Son las subdivisiones verticales de las casillas cabecera. Se incluyen tantas columnas en una casilla cabecera como categorías le correspondan.

6. **Renglones**. Son las divisiones horizontales que corresponden a cada criterio en que es clasificada una variable.

7. **Espacio entre renglones**. Tienen por objeto hacer más clara la presentación de los datos, facilitando así su lectura.

8. **Líneas de cabecera**. Son las líneas que se trazan para dividir las casillas de cabecera de los renglones.

9. **Cabeza del cuadro**. Está formada por el conjunto de casillas cabecera y encabezados de columnas.

10. **Casillas**. Es la intersección que forman cada columna con cada renglón en el cuadro. Las casillas contienen datos o bien los resultados de cálculos efectuados con ellos.

11. **Cuerpo del cuadro**. Está formado por todos los datos sin considerar la cabeza del cuadro y los renglones de totales.

12. **Renglón de totales**. Es un elemento opcional en los cuadros estadísticos.

13. **Línea final de cuadro**. Es la línea que se traza al final del cuerpo del cuadro y en su caso al final del renglón de totales.

14. **Notas al pie del cuadro**. Se usan para calificar o explicar un elemento particular en el cuadro que presente una característica distinta de clasificación.

15. **Fuente**. Es el último elemento de un cuadro estadístico. Tiene por objeto indicar el origen de los datos.

En un ejemplo de ausencias a clase en una determinada semana, la tabla correspondiente será:

Ausencia (valores de variable)	**Número de alumnos (Frecuencia)**
0	11
1	4
2	2
3	2
4	1
5	1

Cuadro 2.7. Tabla de frecuencias

La presentación de datos cualitativos suele hacerse de forma análoga a la de las variables, indicando las distintas clases o atributos observados y sus frecuencias de aparición, tal como se recoge en la tabla siguiente sobre color de pelo en un grupo de 100 turistas italianos:

Color de pelo	Número de personas
Negro	60
Rubio	25
Castaño	15

Cuadro 2.8. Tabla de datos en clases

2.2.3. Frecuencias absolutas y relativas

La **frecuencia absoluta** es el número que indica cuántas veces el valor correspondiente de una variable de medición (dato) se presenta en la muestra y también se le conoce simplemente como frecuencia de ese valor de "x" (dato) en la muestra.

Si ahora dividimos la frecuencia absoluta entre el tamaño de la muestra "n" obtenemos la **frecuencia relativa** correspondiente.

A manera de teorema podemos decir que la frecuencia relativa es por lo menos igual a 0 y cuando más igual a 1. Además, la suma de todas las frecuencias relativas en una muestra siempre es igual a 1.

2.3 Presentación gráfica de datos

Es importante construir gráficas de diversos tipos que permitan explicar más fácilmente el comportamiento de los datos en estudio. Una **gráfica** permite **mostrar, explicar, interpretar y analizar** de manera sencilla, clara y efectiva los **datos** estadísticos mediante formas geométricas tales como líneas, áreas, volúmenes, superficies, etcétera. Las gráficas permiten además la comparación de magnitudes, tendencias y relaciones entre los valores que adquiere una variable.

"Un dibujo vale más que diez mil palabras", dice el viejo proverbio chino,;este principio es tan cierto con respecto a números como a dibujos. Frecuentemente, es posible resumir toda la información importante que se tiene de una gran cantidad de datos en un dibujo sencillo. Así, uno de los métodos más ampliamente utilizados para representar datos es mediante gráficas.

2.3.1. Histogramas y polígonos de frecuencias

Un **histograma de frecuencias** es un gráfico de rectángulos que tiene su base en el eje de las abscisas (eje horizontal o eje de las equis), con anchura igual cuando se trata de representar el comportamiento de una variable discreta y anchura proporcional a la longitud del intervalo cuando se desea representar una variable continua. En este último caso, el punto central de la base de los rectángulos equivale al punto medio de cada clase.

Las alturas de los rectángulos ubicadas en el eje de las ordenadas (de las Y o eje vertical) corresponde a las frecuencias de las clases. El área de los rectángulos así formados es proporcional a las frecuencias de las clases.

Los histogramas de frecuencias pueden **construirse** no sólo con las **frecuencias absolutas**, sino también con **las frecuencias acumuladas** y las **frecuencias relativas**. En este último caso el histograma recibe el nombre de Histograma de frecuencias relativas, Histograma de porcentajes o Histograma de proporciones, según el caso.

El histograma es similar al diagrama de barras o rectángulos, aunque con una diferencia importante: mientras que en los diagramas sólo estamos interesados en las alturas de las barras o rectángulos, en el histograma son fundamentales tanto la altura como la base de los rectángulos, haciendo el área del rectángulo proporcional a su frecuencia.

Como ya se indicó previamente, las variables cualitativas no tienen intervalos de clase por carecer éstos de sentido. Tampoco en ellas se calcula la frecuencia acumulada; por lo tanto, para las variables cualitativas sólo existe la construcción de los histogramas de frecuencia absoluta y los histogramas porcentuales o de frecuencia relativa. Para variables cualitativas no existe polígono de frecuencias.

2.3.2. Pasos a seguir para la elaboración de un diagrama de frecuencias (o polígono de frecuencias) y un histograma.

Considere el siguiente conjunto de datos:

8.9	8.3	9.2	8.4	9.1
8.6	8.9	9.1	8.8	8.8
8.8	9.1	8.9	8.7	8.8
8.9	9.0	8.6	8.7	8.4
8.6	9.0	8.8	8.9	9.1
9.4	9.0	9.2	9.1	8.8
9.1	9.3	9.0	9.2	8.8
9.7	8.9	9.7	8.3	9.3
8.9	8.8	9.3	8.5	8.9
8.3	9.2	8.2	8.9	8.7
8.9	8.8	8.5	8.4	8.0
8.5	8.7	8.7	8.8	8.8
8.3	8.6	8.7	9.0	8.7
8.4	8.8	8.4	8.6	9.0
9.3	8.8	8.5	8.7	9.6
8.5	9.1	9.0	8.8	9.1
8.6	8.6	8.4	9.1	8.5
9.1	9.2	8.8	8.5	8.3
9.3	8.6	8.7	8.7	9.1
8.8	8.7	9.0	9.0	8.5
8.5	8.8	8.9	8.2	9.0
9.0	8.7	8.7	8.9	9.4
8.3	8.6	9.2	8.7	8.7
8.7	9.7	8.9	9.2	8.8
8.3	8.6	8.5	8.6	9.7

9.7	9.7	9.7	9.2	9.2	máximo = 9.7
8.3	8.3	8.2	8.2	8.0	mínimo = 8.0

Paso 1. Cuente el número de datos en la población o muestra; en este caso son 125 lecturas, por lo tanto, n=125.

Paso 2. Calcule el rango de los datos (R).

Para determinar el rango de los datos lo único que se debe hacer es encontrar el número mayor y el número menor de las 125 lecturas que se tienen en la tabla. Para hacer esto, el doctor Kaouru Ishikawa recomendó lo siguiente:

Se toman filas o columnas, en este caso columnas, y se identifica mínimo y máximo por columna, creando dos filas. En este caso, una de máximos y otra de mínimos. Estos dos números podrán ser identificados como el *máximo y mínimo* de las lecturas en la tabla, en este caso: MÁX = 9.7 y MÍN = 8.0, por lo que el rango se calcula de la siguiente forma; R=MÁX–MÍN. R=9.7-8.0=1.7, así que R=1.7 unidades.

Paso 3. Determine el número de clases, celdas o intervalos.

En la construcción de un diagrama de frecuencias o de un histograma es necesario encasillar las lecturas. Aunque existe un fórmula matemática para el cálculo del número de clases que debe tener el diagrama de frecuencias o el histograma, que depende del número de elementos en la población, hay un método práctico por seguir que dice que el número de clases en un diagrama de frecuencias o histograma no debe ser menos de 6 ni más de 15.

Se llamará “Q” a la cantidad de clases que tendrá el histograma; se recomienda lo siguiente:

Número de lecturas	Número de clases
< 50	6 - 8
50 - 100	9 – 11
100 - 250	8 – 13
> 250	10 - 15

Cuadro 2. 9. Tabla de clases.

Paso 4. Determinación de “c” (ancho de celda, clase o intervalo).

Para este caso utilizamos la siguiente fórmula:

$$C = \frac{R}{Q} = \frac{1.7}{10} = 0.17$$

Generalmente es necesario redondear “c” para trabajar con un número cómodo en las gráficas y en los cálculos. En esta ocasión daremos un valor de c=0.20 unidades que debe mantenerse constante a lo largo del rango, que en este caso es de R=1.7

Paso 5. Fijación de los límites de clase.

En muchos casos esto sucede automáticamente y depende de la costumbre. Por ejemplo, si se le pregunta su edad a una persona, ésta contestará con el número de años que tiene. En este caso, el ancho de clase es automáticamente de un año aunque la persona haya cumplido años ayer o hace 11 meses. En otras ocasiones, la resolución en los instrumentos de medición es la que determina el ancho de clase aun cuando es necesario dar una regla general que se mantenga para lograr una normalización del histograma. En el ejemplo, la lectura menor fue de 8.0, por lo que se podría fijar el ancho de clase “c” y repetir esta operación hasta que todos los valores de la tabla queden contenidos.

Paso 6. Construcción de la distribución de frecuencias:

Clase	Límite de clase	Marca de clase	**Frecuencia**	total
1	**8.00-819**	**8.1**	**I**	**1**
2	**8.20-839**	**8.3**	**IIIII IIII**	**9**
3	**8.40-859**	**8.5**	**IIIII IIIII IIIII I**	**16**
4	**8.60-879**	**8.7**	**IIIII IIIII IIIII IIIII IIIII II**	**27**
5	**8.80-8.99**	**8.9**	**IIIII IIIII IIIII IIIII IIIII IIIII I**	**31**
6	**9.00-9.19**	**9.1**	**IIIII IIIII IIIII IIIII III**	**23**
7	**9.20-9.39**	**9.3**	**IIIII IIIII II**	**12**
8	**9.40-9.59**	**9.5**	**II**	**2**
9	**9.60-9.79**	**9.7**	**IIII**	**4**
10	**9.80-9.99**	**9.9**		**0**
			Suma de "f" = N =	**= 125**

Cuadro 2. 10. Tabla de distribución de frecuencias

Al graficar los datos anteriores obtenemos la siguiente gráfica:

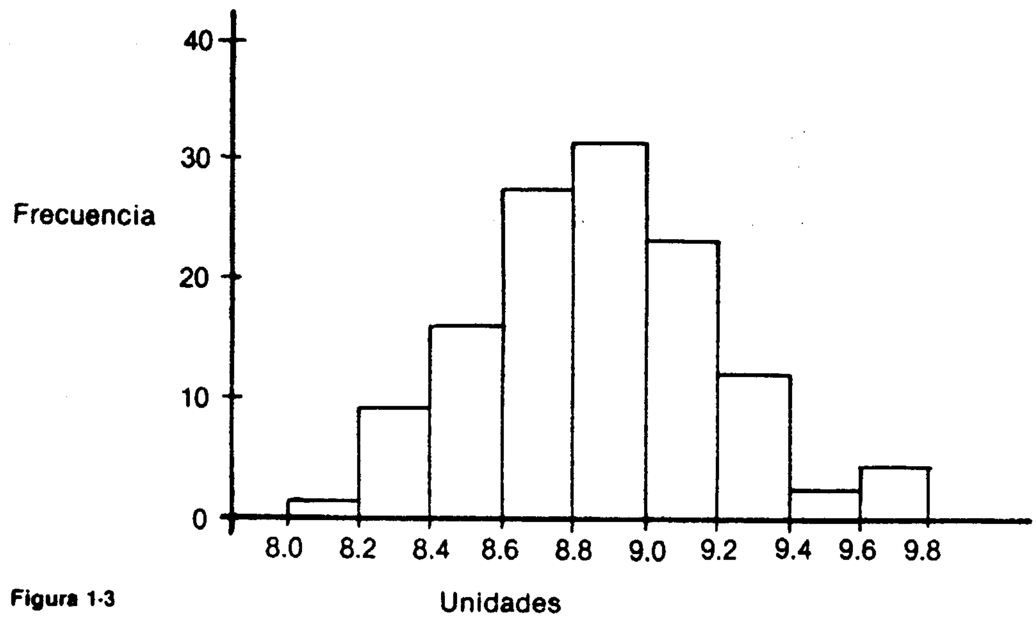


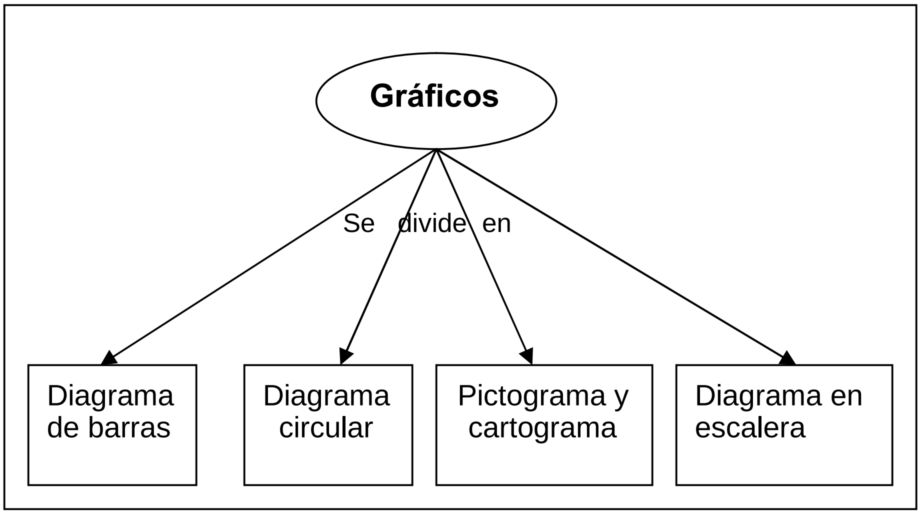
FIGURA 2.1. Histograma de frecuencias

La forma más habitual de representar la información contenida en una tabla es a partir de un sistema de ejes cartesianos. Hay, no obstante, otras formas de representar datos, como posteriormente veremos, que están básicamente orientadas a caracteres no cuantitativos o atributos. Para hacer más clara la exposición de las diferentes representaciones gráficas, distinguiremos las referentes a dos tipos de distribuciones:

- Distribuciones sin agrupar
- Distribuciones agrupadas en intervalos

2.3.3. Gráficas para distribuciones de frecuencias no agrupadas

Para representar este tipo de distribuciones, los gráficos más utilizados son:



Cuadro 2.11. División de los gráficos.

- **a)** El diagrama de barras, que se emplea para distribuciones tanto de variables estadísticas como de atributos.
- **b)** El diagrama circular, que es el más comúnmente utilizado para distribuciones de atributos.
- **c)** El pictograma y el cartograma.
- **d)** Diagrama en escalera, empleado para frecuencias acumuladas.

a) Diagrama de barras. Es la más sencilla de las gráficas y consiste en representar datos mediante una barra o columna simple, la cual puede ser colocada horizontal o verticalmente.

Este gráfico permite comparar las proporciones que guardan cada una de las partes con respecto al todo, por lo que pueden construirse usando valores absolutos, proporciones o bien porcentajes. Suelen utilizarse cuando se comparan gráficamente las distribuciones de iguales conceptos en dos o más periodos.

Asimismo, constituye la representación gráfica más utilizada, por su capacidad para adaptarse a numerosos conjuntos de datos.

La forma de elaborar estos diagramas es la siguiente:

1. Sobre unos **ejes de coordenadas** se representan en las abscisas los diferentes valores de la variable y en las ordenadas las frecuencias.
2. Sobre cada **valor de la variable** se levanta una barra cuya altura sea la frecuencia correspondiente.
3. Esta representación será un conjunto de barras; por ello se denomina diagrama de barras.

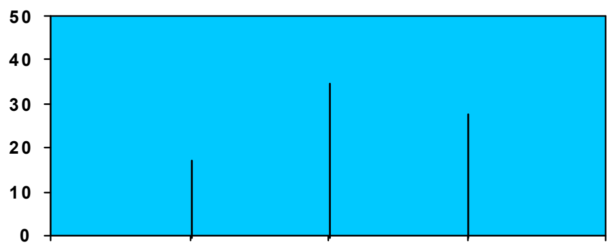


Figura 2.2. Diagrama de barras

A partir de este diagrama, es fácil darse cuenta de en qué valores de la variable se concentra la mayor parte de las observaciones.

Una variante de este diagrama, tal vez más utilizada por ser más ilustrativa, es el **diagrama de rectángulos**. Consiste en representar en el eje de las abscisas los valores de la variable y en el de las ordenadas las frecuencias. Pero ahora, sobre cada valor de la variable se levanta un rectángulo con base constante y altura proporcional a la frecuencia absoluta.

Aunque los datos gráficos son equivalentes, generalmente se opta por el de rectángulo por ser, a simple vista, más ilustrativo.

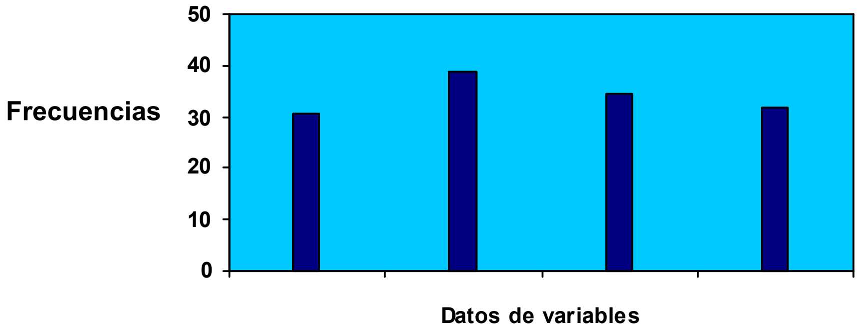
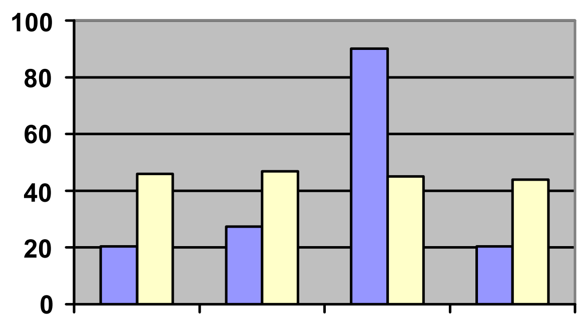


Figura 2.3 Diagrama de rectángulos

Además, el diagrama de rectángulos es especialmente útil cuando se desea comparar, en un mismo gráfico, el comportamiento del fenómeno en dos situaciones o ámbitos distintos.

Pues bien, la adopción del diagrama de rectángulos que utilice dos colores distintos, permite ofrecer una visión simplificada y conjunta de lo que ocurre en ambos casos por tratar.

Ejemplos de análisis comparativo pueden ser representados con rectángulos de dos tonos.



Este	20.4	27.4	90	20.4
Norte	45.9	46.9	45	43.9

Figura 2.4 Diagrama de rectángulos

b) Diagrama circular. Esta representación gráfica es especialmente adecuada en aquellos casos en que se desea que los datos estadísticos lleguen a todo tipo de persona, incluso a las que no tienen por qué tener una formación científica.

Este tipo de diagrama muestra la importancia relativa de las diferentes partes que componen un total. La forma de elaborarlo es la siguiente:

- Se toma como total el círculo.
- A continuación, se divide éste en tantas partes como componentes haya; el tamaño de cada una de ellas será proporcional a la importancia relativa de cada componente. Más concretamente, como el círculo tiene 360ª, éstos se reparten proporcionalmente a las frecuencias absolutas de cada componente.

La ventaja intrínseca de este tipo de representaciones no debe hacer olvidar que plantea ciertas desventajas que enumeramos a continuación:

1. Requiere cálculos adicionales.
2. Es más difícil comparar segmentos de un círculo que comparar alturas de un diagrama de barras.
3. No da información sobre las magnitudes absolutas, a menos que las incorporemos en cada segmento.

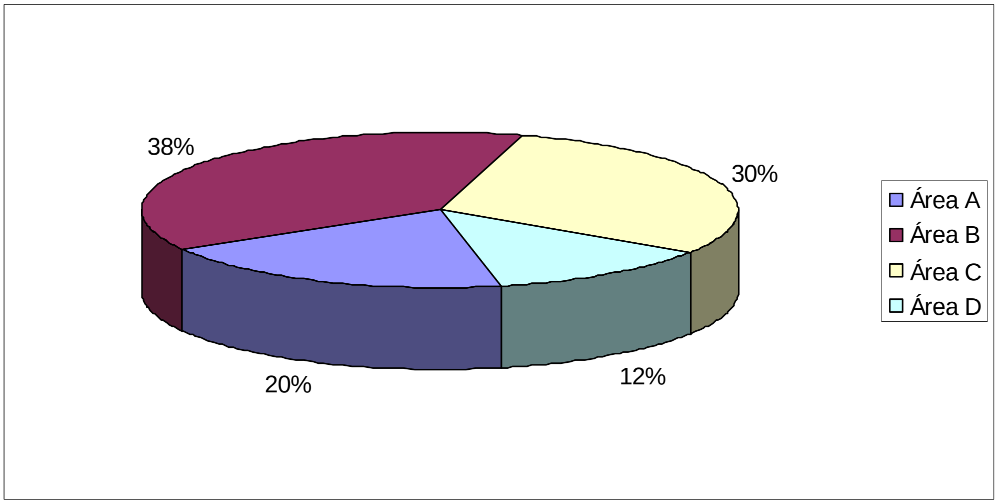


Figura 2.5. Grafica circular o pastel

c) Pictograma. Es otra forma de representar distribuciones de frecuencias. Consiste en tomar como unidad una silueta o símbolo que sea representativo del fenómeno que se va a estudiar.

Por ejemplo:

 = 100 viviendas

Y para representar 300 viviendas

 = 300 viviendas

c1) Cartograma. Son especialmente útiles en estudios de carácter geográfico. La forma de construirlos es la siguiente: se colorea o se raya con colores e intensidades diferentes los distintos espacios o zonas (que pueden ser comunidades autónomas, provincias, ríos, etc.) en función de la mayor o menor importancia que tenga la variable o atributo en estudio.

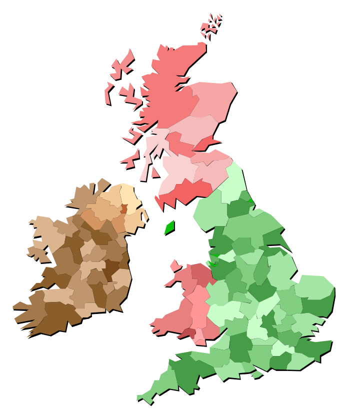


Figura 2.6.Fuente: Revista *Expansión*, No. 852 (octubre 30 del 2002), p. 69.

d) Diagrama en escalera. Su nombre responde a que la representación tiene forma de escalera. Se utiliza para representar frecuencias acumuladas. Su construcción es similar a la del diagrama de barras; y se elabora de la forma siguiente:

- En el eje de las abscisas se miden los valores de la variable o las modalidades del atributo; en el de las ordenadas, las frecuencias absolutas acumuladas.
- Se levanta, sobre cada valor o modalidad, una barra, cuya altura es su frecuencia acumulada.
- Por último, se unen mediante líneas horizontales cada frecuencia acumulada a la barra de la siguiente.
- Los pasos anteriores conducen a la escalera; la última ordenada corresponderá al número total de observaciones.

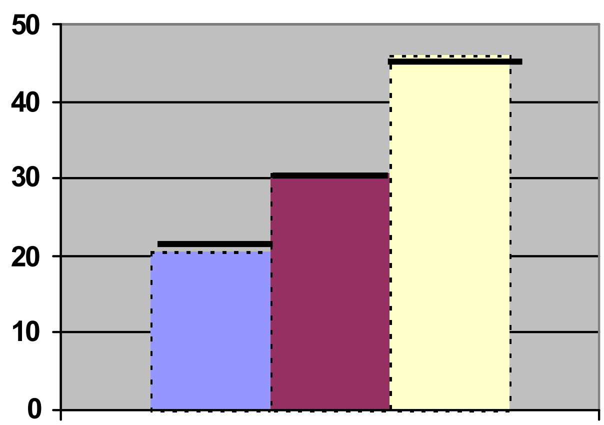


Figura 2.7. Diagrama en escalera

2.3.4. Gráficas para distribuciones de frecuencias agrupadas en clases o eventos

Para distribuciones agrupadas en intervalos existen básicamente tres tipos de representaciones gráficas: el histograma, el polígono de frecuencias y las ojivas.

Polígono de frecuencias. Es un gráfico de línea que se construye, sobre el sistema de coordenadas cartesianas, al colocar sobre cada marca de clase un punto a la altura de la frecuencia asociada a esa clase; posteriormente, estos puntos se unen por segmentos de recta. Para que el polígono quede cerrado se debe considerar un intervalo más al inicio y otro al final con frecuencias cero.

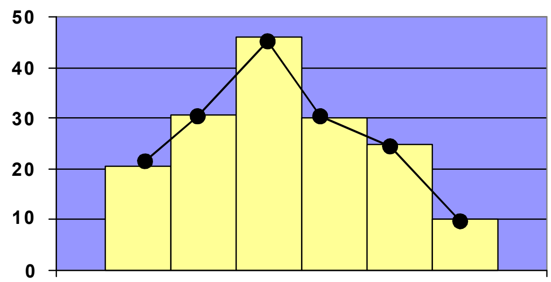


Figura 2.8. Polígono de frecuencias

Ojivas. Si en lugar de frecuencias absolutas utilizamos las acumuladas, obtendremos, en vez del histograma, una representación gráfica en forma de línea creciente que se conoce con el nombre de ojiva. Estos gráficos son especialmente adecuados cuando se tiene interés en saber cuántas observaciones hay en las zonas izquierda o inferior del límite superior de cualquier intervalo.

La ojiva es el polígono que se obtiene al unir por segmentos de recta los puntos situados a una altura igual a la frecuencia acumulada a partir de la marca de clase, en la misma forma en que se realizó para construir el polígono de frecuencias.
La ojiva también es un polígono que se puede construir con la frecuencia acumulada relativa.

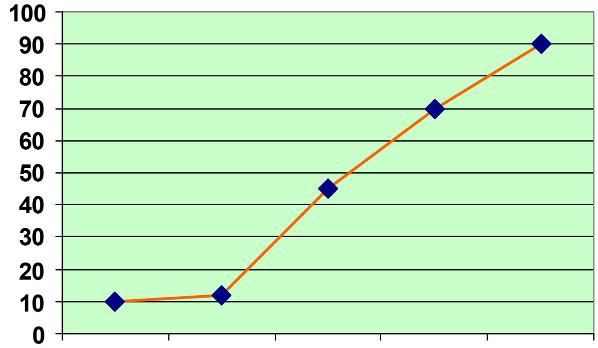


Figura 2.9 Ojivas

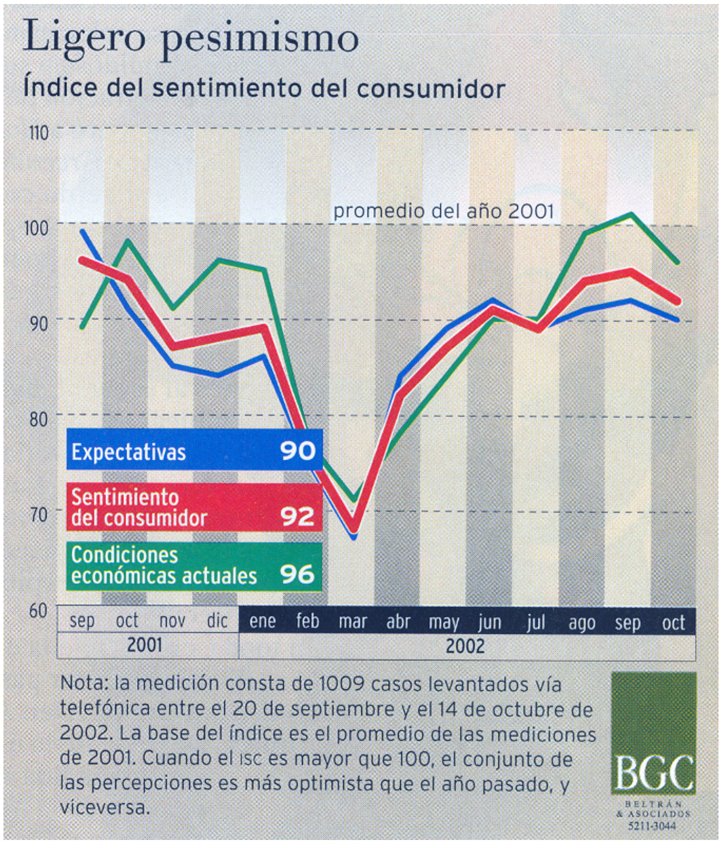
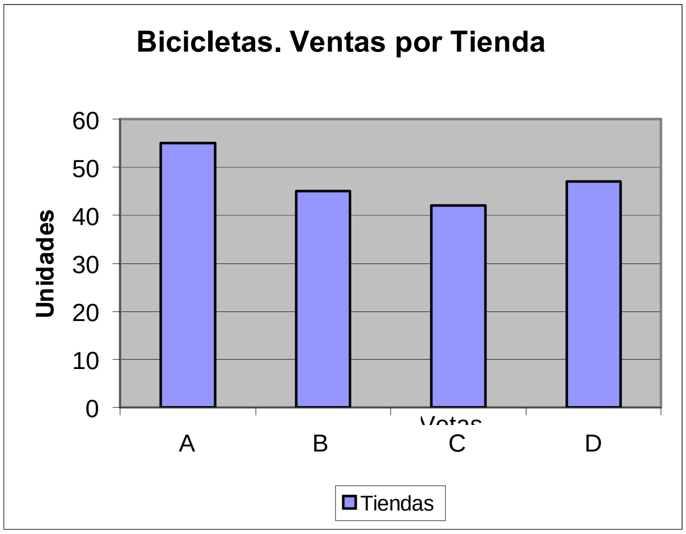


FIGURA 2.10. Fuente: Revista *Expansión*, No. 852, (octubre 30 del 2002) p. 14.

En los siguientes ejemplos se observan los tipos de gráficas estudiadas:

Columnas.

Este tipo de gráficas nos permite visualizar información de categorías con mucha facilidad.



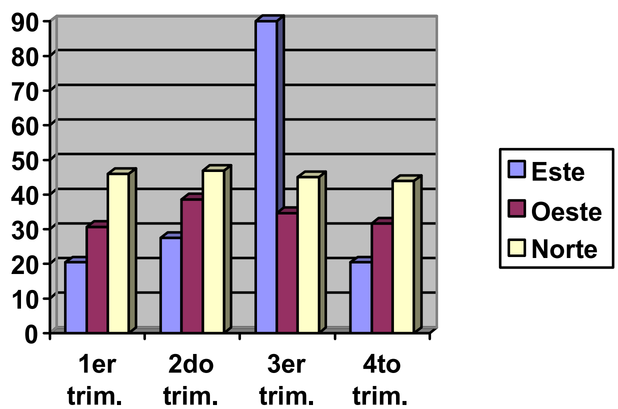


Figura 2.11 Diagrama de columnas

Barras.

Tiene la misma utilidad que el de columnas, pero en este caso con un formato horizontal.

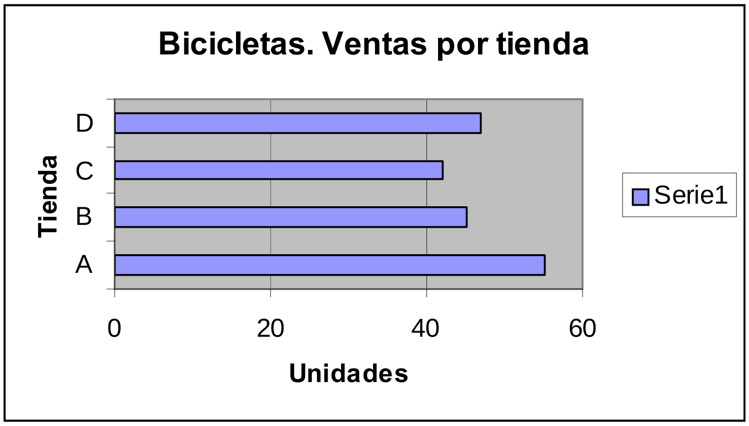


FIGURA 2.12. Diagrama de barras

Circular.

Presenta de una manera muy objetiva las proporciones que tiene cada una de las categorías en el total, como si fueran las tajadas de un pastel.

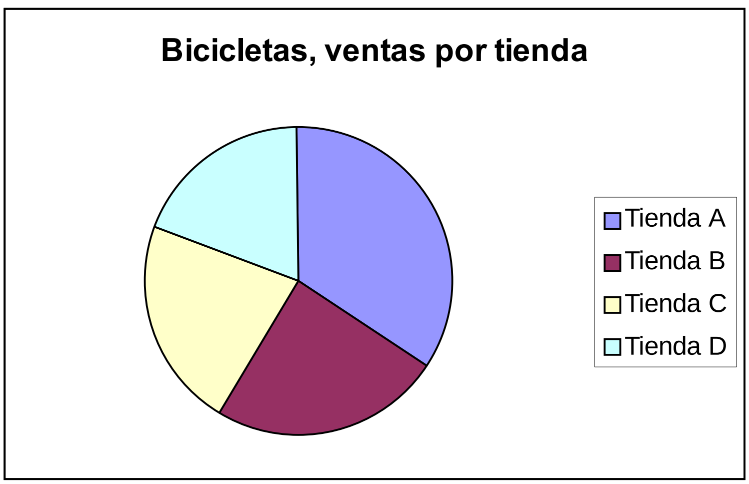


FIGURA 2.13. Diagrama circular

2.4. Medidas de tendencia central

Hemos visto que tanto las tablas como las gráficas pueden sernos útiles para representar y comprender información numérica. Existen, sin embargo, circunstancias en las que ni las tablas ni las gráficas nos dan información suficiente para tomar decisiones. En esos casos debemos procesar nuestra información de diversas maneras para obtener datos de la ésta. A estas medidas se les llama "parámetros" de acuerdo con lo visto en la unidad 1. Se dividen en **medidas de posición y medidas de dispersión**.

2.4.1. Medidas de posición

Son aquellas que nos definen (o nos informan) del valor de datos que ocupan lugares importantes en nuestra distribución de datos; las podemos dividir de la siguiente forma: a unas en medidas de tendencia central y a otras medidas de posición.

Las **medidas de tendencia central** son aquellas que nos indican cuál es el dato más representativo de una distribución. Por el contrario, las medidas de dispersión tienen el objetivo de localizar diversos puntos de interés; por ejemplo, el punto que divide la distribución en dos partes: a la izquierda (datos más pequeños) el 25% de la información y a la derecha (datos más grandes), el 75% de la información. A este punto se le denomina primer cuartil o Q1.

A continuación daremos las definiciones y algunos ejemplos de las medidas de tendencia central y concluiremos el apartado con las medidas de posición.

Las medidas de tendencia central que se contemplan en este material son: la media aritmética, la mediana y la moda.

Media aritmética

La media aritmética es el promedio que todos conocemos desde nuestros años de infancia. Se obtiene sumando todos los datos y dividendo el total entre el número de datos. La manera formal de expresar este concepto es la siguiente:

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

Esta fórmula nos dice que la media aritmética, que está representada por la letra griega μ, se obtiene sumando todos los datos a los que llamamos X subíndice i para, posteriormente, dividir el resultado entre “N”, que es el número total de datos con los que se cuenta.

Considere el siguiente ejemplo: Las calificaciones en los dos primeros semestres de un alumno que estudia la licenciatura en Administración se listan a continuación: 9, 10, 8, 8, 9, 7, 6, 10, 8, 8,7.

La media aritmética está dada por la siguiente expresión:

$$\mu = (9+10+8+8+9+7+6+10+8+8+7)/11$$

Haciendo las operaciones encontramos que la media aritmética es aproximadamente de 8.18.

Mediana

Es el valor que divide la distribución en dos partes iguales. Para obtenerla se deben ordenar los datos (puede ser de menor a mayor o viceversa, no importa) y se encuentra el dato medio. En el caso de las calificaciones del estudiante indicadas arriba, los datos ordenados tendrían el siguiente aspecto:

6, 7, 7, 8, 8, **8**, 8, 9, 9, 10 ,10

El dato que divide la distribución a la mitad se señala con una flecha. Este dato corresponde a la mediana. Como se puede ver a la izquierda del 8 encontramos cinco datos y, a su derecha encontramos otros cinco datos. Este dato es, entonces, el correspondiente a la mediana; así, Md=8.

Cuando en lugar de un número non de datos (como en nuestro ejemplo anterior), nos encontramos con un número par de observaciones, lo que se hace es promediar los dos datos medios. El procedimiento se muestra en el siguiente ejemplo:

Las ventas diarias de una pequeña tienda durante una corta temporada vacacional se consignan a continuación. Ya se ordenaron de menor a mayor para facilitar el trabajo posterior:

3,200; 3,500; 3,650: **3,720; 3,750**; 3,810; 3,850; 3,915

Puede verse fácilmente que no hay un dato central que divida la distribución en dos, por ello se toman los dos datos centrales y se promedian. En este caso la mediana es de 3,735, que es la media aritmética de los dos datos centrales.

Moda

Es el dato más frecuente de nuestro conjunto. En el caso de las calificaciones del estudiante el dato más frecuente es “8”, como se puede ver si repetimos nuestro conjunto de datos.

6, 7, 7, **8, 8, 8, 8**, 9, 9, 10 ,10.

En el caso de las ventas de la tienda, se puede ver que nos hay dos datos iguales; por lo mismo, este conjunto de datos no tiene moda.

Puede darse el caso, en conjuntos más grandes de datos, que el “honor” de ser el valor más frecuente sea compartido por dos datos. En ese caso se afirma que la distribución es **bimodal**, pues tiene dos modas. Algunos autores llegan a hablar de distribuciones **trimodales** e incluso más.

Cuartiles

Así como la mediana divide la distribución de nuestros datos en dos partes iguales, existen medidas de posición llamadas **cuartiles**. Hay tres cuartiles en cada distribución de datos; el **primer cuartil** o Q1 divide la distribución en dos partes: a la izquierda está la cuarta parte (de allí su nombre) o el 25% de los datos. El **segundo cuartil** o Q2 se asimila a la mediana y divide la distribución de nuestros datos en dos partes iguales. El **tercer cuartil** o Q3 hace la misma función, pues divide nuestra distribución de datos en dos partes, la parte izquierda agrupa al 75% de los datos más pequeños y la parte derecha el 25% de los datos más grandes. El siguiente esquema puede aclarar la situación de los cuartiles:

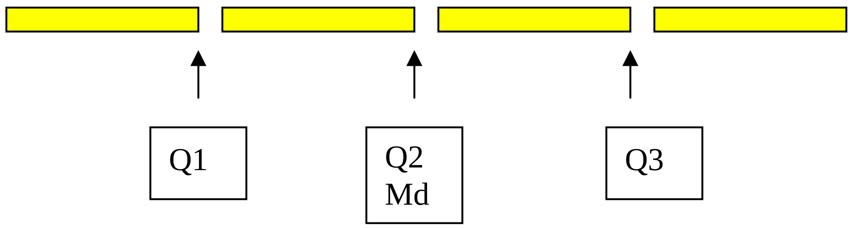


Figura 2.14. Posición de cuartiles

Cada una de las barras amarillas representa un 25% de los datos.

Un concepto parecido al de cuartiles se tiene con el de "**deciles**" y con el de "**percentiles**", pero en lugar de separar los datos en grupos del 25% lo hacen en grupos del 10% y del 1% respectivamente. Desde luego, para que los cuartiles, deciles y percentiles tengan algún sentido se requiere tener conjuntos grandes de datos.

Por ejemplo, no tiene ningún objeto hablar de percentiles si se tienen 14 datos. La manera de encontrar los cuartiles, deciles o percentiles sería, en teoría, la misma; es decir, alinear los datos de menor a mayor y contar cuál de ellos es el que cumple el requisito de dividir la distribución de la manera que queremos, pero este método es completamente impráctico, por lo que nos ocuparemos de su obtención cuando trabajemos datos agrupados.

2.5 Medidas de dispersión

Saber cuál es el dato central de una distribución es importante, pero también lo es saber qué tan concentrada o extendida está nuestra información. Por ejemplo, saber que una tienda tiene ingresos diarios medios de $10,000 es interesante, pero además es importante saber si todos los días esas ventas están muy cerca de los diez mil pesos o, en realidad, se alejan mucho. Enseguida damos los datos de dos tiendas que tienen la misma media de ventas diarias.

Tienda A. \$10,000; \$10500; \$11,000; \$9,000; \$9,500.
Tienda B. \$10,000; \$5,000; \$15,000; \$19,000; \$1,000

Es fácil observar que ambas tiendas tienen las mismas ventas medias (\$10,000). Sin embargo, en la tienda A la planeación de flujo de efectivo es más sencilla que en la tienda B. En la primera podemos contar con un flujo más o menos constante de efectivo que nos permite afrontar los compromisos diarios; en la segunda podemos tener un flujo muy abundante o casi nada. Eso nos lleva a tener que prever cómo invertir excedentes temporales y cómo cubrir faltantes en el corto plazo.

Las medidas que nos permiten cuantificar la dispersión de los datos son cuatro: el rango o recorrido, la varianza, la desviación estándar y el coeficiente de variación. A continuación definimos cada una de ellas.

Rango o recorrido

Es la diferencia entre el dato mayor y el dato menor. En el ejemplo de las tiendas sus rangos son:

Tienda A: 11,000-9,000= 2,000
Tienda B: 19,000-1,000= 18,000.

El rango se expresa frecuentemente con la siguiente fórmula:

$$\boxed{R=X_M - X_m}$$

En esta fórmula R representa al rango; X_M al dato mayor y X_m al dato menor.

El rango es una medida de dispersión que es muy fácil de obtener, pero es un tanto burda, pues solamente toma en cuenta los datos extremos y **no** hace **caso de los datos que están en medio**. Para tomar en cuenta todos los datos se inventaron las siguientes medidas de dispersión que son la varianza y la desviación estándar.

Varianza y desviación estándar

La varianza es una medida de dispersión y es el promedio aritmético de la sumatoria de las desviaciones de cada valor de variable con respecto a la media de los datos.

La desviación estándar se conoce también como desviación típica y es simplemente la raíz cuadrada de la varianza.

Supongamos las ventas de las siguientes dos tiendas:

Tienda C: \$5,000; \$10,000; \$10,000; \$10,000; \$15,000.

Tienda D: \$5,000; \$6,000; \$10,000; \$14,000; \$15,000.

Ambas tiendas tienen una media de \$10,000 y un rango de \$10,000, como fácilmente el alumno puede comprobar; sin embargo, podemos darnos cuenta de que en la tienda D la información está un poco más dispersa que en la tienda C, pues en esta última, si exceptuamos los valores extremos, todos los demás son diez mil; en cambio, en la tienda D existe una mayor diversidad de valores.

Un enfoque que nos puede permitir tomar en cuenta todos los datos es el siguiente:

Supongamos que deseamos saber qué tan alejado está cada uno de los datos de la media. Para ello podemos sacar la diferencia entre cada uno de los datos y esa media para, posteriormente, promediar todas esas diferencias y ver, en promedio, que tan alejado está cada dato de la media ya citada. En la siguiente tabla se realiza ese trabajo.

Tienda C			Tienda D		
datos	cada dato menos la media		datos	cada dato menos la media	
5000	5000-10000=	-5000	5000	5000-10000=	-5000
10000	10000-10000=	0	6000	6000-10000=	-4000
10000	10000-10000=	0	10000	10000-10000=	0
10000	10000-10000=	0	14000	14000-10000=	4000
15000	15000-10000=	5000	15000	15000-10000=	5000
	suma	**0**		**suma**	**0**

Cuadro 2.12. Tabla de desviaciones de datos

Como se puede apreciar la suma de las diferencias entre la media y cada dato tiene cero como resultado por lo que entonces, se elevan las diferencias al cuadrado para que los resultados siempre sean positivos.

A continuación se muestra este trabajo y la suma correspondiente.

Tienda C				Tienda D			
datos	cada dato menos la media		Cuadrado de lo anterior	datos	cada dato menos la media		Cuadrado de lo anterior
5000	5000-10000=	-5000	25,000,000	5000	5000-10000=	-5000	25,000,000
10000	10000-10000=	0	0	6000	6000-10000=	-4000	16,000,000
10000	10000-10000=	0	0	10000	10000-10000=	0	0
10000	10000-10000=	0	0	14000	14000-10000=	4000	16,000,000
15000	15000-10000=	5000	25,000,000	15000	15000-10000=	5000	25,000,000
	suma	**0**	**50,000,000**		**suma**	**0**	**82,000,000**

Cuadro 2.13. Tabla de desviaciones cuadráticas

En este caso, ya la suma de las diferencias entre cada dato y la media (elevadas al cuadrado) nos da un valor diferente de cero con el que podemos trabajar. A este último dato (el de la suma), dividido entre el numero total de datos lo conocemos como varianza (o variancia, según el libro que se consulte).

Por lo mismo, tenemos que la varianza de los datos de la tienda C es igual a 50,000,000/5, es decir 10,000,000. Siguiendo el mismo procedimiento podemos obtener la varianza de la tienda D, que es igual a 82,000,000/5, es decir, 16,500,000.

Es en este punto cuando nos podemos percatar que la varianza de la tienda D es mayor que la de la tienda C, por lo que la información de la primera de ellas (D) está más dispersa que la información de la segunda (C).

La fórmula para la varianza de una población es un resumen del procedimiento que acabamos de seguir y es la siguiente:

$$\sigma^2 = \sum_{1}^{N} (x_i - \mu)^2 / N$$

En donde:

σ^2 es la varianza de datos.

$\sum$ indica una sumatoria.

x_i variable o dato.

μ media de datos.

N número de datos en una población.

La **varianza** es una medida muy importante y tiene interesantes aplicaciones teóricas. Sin embargo, es difícil de comprender de manera intuitiva, entre otras cosas porque al elevar las diferencias entre el dato y la media al cuadrado, las unidades de medida son las de los datos al cuadrado y no es nada fácil captar lo que significan pesos al cuadrado o (en algún otro problema) focos al cuadrado. Por ello se determinó obtener la raíz cuadrada de la varianza. De esta manera las unidades vuelven a expresarse de la manera original y su sentido es menos difícil de captar.

En el caso de nuestras tiendas, las desviaciones estándar son para la tienda C \$3,162.28 y para la tienda D \$4,062.02.

La fórmula para la desviación estándar es:

$$\sigma = \sqrt{\sum_{1}^{N}(x_i - \mu)^2 / N}$$

El alumno podrá observar que la sigma ya no está al cuadrado, lo que es lógico, pues si la varianza es sigma al cuadrado, la raíz cuadrada de la misma es, simplemente sigma. Es importante precisar que ésta es la fórmula de la desviación estándar para una población.

En estadística inferencial es importante distinguir los símbolos para una muestra y para una población. La desviación estándar para una muestra tiene una fórmula cuyo denominador es (n-1) siendo "n" el tamaño de la muestra.

El estudiante deberá notar que al total de la población se le denota con "N" mayúscula y al total de datos de la muestra se le denota con "n" minúscula.

El coeficiente de variación

Dos poblaciones pueden tener la misma desviación estándar y, sin embargo, podemos percatarnos intuitivamente que la dispersión no es la misma para efectos de una toma de decisiones.

El siguiente ejemplo aclara estos conceptos.

Un comercializador de maíz vende su producto de dos maneras distintas:

a) En costales de 50 Kg.
b) A granel, en sus propios camiones repartidores que cargan 5 toneladas (5000) Kg.

Para manejar el ejemplo de manera sencilla, supongamos que en un día determinado solamente vendió tres costales y que además salieron tres camiones cargados; para verificar el trabajo de los operarios, se pesaron tanto unos como otros en presencia de un supervisor. Sus pesos, la media de los mismos y sus desviaciones estándar aparecen en la siguiente tabla (el alumno puede comprobar las medias y las desviaciones estándar calculándolas él mismo, a manera de ejercicio):

Peso de los costales	Peso de los camiones
40 Kg	4990Kg
50 Kg.	5000 Kg.
60 Kg.	5010 Kg.

Cuadro 2.14. Tabla de datos

Media de los costales 50 Kg.

Media de los camiones 5000 Kg.

Desviación estándar de los costales 8.650 Kg.

Desviación estándar de los camiones 8.650 Kg.

Podemos percatarnos de que las variaciones en el peso de los camiones son muy razonables, dado el peso que transportan. En cambio, las variaciones en el peso de los costales son muy grandes, en relación con lo que debería de ser. Los operarios que cargan los camiones pueden ser felicitados por el cuidado que ponen en su trabajo, en cambio podemos ver fácilmente que los trabajadores que llenan los costales tienen algún problema serio, a pesar de que la variación (la desviación estándar) es la misma en ambos casos.

Para formalizar esta relación entre la variación y lo que debe de ser, se trabaja el coeficiente de variación o dispersión relativa, que no es otra cosa que la desviación estándar entre la media y todo ello por cien. En fórmula lo expresamos de la siguiente manera:

$$C.V. = (\sigma / \mu)100$$

Siendo:

$C.V.$ coeficiente de variación.

σ desviación estándar.

μ media de la población.

En el caso de los costales tendíamos:
C.V.=(8.165/50)100=16.33.

Esto nos indica que la desviación estándar del peso de los costales es del 16.33% del peso medio (una desviación importante de lo que debía de ser).

Por otra parte, en el caso de los camiones, el coeficiente de variación nos arroja:
C.V.=(8.165/5000)100= 0.1633

Esto nos indica que la desviación estándar del peso de los camiones es de menos del uno por ciento del peso medio (una desviación realmente razonable).

Datos agrupados en clases o eventos

Cuando se tiene un fuerte volumen de información y se debe trabajar sin ayuda de un paquete de computación, no es práctico trabajar con los datos uno por uno, sino que conviene agruparlos en subconjuntos llamados "**clases**", pues así es más cómodo manipularlos aunque se pierde alguna precisión.

Imagine que se tienen 400 datos el trabajo que representaría ordenarlos uno por uno para obtener la mediana. Por ello se han desarrollado un conjunto de técnicas que permite el trabajo rápido mediante agrupamiento de datos. A continuación se dan algunas definiciones para, posteriormente, pasar a revisar las técnicas antes citadas.

Clase: Cada uno de los subconjuntos en los que dividimos nuestros datos.

Número de clases: Debemos definirlo con base en el número total de datos.

Diversos autores proporcionan diferentes criterios:

- Algunos dicen que el número de clases debe ser aproximadamente la raíz cuadrada del número de datos.

- Otros afirman que el número de clases es aproximadamente el logaritmo del número de clases entre el logaritmo de 2. Normalmente se afirma que las clases no deben ser ni menos de cinco ni más de veinte.

De cualquier manera, el responsable de trabajar con los datos debe utilizar su criterio para lograr los resultados que desea. A continuación se dan algunos ejemplos del número de clases que se obtienen según los dos criterios antes señalados.

Número de datos	**número de clases**	
	(criterio de la raíz cuadrada)	**(criterio del logaritmo)**
50	Aproximadamente 7	6
100	Aproximadamente 10	7
150	Aproximadamente 12	7
200	Aproximadamente 14	8

Cuadro 2.15. Tabla de N° de datos.

Supongamos que tenemos 44 datos —los cuales aparecen en la tabla que se presenta a continuación—, que corresponden a las ventas diarias de una pequeña miscelánea y que seguimos el criterio de los logaritmos. El número de clases será logaritmo de 44 entre logaritmo de 2. Si denominamos K al número aproximado de clases tendremos: K=log 44/log 2=1.6434/0.3010 =5.46, es decir, aproximadamente 5 clases.

Miscelánea "La Esperanza" Ventas de 44 días consecutivos							
día	Venta	día	Venta	día	Venta	día	Venta
1	508	**13**	628	**25**	671	**37**	951
2	918	**14**	935	**26**	965	**38**	667
3	911	**15**	606	**27**	816	**39**	897
4	639	**16**	680	**28**	525	**40**	742
5	615	**17**	993	**29**	846	**41**	1000
6	906	**18**	693	**30**	773	**42**	800
7	638	**19**	586	**31**	547	**43**	747
8	955	**20**	508	**32**	624	**44**	500
9	549	**21**	885	**33**	524		
10	603	**22**	590	**34**	603		
11	767	**23**	763	**35**	890		
12	532	**24**	829	**36**	772		

Cuadro 2.16. Tabla de ventas

Ancho de clase

Es el tamaño del intervalo que va a ocupar cada clase. Se considera que el ancho de clase se obtiene dividiendo el rango entre el número de clases. Así, en el ejemplo de la miscelánea nuestro dato mayor es 999.70, nuestro dato menor es 500 y anteriormente habíamos definido que necesitábamos cinco clases, por lo que el ancho de clase es el rango (499.70 o prácticamente 500) entre el número de clases (5). Por tanto, el ancho de clase es de 100.

Límites de clase

Es el punto en el que termina una clase y comienza la siguiente. En el ejemplo del párrafo anterior podemos resumir la información de la siguiente manera:

Primera clase: comienza en 500 y termina en 600

Segunda clase: comienza en 600 y termina en 700

Tercera clase: comienza en 700 y termina en 800

Cuarta clase: comienza en 800 y termina en 900

Quinta clase: comienza en 900 y termina en 1,000

Estas clases nos permitirán clasificar nuestra información. Si un dato, por ejemplo, tiene el valor de 627.50, lo colocaremos en la segunda clase. El problema que tiene esta manera de clasificar la información es que en los casos de datos que caen exactamente en los límites de clase, no sabríamos en cual de ellas clasificarlos. Si un dato es exactamente 700, no sabríamos se debemos asignarlo a la segunda o a la tercera clase. Para remediar esta situación existen varios caminos, pero el más práctico de ellos (y el que usaremos para los efectos de este trabajo) es el de hacer intervalos abiertos por un lado y cerrados en el otro.

Esto se logra de las siguientes maneras:

Clase	**Incluye datos iguales o mayores a:**	**Incluye datos menores a:**
Primera	500	600
Segunda	600	700
Tercera	700	800
Cuarta	800	900
Quinta	900	1000

Cuadro 2.17. Tabla de clases

Como vemos, los intervalos de cada clase están cerrados del lado izquierdo y abiertos del derecho. Se puede tomar la decisión inversa y dejar abierto el intervalo del lado izquierdo y cerrado del lado derecho. Este enfoque se ejemplifica en la siguiente tabla.

Clase	**Incluye datos mayores a:**	**Incluye datos menores o iguales a:**
Primera	500	600
Segunda	600	700
Tercera	700	800
Cuarta	800	900
Quinta	900	1000

Cuadro 2.18. Tabla de clases

En lo único que se debe tener cuidado es en no excluir alguno de nuestros datos al hacer la clasificación. En el caso de la última tabla, por ejemplo excluimos a los datos cuyo valor es exactamente de 500. Podemos dejarlo así partiendo de la base de que el impacto en nuestro trabajo es, para efectos prácticos, despreciable o modificar los límites para dar cabida a todos los datos. A continuación se presenta un ejemplo de esta segunda alternativa.

Clase	Incluye datos iguales o mayores a:	Incluye datos menores a:
Primera	499.99	599.99
Segunda	599.99	699.99
Tercera	699.99	799.99
Cuarta	799.99	899.99
Quinta	899.99	999.99

Cuadro 2.19. Tabla de clases

De esta manera, tenemos contemplados todos nuestros datos. El investigador deberá definir cuál criterio prefiere con base en el rigor que desea y de las consecuencias prácticas de su decisión. Posteriormente, conforme desarrollemos el ejemplo, se verá el impacto por elegir una u otra de las alternativas.

Marca de clase

La marca de clase es, por así decirlo, la representante de cada clase. Se obtiene sumando el límite inferior y el superior de cada clase y promediándolos. A la marca de clase se le conoce como Xi. A continuación se calcula para las clases de nuestro ejemplo:

Clase	Incluye datos iguales o mayores a:	Incluye datos menores a:	Xi	Xi
Primera	500	600	(500+600)/2=	550
Segunda	600	700	(600+700)/2=	650
Tercera	700	800	(700+800)/2=	750
Cuarta	800	900	(800+900)/2=	850
Quinta	900	1000	(900+1000)/2=	950

Cuadro 2.20. Marcas de clase

Podemos ver con facilidad que las marcas de clase de la cuadro 2.17 son iguales a éstos; en cuanto a las marcas de clase de la cuadro 2.18, éstas se calculan a continuación.

Clase	Incluye datos iguales o mayores a:	Incluye datos menores a:	Xi	Xí
Primera	499.99	599.99	(499.99+599.99)/2=	549.99
Segunda	599.99	699.99	(599.99+699.99)/2=	649.99
Tercera	699.99	799.99	(699.99+799.99)/2=	749.99
Cuarta	799.99	899.99	(799.99+899.99)/2=	849.99
Quinta	899.99	999.99	(899.99+999.99)/2=	949.99

Cuadro 2.21. Marcas de clase

Podemos ver que la diferencia entre la marca de clase de las dos primeras tablas y la tercera es de solamente un centavo. Veremos en el resto del ejemplo las consecuencias que tiene esa diferencia en el desarrollo del trabajo.

Una vez que se tiene la “armadura” o estructura en la que se van a clasificar los datos, se procede a clasificar éstos. Para esto usaremos una de las clasificaciones ya especificadas:

Clase	Incluye datos mayores a:	Incluye datos menores o iguales a:	Frecuencia en clase (Fi)
Primera	500	600 IIIII IIIII I	11
Segunda	600	700 IIIII IIIII I	11
Tercera	700	800 IIIII II	7
Cuarta	800	900 IIIII I	6
Quinta	900	1000 IIIII IIII	9
Total:			44

Cuadro 2.22. Tabla de frecuencias

Para calcular las medidas de tendencia central y de dispersión en **datos agrupados en clases** se utilizan fórmulas similares a las ya estudiadas y la única diferencia es que se incluyen las frecuencias de clase.

A continuación se maneja un listado y un ejemplo de aplicación:

Medidas de tendencia central:

a) Media: $$\bar{x} = \frac{\sum_{x=1}^{N} x_i . f_i}{n}$$

En donde: x_i es la marca de clase.

f_i es la frecuencia de clase.

N es el número de clases.

n es el número de datos.

b) Mediana: $$Md = L_M + \frac{n/2 - F_M}{f_M} \cdot i$$

En donde:

L_M es el límite inferior del intervalo que contiene a la mediana.

F_M es la frecuencia acumulada hasta el intervalo que contiene a la mediana.

f_M es la frecuencia absoluta del intervalo que contiene a la mediana.

i es el ancho del intervalo que contiene a la mediana.

c) Moda o modo: $Mo = L_{Mo} + \frac{d_1}{d_1 + d_2} \cdot i$ $\quad d_1 = f_{Mo} - f_1$ $\quad d_2 = f_{Mo} - f_2$

En donde:

L_{Mo} es límite inferior del intervalo que contiene el modo.

d_1 es la diferencia entre la frecuencia de clase (f_{Mo}) que contiene al modo y la frecuencia de la clase inmediata anterior (f_1).

d_2 es la diferencia entre la frecuencia de clase (f_{Mo})que contiene al modo y la frecuencia de la clase inmediata posterior (f_2)

Medidas de dispersión:

a) Rango: Es la diferencia entre el limite superior del último intervalo de clase y el límite inferior del primer intervalo de clase.

b) Varianza: $\sigma^2 = \frac{\sum (x_i - \bar{x})^2 f_i}{n}$

En donde:

x_i es la marca de clase.

f_i es la frecuencia de clase.

$\bar{x}$ es la media.

n es el número de datos.

c) Desviación estándar:

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

d) Coeficiente de variación:

$$V.I. = \frac{\sigma}{\bar{x}}$$

Se puede utilizar indistintamente la simbología de estadísticos o parámetros, si no es necesario distinguir que los datos provienen de una muestra o de una población. En la estadística inferencial si es importante manejar esta distinción ya que se trabaja con muestras para inferir los parámetros poblacionales.
En el ejemplo siguiente se muestra la utilización de las fórmulas descritas:

En un laboratorio se estudiaron 110 muestras para determinar el número de bacterias por cm^3 de agua contaminada en diversas localidades de un estado del país. En la siguiente tabla de trabajo, se muestran las frecuencias encontradas f_i y los diversos cálculos para determinar las medidas de tendencia central y de dispersión de estas muestras:

Límites reales	x_i	f_i	$f_i acum$	$x_i f_i$	$\left(x_i - \bar{x}\right)^2 f_i$
50 - 55	52.5	4	4	210.0	2,260.57
55 - 60	57.5	7	11	402.5	2,466.91
60 - 65	62.5	9	20	562.5	1,707.19
65 - 70	67.5	12	32	810.0	923.53
70 – 75	72.5	15	**47**	1,087.5	213.50
Md 75 – 80	77.5	**18**	65	1,395.0	27.11
Mo 80 – 85	82.5	20	85	1,650.0	775.58
85 – 90	87.5	13	98	1,137.5	1,638.67
90 – 95	92.5	7	105	647.5	1,843.27
95 - 100	97.5	5	110	487.5	2,252.99
$\sum$		110		8,390.0	14,109.32

Medidas de tendencia central:

a) Media: $$\bar{x} = \frac{\sum_{x=1}^{N} x_i f_i}{n} = \frac{8,390.0}{110} = 76.27$$

El promedio de agua contaminada de todas las muestras es de 76.27 bacterias por cm^3.

b) Mediana: $Md = L_M + \dfrac{n/2 - F_M}{f_M} \cdot i = 75 + \dfrac{55 - 47}{18} \cdot 5 = 77.22$

Se identifica el intervalo que contiene a la mediana (75 – 80) y las frecuencias del límite superior del intervalo anterior del que contiene a la mediana (47) y la frecuencia del propio intervalo (18).

El punto medio de estas muestras es de 77.22 bacterias por cm^3.

c) Moda o modo: $Mo = L_{Mo} + \dfrac{d_1}{d_1 + d_2} \cdot i$ $\qquad d_1 = f_{Mo} - f_1$, $\quad d_2 = f_{Mo} - f_2$

$Mo = 80 + \dfrac{2}{2+7} \cdot 5 = 80.11$ en donde: $d_1 = 20 - 18 = 2$, $d_2 = 20 - 13 = 7$ y $f_{Mo} = 20$, $f_1 = 18$, $f_2 = 13$, $i = 5$

El valor modal se encuentra en el intervalo 80 – 85 y exactamente corresponde a 80.11 bacterias por cm^3.

Medidas de dispersión:

a) Rango: 100 – 50 = 50 . La diferencia es de 50 bacterias por cm^3 entre la muestra menos contaminada y la más contaminada.

b) Varianza: $\sigma^2 = \dfrac{\sum (x_i - \bar{x})^2 f_i}{n} = \dfrac{14,109.32}{110} = 128.27$

La desviación cuadrática de las muestras con respecto a su media es de 128.7 bacterias por cm^3.

c) Desviación estándar:

$$\sigma = \sqrt{128.27} = 11.32$$

La desviación lineal de las muestras con respecto a su media es de 11.32 bacterias por cm^3.

d) Coeficiente de variación:

$$V.I. = \frac{\sigma}{\bar{x}} = \frac{11.32}{76.27} = 0.148 = 14.8\%$$

Este resultado indica que el promedio de la desviación de los datos con respecto a su media se encuentran en un porcentaje aceptable (<25%) para utilizar esta distribución para fines estadísticos.

2.6 Teorema de Tchebysheff y regla empírica

El teorema de Tchebysheff y la regla empírica nos permiten inferir el porcentaje de elementos que deben quedar dentro de una cantidad específica de desviaciones estándar respecto a la media. Se utilizan principalmente para estimar el número aproximado de datos que se encuentran en determinadas áreas de la distribución de datos.

2.6.1. Teorema de Tchebysheff o (Chebyshev).

Cuando menos $1 - \frac{1}{k^2}$ de los elementos en cualquier conjunto de datos debe estar a menos de "**k**" desviaciones estándar de separación respecto a la media, "**k**" puede ser cualquier valor mayor que **1.**

Por ejemplo, veamos algunas implicaciones de este teorema con k=2, 3, y 4 desviaciones estándar:

- cuando menos el 0.75 o 75% de los elementos deben estar a menos de **z=2** desviaciones estándar del promedio.
- cuando menos el 0.89 u 89% de los elementos deben estar a menos de **z=3** desviaciones estándar del promedio.
- cuando menos el 0.94 o 94% de los elementos deben estar a menos de **z=4** desviaciones estándar del promedio.

Como ejemplo de aplicación, supongamos que las calificaciones de 100 alumnos en un examen parcial de estadística tuvieron un promedio de 70 y una desviación estándar de 5. ¿Cuántos alumnos tuvieron calificaciones entre 60 y 80? ¿Cuántos entre 58 y 82?

Solución:

Para las calificaciones entre 60 y 80 vemos que el valor de 60 está a 2 desviaciones estándar abajo del promedio y que el valor de 80 está a dos desviaciones estándar arriba. Al aplicar el teorema de Tchebysheff cuando menos el 0.75 o 75% de los elementos deben tener valores a menos de dos desviaciones estándar del promedio. Así, cuando menos 75 de los 100 alumnos deben haber obtenido calificaciones entre 60 y 80.

Para las calificaciones entre 58 y 82. (58-70)/5=2.4 indica que 50 está a 2.4 desviaciones estándar abajo del promedio, y que (82-70)/5=2.4 indica que 82 está a 2.4 desviaciones estándar arriba del promedio. Al aplicar el teorema de Tchebysheff con z=2.4 tenemos que:

$$1-\frac{1}{k^2}=1-\frac{1}{2.4^2}=0.826$$

Cuando menos el 82.6% de los alumnos deben tener calificaciones entre 58 y 82.

Como podemos ver, en el **teorema de Tchebysheff** se requiere que **z sea mayor que uno,** pero no necesariamente debe ser un entero.

Una de las ventajas del teorema de Tchebysheff es que se aplica a cualquier conjunto de datos, independientemente de la forma de la distribución de los mismos.

Sin embargo, en las aplicaciones prácticas se ha encontrado que muchos conjuntos de datos tienen una distribución en forma de colina o de campana, o sea, que tienen una distribución normal. Cuando se cree que los datos tienen aproximadamente esa distribución se puede aplicar la regla empírica para determinar el porcentaje de elementos que debe estar dentro de determinada cantidad de desviaciones estándar respecto del promedio.

2.6.2. La regla empírica

La regla empírica dice que para los datos que se distribuyen de una manera normal (en forma de campana):

- aproximadamente el 68% de los elementos están a menos de una desviación estándar de la media.
- aproximadamente el 95% de los elementos están a menos de dos desviaciones estándar de la media.
- casi todos los elementos están a menos de tres desviaciones estándar de la media.

Ejemplo:

En una línea de producción se llenan, automáticamente, envases de plástico con detergente líquido. Con frecuencia, los pesos de llenado tienen una distribución en forma de campana. Si el peso promedio de llenado es de 16 onzas y la desviación estándar es de 0.25 onzas, se puede aplicar la regla empírica para sacar las siguientes conclusiones:

- aproximadamente el 68% de los envases llenos tienen entre 15.75 y 16.25 onzas (esto es, a menos de una desviación estándar del promedio)
- aproximadamente el 95% de los envases llenos tienen entre 15.50 y 16.50 onzas (esto es, a menos de dos desviaciones estándar del promedio)
- casi todos los envases llenos tienen entre 15.25 y 16.75 onzas (esto es, a menos de tres desviaciones estándar del promedio).

El estudio y conocimiento de una adecuada recolección, análisis y procesamiento de datos, constituyen una plataforma básica para profundizar en otros requerimientos estadísticos de orden superior.

La presentación gráfica de datos es muy útil para visualizar su comportamiento y distribución y también para determinar la posición de las medidas de tendencia central y la magnitud de su dispersión.

Por lo tanto el dominio que se alcance para calcular estas medidas de datos no agrupados y datos agrupados en clases, así como su correcta interpretación, ayudarán a tomar mejores decisiones en cualquier ámbito personal, social o profesional.

Bibliografía del tema 2

BERENSON L., Mark, David Levine M. y Timothy Krehbiel C., *Estadística para administración*, 2ª edición, México, Prentice Hall, 2001.

LEVIN, Richard I. y David S. Rubin, *Estadística para administradores*, 6a. Edición, México, Prentice Hall, 1996.

MASON D., Robert, Douglas Lind A. y William Marchal G, *Estadística para administración y economía*, 11ª edición, Colombia, Alfaomega, 2004.

Actividades de aprendizaje

2.1 A partir del estudio de la bibliografía específica sugerida para el tema 2, describe en una tabla cada uno de los conceptos principales y sus aplicaciones.

2.2 Selecciona y describe ejemplos de recolección y procesamiento de datos en el sector comercial, empresarial e industrial.

2.3 Presenta estos datos en forma de gráficas y tablas e interpreta su comportamiento.

2.4 Obtén sus medidas de tendencia central y de dispersión con base en datos brutos.

2.5 Agrupa estos datos en clases determinando el número de intervalos más conveniente y presenta sus frecuencias absolutas, relativas, y acumuladas en tablas y gráficas.

2.6 Calcula sus medidas de tendencia central y de dispersión de estos datos agrupados en clases y compara tus resultados con los obtenidos para datos brutos.

Investiga otras aplicaciones prácticas del teorema de Tchebysheff y presenta sus resultados.

2.7 Investigue otras aplicaciones prácticas del la regla empírica y presente sus resultados.

2.8 Realiza un resumen de los conceptos más importantes de este tema y presenta tus propias conclusiones.

2.9 Visita la página www.aulafacil.com (Curso de estadística, clases 4, 5, y 6), compara los temas estudiados con la propuesta que se expresa e escribe tus conclusiones.

Cuestionario de autoevaluación

1. ¿Por qué es necesario organizar un conjunto de datos recopilados?
2. ¿Cuál es la diferencia entre datos nominales y datos ordinales?
3. Defina las características de una escala numérica, una escala de intervalo y una escala de razón.
4. ¿Cuáles son los principales elementos para elaborar una tabla de distribución de frecuencias?
5. ¿Cuáles son las principales diferencias entre un cuadro estadístico de trabajo y un cuadro estadístico de referencia?
6. Indique las diferencias entre un diagrama de barras, un histograma y un diagrama circular y sus aplicaciones más frecuentes.
7. ¿Cuáles son las características más importantes de la media o promedio aritmético, la mediana y la moda de un conjunto de datos?
8. ¿En qué consisten los cuartiles, deciles y percentiles en un conjunto de datos?
9. Explique qué es el rango y el recorrido intercuartílico.
10. Describa las fórmulas de la varianza, de la desviación estándar y del coeficiente de variación de un conjunto de datos, así como la interpretación de cada una y sus posibles aplicaciones.

Examen de autoevaluación

1. Cuando en un estudio estadístico nos referimos a la organización de los datos en categorías y tomamos en cuenta el número de datos que se ubica dentro de cada una de ellas nos referimos a la:
   a. Media
   b. Desviación estándar
   c. Distribución de frecuencias
   d. Moda
   e. Varianza

2. La escala nominal puede ser aplicada a:
   a. La cantidad de fieles que tienen ciertas sectas en Chiapas
   b. La población de los estados de la República Mexicana
   c. Los nombres de los estados de la República Mexicana
   d. Las temperaturas mínimas históricas del desierto de Chihuahua
   e. El peso de los integrantes de una familia

3. El número de alumnos inscritos en la Facultad de Contaduría y Administración puede ser representado por una variable de tipo:
   a. Nominal
   b. Discreto
   c. Continua
   d. De razón

4. Al punto medio de una categoría o clase se le denomina:
   a. Media
   b. Promedio de clase
   c. Mediana
   d. Marca de clase
   e. Centro

5. Una de las representaciones gráficas de la distribución de frecuencias es:

   a. Diagrama de árbol
   b. Distribución normal
   c. Gráfica de dispersión
   d. Ojiva
   e. Histograma

6. Los resultados de rendimiento de combustible de un vehículo en un viaje por la República Mexicana fueron:

Culiacán:	7.6 km/l
Los Cabos	9.2 km/l
Torreón	9.7 km/l

Oaxaca	9.2 km/l
Toluca	6.5 km/l
Chihuahua	9.6 km/l

¿Cuál es la moda de esta muestra?

   a. 8.63 km/l
   b. 9.20 km/l
   c. 8.52 km/l
   d. 6.50 km/l
   e. 9.70 km/l

7. De la siguiente muestra calcule la media aritmética:

Categoría	Límite inferior	Límite superior	Frecuencia
1	600	799	3
2	800	999	7
3	1,000	1,199	11
4	1,200	1,399	22
5	1,400	1,599	40
6	1,600	1,799	24
7	1,800	1,999	9
8	2,000	2,199	4

a. 1,367.82
b. 1,423.74
c. 1,461.17
d. 1,484.50
e. 1,502.18

8. De la siguiente muestra calcule la moda:

Categoría	Límite inferior	Límite superior	Frecuencia
1	600	799	3
2	800	999	7
3	1,000	1,199	11
4	1,200	1,399	22
5	1,400	1,599	40
6	1,600	1,799	24
7	1,800	1,999	9
8	2,000	2,199	4

a. 1,367.82
b. 1,423.74
c. 1,461.17
d. 1,484.50
e. 1,502.18

9. En el censo de población del año 2000 se solicitó la edad del jefe de familia; una muestra de 40 familias mostró el registro de edades siguiente:

Calcule el valor del primer y tercer cuartil.

42	29	31	38	55	27	28	33	49	70
25	21	38	47	63	22	38	52	50	41
19	22	29	81	52	26	35	38	29	31
48	26	33	42	58	40	32	24	34	25

a. 27.50, 47.25

b. 28.52, 45.25

c. 26.53, 48.54

d. 27.55, 47.85

e. 28.52, 47.25

10. Con relación al problema anterior, determine la media, la mediana y desviación estándar.

a. 38.52,35.25,13.89

b. 38.08, 34.50,13.90

c. 38.12, 34.56,13.85

d. 38.52, 35.65, 13.90

e. 38.56, 35.50, 13.75

Tema 3. Análisis combinatorio

Objetivo particular

Al finalizar el tema, el alumno será capaz de aplicar el principio fundamental de conteo, así como las fórmulas de permutaciones y combinaciones en la solución de problemas reales relacionados con las áreas contable y administrativa.

Temario detallado

3.1 Principios fundamentales

- 3.1.1 Principio de multiplicación
- 3.1.2 Principio de adición

3.2 Ordenaciones, permutaciones y combinaciones

- 3.2.1. Ordenaciones
- 3.2.2. Permutaciones
- 3.2.3. Combinaciones

Introducción

Desde los tiempos prehistóricos el hombre realiza juegos de azar, por lo que hubo la necesidad de plantear soluciones a dichos problemas que se presentaban en estos juegos; por ello se hizo una prueba para saber la probabilidad de dicho resultado y se enunciaron estos principios para así obtener la máxima probabilidad de éxito de una estrategia aplicada. Mediante este empleo de principios fundamentales se consideran las diversas necesidades de poder plantear una manera adecuada y satisfactoria con la que no se había notado en otras épocas o que estaban planteadas, pero nadie sabía las diversas formas de aplicarlas.

En la mayoría de los problemas de análisis combinatorio se observa que una operación o actividad aparece en forma repetitiva y es necesario conocer las formas que se puede realizar dicha operación. Para dichos casos es útil conocer determinadas técnicas o estrategias de conteo que facilitarán el cálculo señalado.

El **análisis combinatorio** también se define como una manera práctica y abreviada de **contar**; las operaciones o actividades que se presentan son designadas como eventos o sucesos.

Muchas de las decisiones comerciales requieren que se cuente el número de subconjuntos que se pueden obtener de un conjunto. Por ejemplo, de las ventas de una línea que consta de 10 productos, ¿cuántos subconjuntos de tres productos se pueden ofrecer a los clientes?; ¿cuantos números telefónicos distintos pueden asignarse a una oficina dados los dígitos 0–9? Existen muchos ejemplos en los cuales será importante utilizar diferentes reglas de conteo para su solución.

Por su carácter empírico, las técnicas de conteo no pueden alcanzar el grado de simplificación abstracta alcanzado por la teoría de las probabilidades. A medida en que se complican las situaciones de los conteos, las sencillas definiciones de permutaciones y de combinaciones se vuelven insuficientes; en este punto entra en juego la compleja problemática de los procesos de modelación probabilística.

El análisis combinatorio comprende el estudio de las relaciones de “n” elementos distintos o de parte de ellos (subconjuntos) tomados en orden o no. Esto es esencial para el cálculo de probabilidades.

3.1 Principios fundamentales

Las técnicas de conteo se basan en dos principios fundamentales: el de la multiplicación y el de la adición.

3.1.1 Principio de multiplicación

Si un evento o suceso "A" puede ocurrir, en forma independiente, de **"m"** maneras diferentes y el suceso “B” de **"n"** maneras diferentes, entonces el número de maneras distintas en que pueden suceder ambos sucesos es **"m x n"**

Ejemplo 1: Si un alumno de la Facultad puede llegar a la escuela por metro, camión o auto y puede entrar por cualquiera de las 2 entradas que existen ¿De cuántas maneras distintas puede hacer su arribo?

Solución:

m= 3 formas de llegar a la escuela

n= 2 entradas

Por lo que mxn=6 Existen 6 maneras de hacer su arribo.

Ejemplo 2: En la etapa final de fútbol profesional de primera, cuatro equipos, CRISTAL (C), BOYS (B), ESTUDIANTES (E), UNIVERSITARIO (U), disputan el primer y segundo lugar (campeón y subcampeón). ¿De cuántas maneras diferentes estos equipos pueden ubicarse en dichos lugares?

Solución:

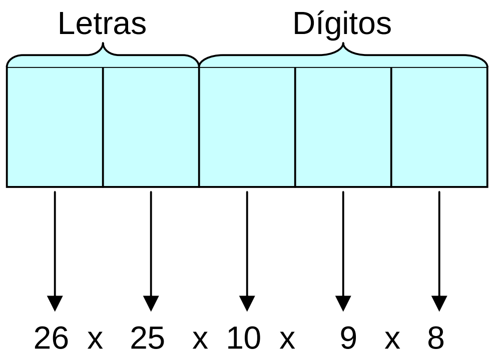
m= 4 equipos

n= 2 lugares

Por lo que mxn=8; así, Existen ocho maneras de ubicación.

Ejemplo 3: ¿Cuántas placas para automóviles pueden hacerse si cada placa consta de dos letras diferentes seguidas de tres dígitos diferentes?, (considere 26 letras del alfabeto).

Solución:



Por lo tanto aplica para
468 000 placas

- El primer casillero puede ser ocupado por cualquiera de las 26 letras
- El segundo casillero puede ser ocupado por cualquiera de las 25 letras que restan
- El tercer casillero puede ser ocupado por cualquiera de los 10 dígitos (del 0 al 9)
- El cuarto casillero lo pueden ocupar los 9 dígitos restantes
- El quinto casiller puede ser ocupado por cualquiera de los 8 dígitos restantes
- Por el principio de multiplicación, el número de placas será 26x25x10x9x8=468,000.

3.1.2 Principio de adición

Supongamos que un evento A se puede realizar de “m” maneras y otro evento B se puede realizar de “n” maneras diferentes, además no es posible que ambos eventos se realicen juntos ($A \cap B = \varnothing$), entonces el evento A o el evento B se realizarán de (m+ n) maneras.

Ejemplo 1: Un repuesto de automóvil se vende en seis tiendas en la Victoria o en ocho tiendas de Breña. ¿De cuántas formas se puede adquirir el repuesto?

Solución:

Por el principio de adición:

Victoria o Breña

6 formas + 8 formas = 14 formas

Ejemplo 2: *Se* desea cruzar un río; para ello se dispone de 3 botes, 2 lanchas y 1 deslizador. ¿De cuántas formas se puede cruzar el río utilizando los medios de transporte señalados?

Solución:

Aplicando el principio de adición se tiene:

$$\begin{array}{ccccc} \text{Bote} & , & \text{lancha} & , & \text{deslizador} \\ \downarrow & & \downarrow & & \downarrow \\ 3 & \text{ó} & 2 & \text{ó} & 1 \end{array}$$

N° de formas = 3+2+1=6

3.2 Ordenaciones, permutaciones y combinaciones

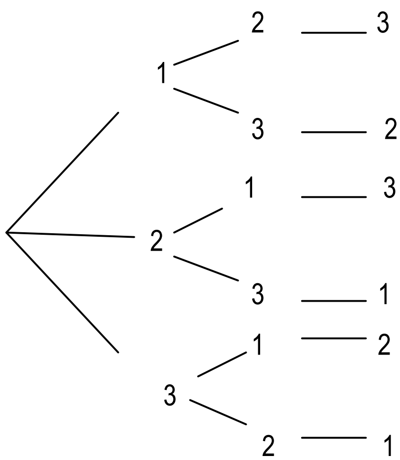
Es importante antes de abordar estos conceptos exponer el concepto de factorial. Los factoriales nos ayudan a **cuantificar rápidamente** el número de posibilidades distintas en que puede ocurrir un evento; los datos que nos aportan son similares a los que obtendríamos de un diagrama de árbol, de hecho podríamos decir que son diagramas de árbol simplificados. El tipo de posibilidades en que se utilice este método dependerá del orden en que se presenten las variables.

El factorial de n se denota como *n!* y se define como: $n! = n(n-1)(n-2)(n-3)\cdots 1$

y *0!* Se define como: $0! = 1$ ya que el *0* es una posibilidad; por lo tanto, asegura que hay una única opción para el evento aunque ésta nunca ocurra.

Ejemplo 1: Utilizar un diagrama para 3! Si utilizamos la ecuación obtendríamos: para 3, 2, 1.

Solución:



$$3! = 3(3-1)(3-2)$$
$$3! = 3*2*1$$
$$3! = 6$$

Como podemos observar con tres variables diferentes podemos obtener 6 eventos distintos.

Ejemplo 2: Obtener 5!

Solución*:*

$$5! = 5(5-1)(5-2)(5-3)(5-4)$$
$$5! = 5*4*3*2*1$$
$$5! = 120$$

Ejemplo 3*:* Calcular 6! multiplicado por 3! y por 0!

Solución:

$$6!*3!*0! = (6*5*4*3*2*1)(3*2*1)(1)$$
$$6!*3!*0! = 720*6*1$$
$$6!*3!*0! = 4320$$

Ejemplo 4*:* Dividir 10! entre 5!

Solución*:*

$$\frac{10!}{5!} = \frac{10*9*8*7*6*5*4*3*2*1}{5*4*3*2*1} = 10*9*8*7*6*1 = 30240$$

3.2.1. Ordenaciones

Se les conoce también como permutaciones sin repetición. El número de ordenaciones de *n* objetos es el número de formas en los que pueden acomodarse esos objetos en términos de orden:

Ordenaciones en n objetos n!=(n)*(n-1)*…*(2)*(1)

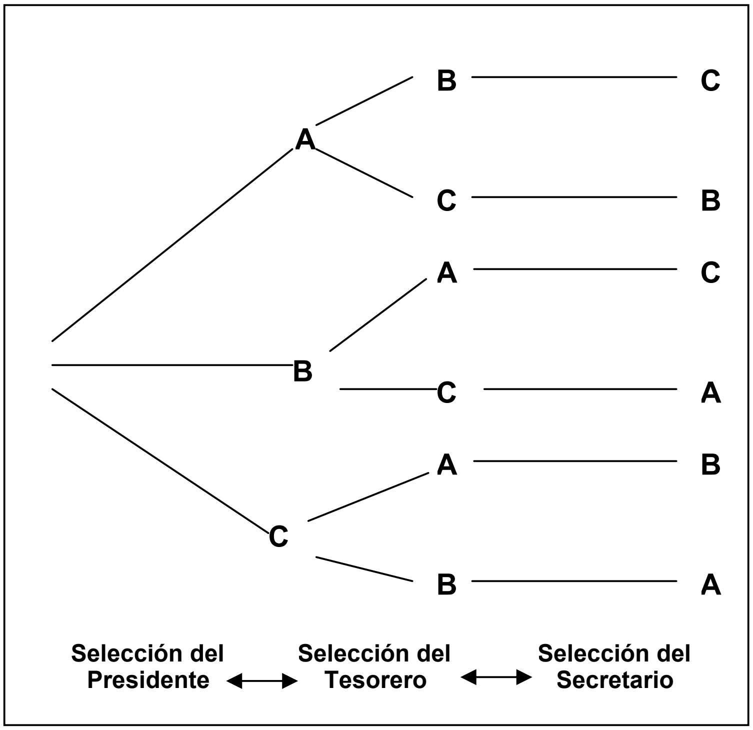
El símbolo *n!* se lee "n factorial".

Ejemplo1: Tres miembros de una organización se han ofrecido a fungir, en forma voluntaria como presidente, tesorero y secretario. Obtener el número de formas en que los tres podrían asumir los puestos.

Solución:

n! =3 = 3x2x1=6 formas

Esto también se puede representar mediante un diagrama, poniéndoles letras a las tres personas A, B Y C.



Ejemplo 2: En una determinada sección de un estante de libros se encuentran cuatro libros. Determinar el número de formas en qué se pueden arreglar ordenadamente.

Solución*:*

n! = 4! = 4x3x2x1=24

Se pueden obtener 24 arreglos de los libros sin repetición.

3.2.2. Permutaciones

Una permutación de un número de objetos es cualquiera de los diferentes arreglos de esos objetos en un orden definido. El número de permutaciones de ***n*** objetos tomados de ***r*** en ***r*** viene dado por la siguiente fórmula general:

$$_nP_r = \frac{n!}{(n-r)!} \quad \text{donde: } r \leq n$$

Ejemplo 1: Una empresa desea colocar tres nuevos gerentes en tres de sus diez plantas. ¿de cuántas maneras diferentes puede hacerlo?

Solución:

Sabemos que la fórmula por aplicar es la anterior, es decir:

$$ {}_{n}P_{r} = \frac{n!}{(n-r)!} \quad \text{donde:} \quad r \leq n $$

Sustituyendo los datos correspondientes tenemos que:

$$ {}_{10}P_{3} = \frac{10!}{(10-3)!} = \frac{10!}{7!} = \frac{10 \cdot 9 \cdot 8 \cdot 7!}{7!} = 10 \cdot 9 \cdot 8 = 720 $$

Como resultado podemos ver que existen 720 formas diferentes de colocar a estos tres gerentes en tres de las diez plantas que posee la empresa.

Ejemplo 2: Supóngase que un club consta de 25 miembros y que se ha de elegir de la lista de miembros un presidente y un secretario. Determine el número total de formas posibles en que estos dos cargos se pueden ocupar.

Solución:

Puesto que los cargos pueden ser ocupados eligiendo uno de los 25 miembros como presidente y eligiendo luego uno de los 24 miembros restantes como secretario, el número total posible de elecciones es de:

$$ {}_{25}P_{2} = (25)(24) = 600 $$

o bien podemos encontrar el resultado utilizando la fórmula:

$$ {}_{n}P_{r} = \frac{n!}{(n-r)!} $$

de donde sustituyendo datos tenemos que:

$$_{25}P_2 = \frac{25!}{(25-2)!}$$

de donde realizando operaciones básicas:

$$_{25}P_2 = \frac{(25)(24)(23!)}{23!}$$

finalmente el resultado se reduce a:

$$_{25}P_2 = (25)(24) = 600$$

Así, el resultado final indica que existen 600 formas diferentes de poder elegir estos cargos de presidente y secretario en el club.

3.2.3. Combinaciones

Una combinación de objetos es cualquier selección de ellos en la que no importa el orden. El número de combinaciones de ***n*** objetos tomados de ***r*** en ***r*** viene dado por la formula general siguiente:

$$_{n}C_{r} = \frac{n!}{r!(n-r)!}$$

Ejemplo 1: Al auditar las 87 cuentas por pagar de una compañía, se inspecciona una muestra de 10 cuentas. ¿Cuántas muestras posibles hay? Suponiendo que 13 de las cuentas contienen un error, ¿cuántas muestras contienen exactamente dos cuentas incorrectas?

Solución:

No hay necesidad de considerar el orden en el que las 10 cuentas se seleccionan, pues todas serán inspeccionadas. Por consiguiente se trata de un problema de combinaciones. Por lo tanto, al aplicar la fórmula correspondiente tenemos que hay:

$$ {}_{87}C_{10} = \frac{87!}{10!(87-10)!} $$

$$ {}_{87}C_{10} = \frac{87!}{10! \cdot 77!} $$

$$ {}_{87}C_{10} \approx 4{,}000{,}000{,}000{,}000 $$ muestras posibles.

Para obtener todas las muestras con dos cuentas incorrectas, podemos combinar cualquiera de las ${}_{13}C_2$ selecciones de dos tomadas de las 13 cuentas incorrectas, con cualquiera de las ${}_{74}C_8$ elecciones de ocho, tomadas de las 74 cuentas incorrectas. Ahora bien, como cada selección de dos cuentas incorrectas se puede acompañar con cualquier elección de ocho cuentas correctas entonces habrá:

${}_{13}C_2$ x ${}_{74}C_8 = 1{,}200{,}000{,}000{,}000$ muestras con dos cuentas erróneas y ocho correctas.

Ejemplo 2: Si un club tiene 20 miembros, ¿Cuántos comités diferentes de cuatro miembros son posibles?

El orden no es importante porque no importa como sean acomodados los miembros del comité. Así, solo tenemos que calcular el número de combinaciones de 20 miembros, tomados de 4 en 4.

Solución:

$$ {}_{20}C_4 = \frac{20!}{4!(20-4)!} $$

$$ {}_{20}C_4 = \frac{20!}{4!16!} $$

$$ {}_{20}C_4 = \frac{20*19*18*17*16!}{4*3*2*1*16!} $$

$$ \boxed{{}_{20}C_4 = 4845} $$

Existen, entonces, 4,845 formas diferentes de conformar el comité.

Los principios y las reglas de conteo como se ha observado en el desarrollo de este tema, constituyen un conocimiento de mecanismos de agrupación de datos y sus relaciones entre sí.

Por el enfoque clásico en el cálculo de probabilidades de ocurrencia de determinados eventos, el valor de probabilidad se basa en la relación de la cantidad de resultados igualmente probables y que sean favorables respecto del número total de resultados posibles. Cuando los problemas o situaciones son sencillas, el número de resultados puede contarse directamente. Sin embargo para problemas o situaciones más complejas se requieren los métodos permutaciones y combinaciones estudiadas para determinar el número de resultados posibles.

Bibliografía del tema 3

ANDERSON R., David, *Estadística para administración y economía*, 8ª edición, Thompson, México, 2004.

MASON D., Robert, Douglas Lind A. y William Marchal G, *Estadística para administración y Economía*, 11ª edición, Alfaomega, Colombia, 2004.

WEBSTER L., Allen, *Estadística aplicada a los negocios y la economía*, 3ª edición, McGraw-Hill, México 2002.

Actividades de aprendizaje

3.1 A partir del estudio de la bibliografía sugerida para el tema 3, presenta por escrito cada uno de los conceptos principales y sus aplicaciones.

3.2 Investiga y realiza por lo menos tres ejercicios o aplicaciones estadísticas de acuerdo con los conceptos estudiados; para ello utilice la bibliografía sugerida.

3.3 Investiga otros tipos de permutaciones que existen y sus principales aplicaciones.

3.4 Expón las diferencias fundamentales entre los diversos tipos de permutaciones investigadas.

3.5 Consulta y expón las principales aplicaciones de los métodos de conteo en la Teoría de probabilidades.

3.6 Explica la utilización de diagramas de árbol en las permutaciones y combinaciones mediante ejercicios específicos.

3.6 Visita la página www.aulafacil.com (Curso de estadística clases 18, 19, y 20) compara los temas estudiados con la propuesta que se expresa e indica tus conclusiones.

Cuestionario de autoevaluación

1. Explique brevemente en qué consiste el análisis combinatorio y sus principales aplicaciones.
2. ¿Cuáles son los principios fundamentales de las reglas de conteo?
3. ¿En qué se basa el principio de multiplicación?
4. ¿En qué se basa el principio de adición?
5. Explique brevemente el concepto de factorial y cuál es la ayuda que brindan.
6. ¿En qué consisten las ordenaciones?
7. ¿Qué es una permutación y porqué es importante el orden?
8. Explique la fórmula de una permutación.
9. ¿Qué es una combinación y porqué no es importante el orden?
10. Explique la fórmula de una combinación y sus diferencias con la de una permutación.

Examen de autoevaluación

1. El producto de 4! por 3! Es igual a:
   a. 49
   b. 81
   c. 121
   d. 144
   e. 169

2. Diga si la siguiente igualdad es correcta: $n! = n(n + 1)(n + 2)......1$
   a. Es correcta
   b. Es correcta sólo para $n = 0$
   c. Es correcta sólo para $n < 0$
   d. Incorrecta
   e. Se requieren más datos

3. Si un problema se puede resolver de tres maneras diferentes y otro problema se puede resolver de cinco formas distintas, ¿de cuántas maneras se pueden resolver ambos problemas?
   a. 15
   b. 120
   c. 81
   d. 720
   e. 1024

4. Si los tres números se toman de dos en dos, ¿cuántas cantidades se pueden formar?
   a. 12
   b. 6
   c. 9
   d. 3
   e. 10

5. Para el cálculo de permutaciones de n objetos tomados de n en n diga si la siguiente fórmula es correcta: ${}_nP_n = n(n-1)(n-2)...1$
   a. Es correcta
   b. Es correcta sólo para $n = 0$
   c. Es correcta sólo para valores negativos de n
   d. Es incorrecta
   e. Se requieren más datos

6. El valor de ${}_7C_4$ es:
   a. 128
   b. 56
   c. 28
   d. 35
   e. 121

7. ¿De cuántas maneras se puede formar un comité de cinco estudiantes si se tiene nueve candidatos?
   a. 45
   b. 135
   c. 128
   d. 99
   e. 126

8. Con un total de cinco profesores de matemáticas y siete de estadística se integra un comité donde deben participar dos de matemáticas y tres de estadística, ¿de cuántas formas puede formarse dicho comité?
   a. 365
   b. 350
   c. 35
   d. 175
   e. 208

9. General Motors de México ofrece cinco modelos de vehículos con tres tipos distintos de equipamiento, ¿cuántos modelos diferentes pueden ofrecerse a sus clientes?
   a. 720
   b. 15
   c. 30
   d. 45
   e. 120

10. Con siete administradores y cinco contadores se quiere formar un consejo que conste de cuatro administradores y tres contadores, ¿de cuántas maneras diferentes se puede integrar?
   a. 121,480
   b. 3,000,000
   c. 50,400
   d. 1,764,000
   e. 1,024,000

Tema 4. Teoría de la probabilidad

Objetivo particular

Al finalizar el tema, el alumno será capaz de aplicar la teoría de probabilidad y los conceptos relacionados con ella en la solución de problemas que tengan relación con su carrera. Reconocerá los diferentes enfoques y las situaciones en las cuales son aplicables cada uno de ellos y desarrollará la habilidad para aplicar las técnicas de cálculo requeridas para la obtención de los diversos valores de probabilidad que se requiera.

Temario detallado

4.1 Probabilidad subjetiva y probabilidad frecuencial

4.2 Espacio muestral y eventos

- 4.2.1. Probabilidad simple (marginal)
- 4.2.2. Probabilidad conjunta

4.3 Reglas de la adición y multiplicación

4.4 Probabilidad condicional

- 4.4.1. Regla de la multiplicación

4.5 Tablas de probabilidad conjunta

- 4.5.1. Probabilidad marginal

4.6 Teorema de Bayes

Introducción

Algunas personas dicen que solamente existen dos cosas en la vida que seguramente nos acontecerán: los impuestos y la muerte. Todos los demás eventos pueden o no sucedernos; es decir, tenemos un cierto nivel de duda sobre su ocurrencia. Para tratar de cuantificar el nivel de duda (o de certeza) que tenemos de que ocurra un determinado fenómeno se creó la teoría de la probabilidad. En esta unidad nos abocaremos en lo que se conoce como probabilidad básica. En ella no existen muchas fórmulas a las cuales recurrir (desde luego existen algunas). La mayor parte de los problemas se resuelven mediante la aplicación de un reducido conjunto de principios básicos y de algo de ingenio.
Para ello es indispensable entender claramente el problema en sí, por lo que la lectura cuidadosa y crítica es indispensable.

A reserva de ahondar más en el tema, podemos adelantar que la **probabilidad siempre es un número entre cero y uno**. Mientras más probable sea la ocurrencia de un evento más se acercará a uno; mientras más improbable sea se acercará más a cero. Las razones de ello se explican en la siguiente sección de este tema.

Es necesario, por último, hacer una advertencia sobre la presentación de datos. Al ser la probabilidad un número entre cero y uno **es frecuente expresarla en porcentaje**. A la mayoría de las personas se nos facilita más la comprensión cuando la cantidad está expresada de esta última manera. Si decimos, por ejemplo, que la probabilidad de que llueva hoy es del 10%, damos la misma información que si decimos que la probabilidad de que llueva hoy es de 0.10. Ambas maneras de presentar la información son equivalentes.

4.1 Probabilidad subjetiva y probabilidad frecuencial

Para determinar la probabilidad de un suceso podemos tomar dos enfoques. El primero de ellos se denomina **objetivo** y tiene, a su vez, **dos enfoques**, que a continuación se detallan.

En el enfoque ***a priori*** (es decir, antes del hecho) se parte de la base de que se conocen todos los resultados posibles y a cada uno de ellos se les asigna una probabilidad de manera directa sin hacer experimento o medición alguna.

Frecuentemente decimos que al arrojar una moneda existen 50% de probabilidades de que salga águila y 50% de probabilidades de que salga sol, basándonos en el hecho de que la moneda tiene dos caras y que ambas tienen las mismas probabilidades de salir. Igual camino seguimos al asignar a cada cara de un dado la probabilidad de un sexto de salir. Razonamos que el dado tiene seis caras y por tanto, si el dado es legal, cada una de ellas tiene las mismas probabilidades.

El otro enfoque objetivo es el llamado ***a posteriori*** (es decir, después del hecho).

Para asignarle probabilidad a un suceso se experimenta antes y con base en los resultados se asignan probabilidades. En el caso de la moneda, este enfoque nos recomendaría hacer un número muy grande de "volados", por ejemplo mil, y con base en ellos definir la probabilidad. Si decimos, por ejemplo, que la probabilidad de que salga águila es de 488/1000 porque lancé mil veces la moneda y esos fueron los resultados, estaremos aplicando la probabilidad *a posteriori*.

En ejemplos menos triviales, las compañías de seguros desarrollan tablas de mortalidad de las personas para diferentes edades y circunstancias con base en sus experiencias. Ese es un caso del enfoque *a posteriori*.

La **probabilidad subjetiva** es una cuestión de opinión y se basa en el hecho de que un experimento se realiza una sola vez. Dos personas, por ejemplo, pueden asignar diferentes probabilidades a un mismo evento, aun cuando tengan la misma información. Tal **diversidad de opiniones** se puede ver en las proyecciones económicas que hacen los asesores en inversiones y los economistas para los años venideros. Aunque muchos de estos individuos trabajan con los mismos datos, ellos se forman distintas opiniones acerca de las condiciones económicas más probables. Tales proyecciones son inherentemente subjetivas.

También se presenta cuando no existen antecedentes para determinarla (como en el caso de las tablas actuariales de las compañías de seguros) ni una base lógica para fijarla *a priori*.

Si pensamos, por ejemplo, en la final de futbol del mundial de 2002, en la que se enfrentaron Brasil y Alemania, vemos que no había historia previa de enfrentamientos entre los dos equipos y había tantos factores en juego que difícilmente se podía dar una probabilidad sobre las bases que anteriormente llamamos objetivas; por lo mismo, se debe recurrir al juicio de las personas para definir las probabilidades. A esta manera de fijar probabilidades se le llama, por este hecho, probabilidad subjetiva.

4.2 Espacio muestral y eventos

Para trabajar con comodidad la probabilidad, vale la pena expresar algunas ideas que necesitaremos posteriormente.

IDEAS	DESCRIPCIÓN
Experimento	Es aquel proceso que da lugar a una medición o a una

	observación.
Experimento aleatorio	Es aquel experimento cuyo resultado es producto de la suerte o del azar. Por ejemplo, el experimento de arrojar un dado.
Evento	El resultado de un experimento.
Evento aleatorio	El resultado de un experimento aleatorio.Por ejemplo A: el evento de que al arrojar un dado salga un número non.
Evento compuesto	El que puede ser descompuesto en eventos más simples. Por ejemplo, el evento A mencionado anteriormente se puede descomponer en los siguientes eventos: E1: el evento de que al arrojar un dado salga un uno. E2: el evento de que al arrojar un dado salga un tres. E3: el evento de que al arrojar un dado salga un cinco.

Cuadro 4.1. Tabla de conceptos estadísticos

Los eventos que no pueden descomponerse en otros más sencillos se conocen como **eventos simples**. El estudiante habrá notado el cambio de nomenclatura, pues los eventos simples además de la letra tienen un subíndice (el número después de la letra mayúscula del evento). Otra manera de denominar a los eventos simples es la de "puntos muestrales". Esta denominación es útil cuando se trata de representar gráficamente los problemas de probabilidad pues cada evento simple (punto muestral) se representa efectivamente como un punto.

En este momento podemos dar ya una definición más formal de probabilidad:

Sea A un evento cualquiera; N el número de veces que repetimos un experimento en el que se puede representar el evento A; n_A el número de veces que efectivamente se representa el evento A; y P(A) a la probabilidad de que se presente el evento A.

Entonces tenemos que P(A)= $\lim_{N \to \infty} n_A / N$

Es decir que la probabilidad de que ocurra el evento A, es la división entre el número de veces que A efectivamente apareció entre el número de veces que se intentó el experimento. N tiende a infinito quiere decir que el experimento se intentó muchas veces). Podemos ver que el menor valor que puede tener P(A) es de cero, en el caso de que en todos los experimentos intentados A no apareciera ni una sola vez. El mayor valor que puede tener P(A) es de uno, en el caso de que en todos los experimentos intentados A apareciera todas las veces, pues en ese caso n_A sería igual a N y todo número dividido entre sí mismo es igual a 1.

En todos los demás casos, la probabilidad de ocurrencia estará entre estos dos números extremos y por eso podemos decir que la **probabilidad de ocurrencia** de cualquier evento estará entre cero y uno. Ésta es la justificación de la afirmación análoga que se realizó al principio de la unidad y también la justificación de la afirmación que se hace frecuentemente de que la probabilidad se expresa como la frecuencia relativa de un evento; es decir, relativa al total de experimentos que se intentaron.

4.2.1. Probabilidad simple (marginal)

La probabilidad simple de un evento es la que tiene éste, sin considerar las conexiones que pueda tener con otros eventos. También se le llama probabilidad marginal. A continuación se define un procedimiento sencillo para calcular la probabilidad simple de un evento.

1. Defina el experimento.
2. Haga la lista de todos los eventos simples asociados con el experimento que definió (es decir, haga la lista de todos los puntos muestrales).
3. Asigne probabilidades a cada uno de los puntos muestrales. La suma total de las probabilidades de todos los puntos muestrales debe ser igual a la unidad.
4. Defina el evento que le interesa como un conjunto de puntos muestrales.
5. Encuentre la probabilidad del evento que le interesa sumando la probabilidad de los puntos muestrales que lo componen.

A continuación se dan varios ejemplos que nos permitirán comprender mejor este procedimiento.

Ejemplo 1.

a. El experimento consiste en arrojar un dado normal y bien balanceado de seis caras.

b. Todos los resultados posibles (los eventos simples o puntos muestrales) se listan a continuación:

E1: que salga un uno
E2: que salga un dos
E3: que salga un tres
E4: que salga un cuatro
E5: que salga un cinco
E6: que salga un seis

c. Para asignar probabilidades a cada evento, es razonable darle la misma probabilidad a cada evento simple; si hay seis resultados posibles, también resulta razonable darle 1/6 a cada uno.

d. A continuación definimos tres eventos como de interés y los definimos como conjuntos de puntos muestrales.

Evento A: que salga un número menor a cuatro. Se compone de los eventos E1, E2 y E3.
Evento B: que salga un número par. Se compone de los eventos E2, E4, E6.
Evento C: que salga un número mayor que seis. Ningún evento lo compone.

e. La probabilidad de A es la suma de las probabilidades de E1,E2 y E3: 1/6+1/6+1/6 = 3/6=1/2.

f. La probabilidad de B es la suma de las probabilidades de E2, E4, E6: 1/6+1/6+1/6 = 3/6=1/2.
g. La probabilidad de C es de cero, pues no existe ningún evento que lo componga.

Ejemplo 2. El comité directivo de la sociedad de padres de familia de una escuela primaria está compuesto por cinco personas: tres mujeres y dos hombres. Se van a elegir al azar dos miembros del comité para solicitar al delegado que ponga una patrulla a vigilar la salida de los niños ¿Cuál es la probabilidad de que el comité esté compuesto por un hombre y una mujer?

Solución:
El experimento es elegir al azar dos personas de las cuales tres son mujeres y dos son hombres.

Para listar todos los eventos simples simbolizaremos a las mujeres con una M y los hombres con una H. Así el comité está compuesto por: M1, M2, M3, H1,H2; M1 es la primera mujer, M2 la segunda, H1 el primer hombre y así sucesivamente.

Los eventos simples que son posibles se listan a continuación:

M1M2; M1M3; M1H1; M1H2
M2M3; M2H1; M2H2;
M3H1; M3H2;
H1H2.

Vemos que pueden darse 10 pares distintos. Si cada par es elegido al azar, es razonable suponer que todos ellos tienen la misma probabilidad de ser seleccionados, por ello podemos afirmar que cada par tiene una probabilidad de 1/10 de ser seleccionado. Los pares que están constituidos por un hombre una mujer son: M1H1 M1H2; M2H1; M2H2; M3H1 y M3H2; es decir, seis de los diez pares posibles.

La probabilidad de nuestro evento de interés es entonces, de seis veces un décimo o 6/10. Expresada en porcentaje, esta probabilidad será del 60%.

Ejemplo 3. Una tienda de electrodomésticos va a recibir un embarque de seis refrigeradores, de los cuales dos están descompuestos. El dueño de la tienda someterá a prueba dos refrigeradores al recibir el embarque y solamente lo aceptará si ninguno de ellos presenta fallas. Diga usted que probabilidades tiene de aceptar el embarque.

Solución:

El experimento es elegir dos refrigeradores al azar para ver si funcionan o no funcionan.

Si llamamos B al refrigerador que trabaja bien y D al descompuesto, podemos listar a todos los refrigeradores del embarque de la siguiente manera:

B1, B2, B3, B4, D1, D2..

Todos los eventos posibles (es decir, todos los pares diferentes que se pueden elegir) se listan a continuación. Los eventos simples de interés (aquellos en que los dos refrigeradores están en buen estado) se resaltan en negritas.

B1B2; B1B3; B1B4; B1D1; B1D2;
B2B3; B2B4; B2D1; B2D2;
B3B4; B3D1; B3D2;
B4D1; B4D2
D1D2

Vemos que existen quince eventos posibles, de los cuales en seis se presenta el caso de que ambos refrigeradores estén en buen estado. Si, como en lo ejemplos anteriores, asignamos una probabilidad igual a todos los eventos simples (en este caso 1/15); nos percataremos que la probabilidad de aceptar el embarque está dada por 6/15.

4.2.2. Probabilidad conjunta

El enfoque visto hasta el momento es útil para comprender el concepto de probabilidad. Sin embargo, la mayoría de los eventos con los que nos podemos encontrar están constituidos por muchos eventos simples y resulta laborioso el cálculo de los mismos. Afortunadamente podemos recurrir a un procedimiento menos incómodo. Para ello debemos recurrir a las nociones de conjuntos que el estudiante aprendió en matemáticas básicas y a algunas definiciones que se detallan a continuación.

Evento compuesto es aquel que consta de más de un punto muestral.

Para componer eventos hacemos uso de las operaciones de conjuntos llamadas **unión e intersección** y bien podemos hacer uso de una combinación de ambas.

Si definimos a los eventos A y B como resultados de un experimento aleatorio y recordamos que todos los **eventos posibles** (el conjunto universal) lo denominamos **espacio muestral** y lo representamos como S; tenemos que la unión de A y B es un evento que contiene todos los puntos muestrales que pertenecen al evento A y/o que pertenecen al evento B. Si usamos la notación de conjuntos nos queda como: $A \cup B$. La probabilidad de $A \cup B$ es la probabilidad de que suceda el evento A o de que suceda el evento B o de que ambos sucedan conjuntamente. Por otra parte, tenemos que la intersección de A y B es la situación en que ambos A y B suceden conjuntamente. La intersección se denota con la simbología de conjuntos como $A \cap B$. Enseguida aparece un ejemplo que nos ayudará a dejar en claro este concepto.

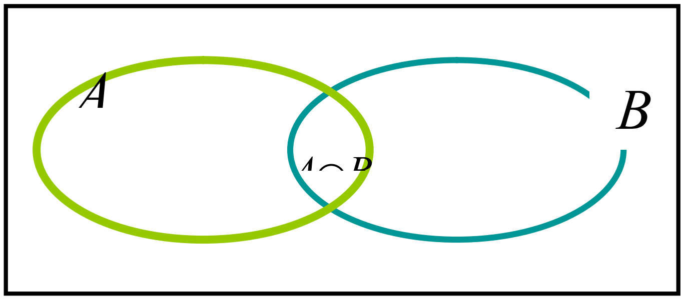


Figura 4.1. Eventos simultáneos.

Ejemplo 1: Si una pareja tiene dos hijos y definimos los siguientes eventos:

A: La pareja tiene por lo menos un varón.

B: La pareja tiene por lo menos una niña.

Defina el espacio muestral y encuentre los puntos muestrales de los siguientes eventos:

Evento A.

Evento B.

Evento $A \cup B$

Si llamamos V al hecho de tener un varón y M al hecho de tener una mujercita, el espacio muestral está integrado por los siguientes eventos:

E1: VV

E2: VM

E3: MV

E4: MM.

Nuestros eventos de interés están integrados de la siguiente manera

Evento A. E1, E2, E3

Evento B. E2, E3, E4

Evento $A \cup B$. E1, E2, E3, E4.

Una vez hecho lo anterior, si asignamos probabilidades iguales a cada uno de los eventos simples, en este caso podemos asignar ¼ o 0.25 a cada uno y por ello tenemos que:

$$P(A)= .75$$
$$P(B)=.75$$
$$P(A \cup B)=1.00$$

Con base en los conceptos que acabamos de desarrollar y de la teoría de conjuntos ya mencionada, podemos clasificar las relaciones entre diversos eventos de la siguiente manera:

Si A y B son eventos aleatorios en el espacio muestral S

Eventos complementarios son todos aquellos que están en el espacio muestral S y que no están en el evento de interés. En nuestro caso, El complemento de A, denotado como A´, son todos los eventos que estando en el espacio muestral no están en A.

Por definición, la probabilidad de A más la probabilidad de su complemento es igual a 1.

Esto lo podemos expresar de la siguiente manera:

P(A) + P(A´)=1. Por lo mismo: P(A)=1 – P(A´).

Si S son todos los resultados posibles al arrojar dos dados, tenemos que

S={2,3,4,5,6,7,8,9,10,11,12}

Si definimos A como el hecho de que el tiro sume menos de cuatro se da que:

A={2,3}; en tanto que A´={4,5,6,7,8,9,10,11,12}

Eventos mutuamente excluyentes. Son aquellos eventos que si se produce uno de ellos, no puede producirse el otro. Dicho en el lenguaje de los conjuntos, podemos afirmar que si dos eventos son mutuamente excluyentes la intersección de ellos está vacía. En terminología de conjuntos también se dice que estos eventos son disjuntos.

Ejemplo 2: Sea S el mismo que el ejemplo anterior.

Sea A: La suma de puntos de los dos dados es de 12.

Sea B: Aparece por lo menos un "uno" en los dados arrojados.

Solución:

Vemos que es imposible que se den A y B simultáneamente, pues para que la suma de doce ambos dados tienen que salir "seis", en tanto que si uno de los dos dados tiene "uno" como resultado, la suma máxima que se puede lograr es de "siete".

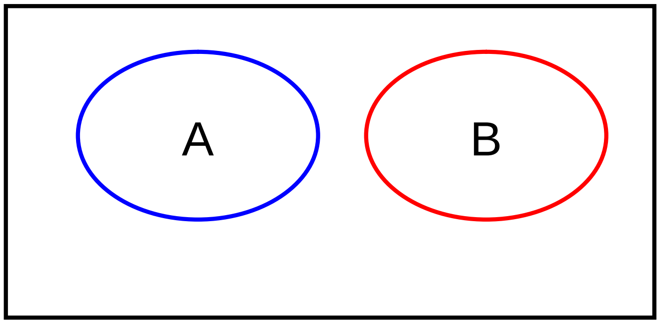


Figura 4.2. Eventos mutuamente excluyentes.

4.3 Regla de la adición y la multiplicación

Si A y B son dos eventos del espacio muestral S, tenemos que la probabilidad de la unión de dos eventos es la suma de las probabilidades de ambos eventos menos la probabilidad de la intersección. Expresado en simbología de conjuntos tenemos que:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Si se da el caso de que A y B sean mutuamente excluyentes, entonces la intersección está vacía y, por lo mismo, la probabilidad de la unión de dos eventos es la suma de las probabilidades de los eventos tomados individualmente. Los siguientes ejemplos nos ayudará a dejar en claro estos conceptos.

Ejemplo 1: En una investigación de mercado se encontró que entre los integrantes de un club, el 30% de los hombres usan loción para después de afeitarse, en tanto que el 40% de ellos utiliza desodorante y el 10% utiliza ambos productos. Si elegimos al azar a un varón de ese club ¿Qué probabilidades existen de que utilice desodorante **o** de que use loción para después de afeitarse?

Sea A: El sujeto usa loción para después de afeitarse.

Sea B: El sujeto usa desodorante.

Solución:

Dado que estamos buscando que utilice, no ambos productos simultáneamente sino cualquiera de los dos de manera indistinta, estamos buscando la unión de los dos eventos. El cálculo de la probabilidad aparece a continuación:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cup B) = 0.40 + 0.30 - 0.10 = 0.60$$

Esto quiere decir que existe un 60% de probabilidades de que un socio de este club elegido al azar use alguno de los dos productos.

En el mismo club de éste ejemplo, el 20% de los socios pertenece al equipo de natación y el 10% al equipo de waterpolo. Ningún socio pertenece a ambos equipos simultáneamente. Diga cual es la probabilidad, si elegimos al azar un socio del club, de que sea integrante de alguno de los dos equipos. El cálculo de probabilidades aparece a continuación. El estudiante tiene que recordar que, dado que ningún socio pertenece a los dos equipos simultáneamente, la intersección está vacía y por lo mismo la probabilidad de su ocurrencia es de cero.

$$P(A \cup B) = 0.20 + 0.10 - 0.0 = 0.30$$

Si tenemos varios eventos mutuamente excluyentes en el espacio de eventos S y queremos saber cuál es la probabilidad de que ocurra cualquiera de ellos es la probabilidad de la unión de los mismos. Dado que al ser eventos mutuamente excluyentes la intersección está vacía, la probabilidad de ocurrencia es simplemente la suma o **adición** de las probabilidades individuales; es por ello que a esta regla se la conoce como regla de la adición.

La regla de la multiplicación se estudiará en el siguiente tema después de la probabilidad condicional.

4.4 Probabilidad condicional

En muchas circunstancias encontramos que la probabilidad de ocurrencia de un **evento** se ve **modificada por** la ocurrencia de **otro evento**. Por ejemplo, la probabilidad de pasar un examen depende del hecho de que el estudiante haya estudiado para el mismo. A esta probabilidad se le conoce como probabilidad condicional. Se expresa simbólicamente de la siguiente manera:

La probabilidad condicional de que ocurra B dado que A ya ocurrió es:

$$P(B/A) = \frac{P(A \cap B)}{P(A)}$$

Es decir, la probabilidad de B dado que A ya ocurrió es igual a la probabilidad de que ocurran ambos eventos simultáneamente dividido por la probabilidad de que ocurra A, que en este caso es el evento antecedente.

La probabilidad condicional de que ocurra A dado que B ya ocurrió es:

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

Es decir, la probabilidad de A dado que B (eso quiere decir la línea vertical entre la B y la A) está dada por la probabilidad de que ocurran ambos eventos simultáneamente dividido por la probabilidad de que ocurra A, que en este caso es el evento antecedente.

El siguiente ejemplo nos ayudará a clarificar estas ideas.
Sea el evento A: Amanece nublado en la región X
Sea el evento B: Llueve en la tarde en la región X.

De acuerdo con información meteorológica de muchos días, tenemos la siguiente información:

Amanece nublado y llueve el 40% de los días.
Amanece nublado y no llueve el 20% de los días.
Amanece despejado y llueve el 10% de los días.
Amanece despejado y no llueve el 30% de los días.

Dado lo anterior, la probabilidad de que llueva en la tarde, es la suma de las probabilidades de que llueva tanto si amaneció despejado como si amaneció nublado. Es decir, 40% más 10%, o sea, 50%. La probabilidad de que no llueva es su complemento, en este caso también el 50%.

Deseamos averiguar lo siguiente.

a) La probabilidad de que llueva en la tarde dado que amaneció nublado.
b) La probabilidad de que llueva en la tarde dado que amaneció despejado.

En el inciso “a” deseamos averiguar B dado que A. Con la información que tenemos ya podemos sustituir directamente en la fórmula.

La probabilidad condicional de que ocurra B dado que A ya ocurrió es:

$$P\left(B / A\right) = \frac{0.40}{0.60} = 0.667 = 66.7\%$$

Es decir, que la probabilidad de que llueva, dado que amaneció nublado, es del 67%. Podemos percatarnos a simple vista de que el hecho de que amanezca nublado efectivamente afecta la probabilidad de que llueva en la tarde. Recordemos que la probabilidad de que llueva simplemente sin tener antecedentes es del 50%. Pasemos al siguiente inciso.

La probabilidad de que llueva en la tarde dado que amaneció despejado.
En este caso estamos buscando B dado que A´ ya ocurrió. Como amanece nublado el 60% de los días y despejado el 40% de ellos, podemos sustituir en la fórmula.

$$P\left(B/A'\right)=\frac{0.10}{0.40}=0.25=25\%$$

Vemos que, si la probabilidad de que llueva cuando amaneció nublado es del 50% y la probabilidad de que llueva estando despejado es de sólo el 25%, el hecho de que amanezca despejado efectivamente afecta las probabilidades de que llueva.

Esto nos lleva al la última definición de relaciones entre eventos que es la de eventos independientes.

Sean dos eventos A y B del espacio de eventos S; decimos que A y B son independientes si la probabilidad de que ocurra A no influye en la probabilidad de que ocurra B y, simultáneamente, la probabilidad de que ocurra B no influye en la probabilidad de que ocurra A. En caso contrario decimos que los eventos son dependientes. Esto lo expresamos simbólicamente del siguiente modo:

Para considerar que A y B son independientes se deben cumplir las dos condiciones siguientes:

$$P\left(B/A\right)=P\left(B\right)\;\; y \;\;\; P\left(A/B\right)=P\left(A\right)$$

Es decir, el hecho de que ocurra un evento no modifica la probabilidad de que ocurra el otro en ningún orden que se tomen.

4.4.1. Regla de la multiplicación

Si tomamos la fórmula de la probabilidad condicional que ya vimos y despejamos la intersección, el resultado es el siguiente:

$$P(A \cap B) = P(B/A) \cdot P(A)$$

Si los eventos A y B son independientes, la probabilidad de B dado que A es igual a la probabilidad de B o P(B), por lo que:

$$P(A \cap B) = P(A) \cdot P(B)$$

Es decir que la probabilidad de que ambos eventos ocurran (ambos juntos no uno o el otro) está dada por el producto de sus probabilidades individuales. A continuación se da un par de ejemplos que nos ayuda a comprender mejor este concepto.

Ejemplo 1: Una tienda de artículos de línea blanca recibe de su proveedor 50 refrigeradores. Entre estos aparatos hay diez que están defectuosos. El comprador inspeccionará dos refrigeradores elegidos al azar ¿Cuál es la probabilidad de que ambos estén defectuosos?

Solución:

Definimos primero nuestros eventos:

A: El primer refrigerador está defectuoso.

B: El segundo refrigerador está defectuoso.

Si hay diez aparatos defectuosos de un total de 50, la probabilidad de que ocurra el evento A es de 10/50 ó sea 0.20.

Si el primer refrigerador sale defectuoso, nos quedan 49 refrigeradores, de los cuales nueve están defectuosos, por ello la probabilidad de B dado que A ocurrió es de 9/49. Esta información se expresa formalmente enseguida y a continuación se hace el cálculo de la probabilidad

$$P(A) = {}^{10}\!/_{50}$$
$$P(B/A) = {}^{9}\!/_{49}$$
$$\therefore\ P(A \cap B) = P(A) \cdot P(B) = {}^{90}\!/_{2,450} = 0.0367 = 3.67\%$$

Esto es aproximadamente el 3.67%

Ejemplo 2: Si se arroja un dado bien balanceado dos veces de manera consecutiva, diga cuál es la probabilidad de que aparezcan dos unos como resultado.

Definimos nuestros eventos:

A: Sale uno en la primera tirada.

B: Sale uno en la segunda tirada.

$$P(A) = {}^{1}\!/_{6}$$
$$P(B/A) = {}^{1}\!/_{6}$$
$$\therefore\ P(A \cap B) = P(A) \cdot P(B) = {}^{1}\!/_{36} = 0.0278 = 2.78\%$$

4.5 Tablas de probabilidad conjunta

Es una tabla donde **se enumeran todos los eventos** posibles para una variable (u observación) en columnas y una segunda variable en filas. El valor en cada celda es la probabilidad de ocurrencia conjunta.

Se empieza elaborando una tabla de contingencia cuyos valores de cada celda se dividen entre el total de datos para obtener los valores de probabilidad correspondientes.

Ejemplo 1: *Se* obtiene una estadística de 300 personas, de acuerdo con su edad y sexo, que entraron en un almacén.

Tabla de contingencia de clientes

Edad/sexo	**Hombre**	**Mujer**	**Total**
< 30 años	35	46	81
30 – 40	42	59	101
> 40 años	51	67	118
Total	128	172	300

Tabla de probabilidad conjunta

Evento	**Edad/sexo**	**Hombre** H	**Mujer** M	**Total**
E_1	< 30 años	0.117	0.153	0.270
E_2	30 – 40	0.140	0.197	0.337
E_3	> 40 años	0.170	0.223	0.393
	Probabilidad marginal	0.427	0.573	1.000

Con esta información obtener la probabilidad de que la siguiente persona que entre al almacén sea:

a) Un hombre menor de 30 años.
b) Una mujer.
c) Una persona de más de 40 años.
d) Habiendo entrado una mujer, que tenga entre 30 y 40 años.
e) Habiendo entrado un hombre, que tenga menos de 30 años.
f) Tenga entre 30 y 40 años, y que sea mujer.
g) Tenga entre 30 y 40 años, y que sea hombre.

Solución:

a) $P(E_1 \cap H) = 0.117 = 11.7\%$

b) $P(M) = 0.573 = 57.3\%$

c) $P(E_3) = .393 = 39.3\%$

d) $P(E_2 / M) = \dfrac{P(E_2 \cap M)}{P(M)} = \dfrac{0.197}{0.573} = 0.344 = 34.4\%$

e) $P(E_1 / H) = \dfrac{P(E_1 \cap H)}{P(H)} = \dfrac{0.117}{0.427} = 0.274 = 27.4\%$

f) $P(M / E_2) = \dfrac{P(E_2 \cap M)}{P(E_2)} = \dfrac{0.197}{0.337} = 0.588 = 58.8\%$

4.5.1. Probabilidad marginal

La probabilidad marginal es la probabilidad incondicional de un evento particular simple, que consiste en una suma de probabilidades conjuntas. Si se toma la probabilidad de que el siguiente cliente sea un hombre (ejercicio anterior), la probabilidad sería:

$$P(H) = P(H \cap E_1) + P(H \cap E_2) + P(H \cap E_3)$$

o sea: $$P(H) = 0.117 + 0.140 + 0.170 = 0.427 = 42.7\%$$

4.6 Teorema de Bayes

Cuando calculamos la probabilidad de B dado que A ya ocurrió, de alguna manera se piensa que el evento A es algo que sucede antes que B y que A puede ser (tal vez) causa de B o puede contribuir a su aparición. También de algún modo podemos decir que A normalmente ocurre antes que B. Pensemos, por ejemplo, que deseamos saber la probabilidad de que un estudiante apruebe el examen parcial de estadística dado que estudió por lo menos veinte horas antes del mismo.

En algunas ocasiones sabemos que ocurrió el evento B y queremos saber cuál es la probabilidad de que haya ocurrido el evento A. En nuestro ejemplo anterior la pregunta sería cuál es la probabilidad de que el alumno haya estudiado por lo menos veinte horas dado que, efectivamente, aprobó el examen parcial de estadística.

Esta probabilidad se encuentra aplicando una regla que se conoce como teorema de Bayes, mismo que se muestra enseguida.

$$P(A_i / B) = \frac{P(B/A_i)\cdot P(A_i)}{P(B/A_1)\cdot P(A_1)+P(B/A_2)\cdot P(A_2)+.............+P(B/A_k)\cdot P(A_k)}$$

En donde:

$P(A_i) =$ **Probabilidad previa**.

Es la probabilidad de un evento posible antes de cualquier otra información.

$P(B/A_i) =$ **Probabilidad condicional**.

Es la probabilidad de que el evento “B” ocurra en cada posible suceso de A_i.

$P(B/A_i)\cdot P(A_i) =$ **Probabilidad conjunta**.

La probabilidad de $(A_i \cap B)$ determinada por la regla general de la multiplicación.

$P(A_i / B) =$ **Probabilidad *a posteriori***.

Combina la información provista en la distribución previa con la que se ofrece a través de las probabilidades condicionales para obtener una probabilidad condicional final.

Ejemplo 1: Un gerente de crédito trata con tres tipos de riesgos crediticios con sus clientes: las personas que pagan a tiempo, las que pagan tarde (morosos) y las que no pagan.

Con base en datos estadísticos, las proporciones de cada grupo son 72.3%, 18.8% y 8.9%, respectivamente.

También, por experiencia, el gerente de crédito sabe que el 82.4% de las personas del primer grupo son dueños de sus casas: el 53.6% de los que pagan tarde, son dueños de sus casas, y el 17.4% de los que no pagan, también son propietarios de sus casas.

Calcular para el gerente de crédito la probabilidad de que un nuevo solicitante de crédito en un futuro, si es dueño de su casa:

a) Pague a tiempo.

b) Pague tarde.

c) No pague.

d) Elaborar su tabla de probabilidades.

Solución:

Definición de eventos:

$P_1 =$ Clientes que pagan a tiempo. $D =$ Clientes dueños de sus casas.

$P_2 =$ Clientes pagan tarde. $D' =$ Clientes no son dueños de sus casas

$P_3 =$ Clientes que no pagan.

Fórmula general: $$P(P_i / D) = \frac{P(D/P_i)\cdot P(P_i)}{P(D/P_1)\cdot P(P_1) + P(D/P_2)\cdot P(P_2) + P(D/P_3)\cdot P(P_3)}$$

a) Probabilidad de que un nuevo solicitante pague a tiempo.

Datos:

$$P_i = P_1 = 0.723$$
$$P_2 = 0.188$$
$$P_3 = 0.089$$
$$P(D/P_1) = 0.824$$
$$P(D/P_2) = 0.536$$
$$P(D/P_3) = 0.174$$

Sustituyendo en la fórmula general:

$$P(P_1/D) = \frac{0.824 \times 0.723}{0.824 \times 0.723 + 0.536 \times 0.188 + 0.174 \times 0.089} = \frac{0.596}{0.712} = 0.837 = 83.7\%$$

Un nuevo solicitante que sea propietario de su casa tendrá un 83.7% de probabilidades de que pague a tiempo.

b) Probabilidad de que un nuevo solicitante pague tarde:

$$P(P_2/D) = \frac{0.536 \times 0.188}{0.824 \times 0.723 + 0.536 \times 0.188 + 0.174 \times 0.089} = \frac{0.101}{0.712} = 0.142 = 14.2\%$$

Un nuevo solicitante que sea propietario de su casa tendrá un 14.2% de probabilidades de que pague tarde (cliente moroso).

c) Probabilidad de que un nuevo solicitante no pague.

$$P(P_3/D) = \frac{0.174 \times 0.089}{0.824 \times 0.723 + 0.536 \times 0.188 + 0.174 \times 0.089} = \frac{0.015}{0.712} = 0.021 = 2.1\%$$

Un nuevo solicitante que sea propietario de su casa tendrá un 2.1% de probabilidades de que nunca pague.

Esta información es de gran utilidad para determinar si aprobar o no una solicitud de crédito.

El denominador de la fórmula representa la probabilidad marginal del evento "D". Se puede indicar que un 71.2% de sus clientes son dueños de sus casas.

Se puede inferir también que una persona no "dueña de su casa" tendrá una probabilidad de pagar a tiempo de solo un 16.3% o de pagar tarde un 85.8% y de no pagar de un 97.9%.

Este análisis se puede elaborar con mayor facilidad si se utiliza una tabla de probabilidades tal como se muestra:

Evento P_i	**Probabilidad Previa** $P(P_i)$	**Probabilidad Condicional** $P(D/P_i)$	**Probabilidad Conjunta** $P(D/P_i)\cdot P(P_i)$	**Probabilidad *a posteriori*** $P(P_i/D)$
P_1	0.723	0.824	0.596	0.837
P_2	0.188	0.536	0.101	0.142
P_3	0.089	0.174	0.015	0.021
Total	1.000		0.712	1.000

Cuadro 4.2. Tabla de probabilidades del Teorema de Bayes.

El interés por el conocimiento de la teoría de la probabilidad nos permite obtener elementos de información verdaderamente útiles para su aplicación en las diversas situaciones de vida de tipo personal, profesional o social. La distinción de las variables aleatorias discretas o continuas así como las reglas de adición y de multiplicación dan como resultado una interpretación adecuada del concepto de probabilidad condicional, la cual tiene gran influencia en múltiples actividades de carácter comercial, industrial, o de servicios.

Las **tablas de probabilidad** conjunta son instrumentos muy valiosos para predecir el grado de probabilidad de ocurrencia de hechos supuestos de antemano. El concepto de probabilidad marginal nos conduce a comprender la probabilidad de un evento simple formado por la sumatoria de varios eventos conjuntos y es la base del Teorema de Bayes.

La utilización de este teorema nos permitirá descubrir la probabilidad de que un cierto evento haya sido la causa del evento que está ocurriendo o está por ocurrir. Los conceptos estudiados en este tema constituyen un importante soporte para el conocimiento de las distribuciones básicas de probabilidad de variables discretas o continuas que se verán más adelante.

Bibliografía del tema 4

BERENSON L., Mark, David LEVINE M. y Timothy Krehbiel C., *Estadística para administración*, 2ª edición, Prentice Hall, 2001.

LEVIN, Richard I. y David S. Rubin, *Estadística para administradores*, 6a. Edición, México, Prentice Hall, 1996.

MASON D., Robert, Douglas LIND A. y William MARCHAL G, *Estadística para administración y economía*, 11ª edición, Colombia, Alfaomega, 2004.

Actividades de aprendizaje

4.1 A partir del estudio de la bibliografía específica sugerida para el tema 4, desarrolla un mapa conceptual relacionando cada uno de los conceptos principales y sus aplicaciones.

4.2 Investiga y realiza por lo menos tres ejercicios o aplicaciones estadísticas de acuerdo con los conceptos estudiados utilizando la bibliografía sugerida.

4.3 Propón ejemplos de aplicación de probabilidad subjetiva en una empresa.

4.4 Proporciona ejemplos prácticos de utilización de las reglas de multiplicación en el ámbito comercial e industrial.

4.5 Estudia y resume el tema de probabilidades bajo condiciones de independencia estadística.

4.6 Estudia y resume el tema de probabilidades bajo condiciones de dependencia estadística.

4.7 Describe cuál es el propósito de la elaboración de diagramas de árbol (o arborigramas) con referencia en su utilidad para indicar la probabilidad conjunta de varios eventos.

4.8 Aplica el Teorema de Bayes en situaciones prácticas del campo empresarial y de negocios.

4.9 Investiga y elabora un diagrama de flujo general de la Teoría de la probabilidad.

Visita la página www.aulafacil.com y compara los temas estudiados con la propuesta que se expresa e indica tus conclusiones.

Cuestionario de autoevaluación

1. Indique la diferencia entre una probabilidad frecuencial y una probabilidad subjetiva.
2. ¿Cuáles son los pasos del procedimiento para calcular la probabilidad simple de un evento?
3. ¿Cuál es la diferencia entre eventos excluyentes y eventos independientes?
4. Explique las características de la regla de la adición.
5. Defina las propiedades de una probabilidad condicional.
6. Explique las características de la regla de la multiplicación.

7. ¿En qué consiste una tabla de probabilidad conjunta?
8. ¿En qué consiste una tabla de contingencia?, ¿cuál es su relación con una tabla de probabilidades?
9. ¿A qué se refiere cuando se habla de una probabilidad marginal?
10. ¿Cuáles son los objetivos de un teorema de Bayes?, ¿qué tipo de probabilidades intervienen?

Examen de autoevaluación

1. Durante el semestre se aplicó una encuesta a 500 estudiantes que cursaban una o más de las siguientes asignaturas: Administración, Contabilidad y Estadística. La distribución de los estudiantes en cada una de ellas fue la siguiente:

Administración	329
Contabilidad	186
Estadística	295

Administración y Contabilidad	83
Administración y Estadística	217
Contabilidad y Estadística	63

¿Cuántos estudiantes cursaban las tres asignaturas?

a. 53
b. 164
c. 103
d. 232
e. 314

2. Durante el semestre se aplicó una encuesta a 500 estudiantes que cursaban una o más de las siguientes asignaturas: Administración, Contabilidad y Estadística. La distribución de los estudiantes en cada una de ellas fue la siguiente:

Administración	329	Administración y Contabilidad	83
Contabilidad	186	Administración y Estadística	217

Estadística	295	Contabilidad y Estadística	63

¿Cuántos estudiantes cursaban Estadística, pero no Contabilidad?

a. 53
b. 164
c. 103
d. 232
e. 314

3. Una urna contiene los nombres de ocho estudiantes de Contaduría, tres de Administración y nueve de Informática. Si se extraen tres nombres al azar, la posibilidad de que éstos pertenezcan al área de Contaduría es:

a. 2.3%
b. 5.7%
c. 3.9%
d. 4.9%
e. 8.1%

4. Los porcentajes de primeros lugares en ventas obtenidos por los integrantes de un equipo de vendedores es el siguiente:

Vendedor	Primeros lugares
1	17%
2	22%
3	34%
4	18%
5	9%

La probabilidad de que los vendedores uno y dos no alcancen el primer lugar en el próximo ciclo es:

a. 17%
b. 22%
c. 39%
d. 61%
e. 42%

5. Van a cambiar al jefe de un área y existe un 60% de probabilidades que el candidato A lo sustituya, un 30% de que el candidato B ocupe el puesto y un 10% de que sea nombrado el candidato C. En función de quien sea el próximo jefe, la probabilidad de aumento de sueldo es la siguiente:

I. Con el candidato A del 5%
II. Con el candidato B del 20%
III. Con el candidato C del 60%

Entonces la probabilidad del aumento de sueldo será de:

a. 0.32
b. 0.195
c. 0.43
d. 0.23
e. 0.15

6. Se aplicó una encuesta en un despacho contable, la cual dio como resultado que el 35 % de los profesionistas son contadores; de ellos, un 30% tiene estudios de maestría. Entonces, la probabilidad de que un trabajador sea contador y tenga estudios de maestría es de:

a. 0.21
b. 0.857
c. 0.105
d. 0.33
e. 0.50

7. Se efectuará una prueba con dos prototipos de súpercomputadoras. La computadora Alfa es nueva con una probabilidad de falla al mes de 1%; la Beta con dos años de uso y una probabilidad de falla al mes del 5%, ¿cuál es la probabilidad de que ambas computadoras fallen el mismo mes?

a. 0.005%
b. 6%
c. 0.5%
d. 0.06%
e. 0.05%

8. Una envasadora de productos alimenticios perecederos estima que menos del 2.5% de los empaques llevan menor peso del establecido y menos del 4.5% llevan de más. Al seleccionar un empaque al azar, ¿cuál es la probabilidad de que esté dentro del peso establecido?

a. 7%
b. 93%
c. 2.5%
d. 4.5%
e. 95%

9. Se aplicó una prueba de máximo esfuerzo a dos equipos, cada uno integrado por 20 elementos. En el equipo "A" hay cinco competidores con antecedentes cardiacos; en el "B", sólo uno. Durante la prueba se detectó un competidor con un problema cardiaco, ¿cuál es la probabilidad de que pertenezca al equipo "A"?

a. 63%

b. 53%

c. 91%

d. 73%

e. 83%

10. En una prueba de esfuerzo máximo, aplicada a dos equipos se encontró que en el equipo "A" cinco de sus 20 competidores tienen antecedentes cardiacos; en el equipo "B", sólo uno. Si se selecciona un competidor al azar y resulta con antecedentes cardiacos, ¿cuál es la probabilidad de que éste sea del equipo "B"?

a. 5%

b. 95%

c. 17%

d. 43%

e. 83%

Tema 5. Distribuciones de probabilidad

Objetivo particular

Al terminar el tema, el alumno comprenderá los conceptos de variable discreta y continua, aprenderá las diferentes distribuciones teóricas de probabilidad que existen asociadas al tipo de variable y reconocerá las situaciones en que son aplicables cada una de ellas. Desarrollará la habilidad para calcular probabilidades haciendo uso de los diferentes tipos de distribuciones como la binomial, la de Poisson, la exponencial y la normal.

Temario detallado

5.1. Variables aleatorias, discretas y continuas

5.2. Distribuciones de probabilidad de variable discreta

5.3. Distribución binomial y distribución de Poisson

5.4. Distribuciones de probabilidad de variable continúa

5.5. Distribución normal y distribución exponencial

5.6. Ley de los grandes números.

Introducción

En esta tema se describen los diferentes tipos de distribuciones de probabilidad que existen, las técnicas para el cálculo o asignación de probabilidades aplicable para cada tipo de dato y cada situación, se analizan sus características y la aplicación de una de ellas en las diferentes situaciones que se presentan en el mundo de los negocios.

Una distribución de probabilidades da toda la gama de valores que pueden ocurrir con base en un experimento, y resulta similar a una distribución de frecuencias. Sin embargo, en vez de describir el pasado, define qué tan probable es que suceda algún evento futuro

5.1 Variables aleatorias, discretas y continuas

Una **variable** es **aleatoria** si los valores que toma corresponden a los distintos resultados posibles de un experimento; por ello el hecho de que tome un valor particular es un evento aleatorio.

Las variables se clasifican en variables cuantitativas o métricas y variables cualitativas o no métricas.

Las **variables cuantitativas** o métricas son aquellas cuya determinación está asociada a una unidad de medida; por ejemplo, la estatura de una persona, el número de habitantes de una población, el ingreso mensual de los individuos de un país, etcétera. Estas variables se subclasifican a su vez en variables discretas y continuas.

- Las **variables discretas o discontinuas** son aquellas que cuantifican la característica por medio de valores enteros y nunca mediante fracciones de los mismos. Como ejemplo de este tipo de variables tenemos el número de clientes de un banco, el número de hijos de una familia, el número de alumnos en un grupo de la universidad, el número de personas en una población rural, el número de automóviles en una ciudad, etcétera.

- Las **variables continuas** son aquellas que pueden tomar cualquier valor numérico, es decir, un valor entero o fraccionario dentro de un intervalo previamente especificado. Así, por ejemplo, la variable tiempo en una investigación podría medirse en intervalos de horas, o bien, en horas y minutos, o bien en horas, minutos y segundos según sea el requerimiento de la misma.

Las **variables cualitativas o no métricas** *e*specifican y miden cualidades en los individuos, lugares o cosas a partir de su descripción con palabras. Ejemplos de ellas son la variable sexo (masculino, femenino), la variable religión (católica, protestante, etc.), la variable idioma (español, inglés, francés, etc.), la variable estatus social (alto, medio, bajo), la variable calidad de un servicio (bueno, regular, malo), etcétera. Al igual que las variables cuantitativas, éstas pueden subclasificarse en variables cualitativas nominales y ordinales.

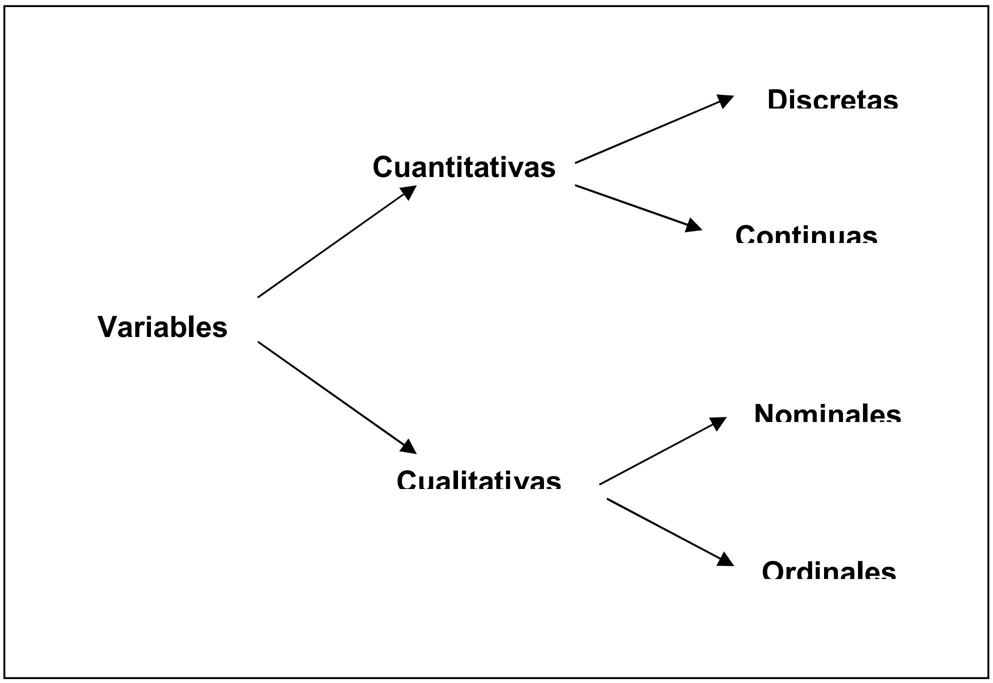
- Las **variables nominales** son aquellas variables no métricas usadas para describir una característica que no puede ser cuantificada numéricamente. Por ejemplo, el sexo de los individuos en un grupo universitario, su idioma, su religión, la carrera que cursan, etcétera; en algunos casos a estas variables se les asignan números, de acuerdo con una regla específica, pero estos números sólo se emplean para diferenciar a los distintos objetos o categorías. Así, la variable sexo podría clasificarse mediante dos categorías: 1 para los hombres y 2 para las mujeres. De esta forma, en una investigación todos los individuos 1 serían hombres y todos los 2 serían mujeres. Esta numeración en las variables nominales no permite ninguna operación aritmética o algebraica.

- Las **variables ordinales** son aquellas variables no métricas que permiten describir la característica de una persona, objeto o lugar a partir de diferenciarla en diversas categorías establecidas en orden de supremacía de acuerdo con un criterio jerárquico. La diferencia que se establece entre las categorías ordinales no tiene significado cuantitativo porque sólo expresan, por ejemplo, que una situación es mejor que otra, pero no cuantifican. Un ejemplo de ello es la evaluación que un cliente puede realizar sobre el servicio en un restaurante. El servicio fue malo, regular, bueno y muy bueno. Como puede observarse, la variable mide diversas categorías que han sido establecidas de acuerdo con un criterio, pero no puede establecer en ellas una cuantificación numérica, ya que la evaluación realizada por el cliente es subjetiva.

La clasificación de las variables anteriormente expuesta, que parte del punto de vista de la estadística, no es única, pues cada disciplina científica acostumbra hacer alguna denominación para las variables que en ella se manejan comúnmente.

Por ejemplo, en el área de las ciencias sociales es común establecer relaciones entre variables experimentales; por ello, estas ciencias, se clasifican desde el punto de vista metodológico en variables dependientes y variables independientes.

La **variable dependiente** es aquella cuyos valores están condicionados por los valores que toman la variable o las variables independientes con las que tiene relación.



Cuadro 5.1. Variables aleatorias

Por lo tanto, la variable o **las variables independientes** son la causa iniciadora de la acción, es decir, condicionan de acuerdo con sus valores a la variable dependiente.

Ejemplo*:* Consideremos el comportamiento del ahorro de un individuo en una sociedad. El modelo económico que explica su ahorro podría ser:

Ahorro = ingreso – gasto

En este modelo, el ahorro es la variable dependiente y presentará una situación específica de acuerdo con el comportamiento que tengan las variables independientes de la relación.

Sin embargo, no sólo es importante identificarlas y clasificarlas, sino también deberán definirse adecuadamente a partir del criterio establecido por el investigador. De acuerdo con Hernández, Fernández y Baptista, su definición deberá establecerse en dos niveles, especificados como nivel conceptual y nivel operacional.

Nivel conceptual. Consiste en definir el término o variable con otros términos. Por ejemplo, el término "poder" podría ser definido como "influir más en los demás que lo que éstos influyen en uno". Este tipo de definición es necesaria, pero insuficiente para definir una variable debido a que no nos relaciona directamente con la realidad, puesto que, como puede observarse, siguen siendo conceptos.

Nivel operacional. Constituye el conjunto de procedimientos que describen las actividades que un observador realiza para recibir las impresiones sensoriales que indican la existencia de un concepto teórico (conceptual) en mayor o menor grado, es decir, consiste en especificar las actividades u operaciones necesarias que deben realizarse para medir una variable.

Con estas dos definiciones, el estudiante o investigador estará ahora en posibilidad de acotar adecuadamente las variables para un manejo estadístico, de acuerdo con el interés que se tiene en ellas, para la realización de un estudio o investigación. Un par de ejemplos de ello los mostraremos a continuación.

Ejemplo 1

Variable: "Ausentismo laboral"

Nivel conceptual: "El grado en el cual un trabajador no se reportó a trabajar a la hora en la que estaba programado para hacerlo”.

Nivel operacional: "Revisión de las tarjetas de asistencia al trabajo durante el último bimestre".

Ejemplo 2:

Variable: "Sexo".

Nivel conceptual: "Condición orgánica que distingue al macho de la Hembra”

Nivel operacional: "Asignación de la condición orgánica: masculino o femenino".

5.2 Distribuciones de probabilidad discretas

La distribución de probabilidad de una variable aleatoria describe cómo se distribuyen las probabilidades de los diferentes valores de la variable aleatoria. Para una variable aleatoria discreta “**X**”, la distribución de probabilidad se describe mediante una **función de probabilidad,** representada por **f(X)**, que define la probabilidad de cada valor de la variable aleatoria.

5.3 Distribución binomial y Distribución Poisson

La distribución binomial se relaciona con un experimento aleatorio conocido como experimento de Bernoulli y que tiene las siguientes características:

- El experimento está constituido por un número de pruebas idénticas (por ejemplo, arrojar una moneda).

- Cada prueba tiene exactamente dos resultados posibles. A uno de ellos se le llama arbitrariamente éxito y al otro se le denomina fracaso.
- La probabilidad de éxito de cada prueba aislada es constante para todas las pruebas y recibe la denominación de “p”.
- La probabilidad de fracaso de cada prueba es, igualmente, constante para todas las pruebas y se le denomina “q”.
- Las características antes mencionadas hacen que p + q =1.
- Por medio de la distribución binomial tratamos de encontrar un número dado de éxitos en un número igual o mayor de pruebas.

La probabilidad de “x” éxitos en n intentos está dada por la siguiente fórmula:

$$P(x)= nCx\ p^{x}q^{(n-x)}.$$

Esta fórmula nos dice que la probabilidad de obtener “x” número de éxitos en n pruebas (como ya se indicó arriba) está dada por la multiplicación de n combinaciones de x (el alumno debe recordar el tema de reglas de conteo) por la probabilidad de éxito elevada al número de éxitos deseado y por la probabilidad de fracaso elevada al número de fracasos deseados.

A continuación se ofrecen varios ejemplos que nos ayudarán a comprender el uso de esta distribución.

Media y varianza de una variable aleatoria binomial

Media: $\mu = np$

Varianza: $\sigma^2 = npq$

Desde luego la desviación estándar es la raíz cuadrada de la varianza.

El siguiente ejemplo nos ayudará a entender este concepto.

Ejemplo 1: De acuerdo con estudios realizados en un pequeño poblado, el 20% de la población tiene parásitos intestinales. Si se toma una muestra de 1,400 ¿Cuántos esperamos que tengan parásitos intestinales?

$$\mu = np = 1,400 \times 0.20 = 280$$

Éste es el número de elementos de la muestra que tendría ese problema.

Usando el teorema de Tchebysheff (que se estudió en el apartado 2.6) podríamos considerar que el valor real estaría a dos desviaciones estándar con un 75% de probabilidades y a tres con un 89%. De acuerdo con ello obtenemos la desviación estándar y posteriormente determinamos los intervalos.

Desviación estándar: $\sigma = \sqrt{npq} = \sqrt{1,400 \times 0.20 \times 0.80} = \sqrt{224} = 14.97$

La media más menos dos desviaciones estándar nos daría el intervalo de 250 a 310 personas que tienen problemas.

La media más menos tres desviaciones estándar nos daría el intervalo de 235 a 325 personas que podrían tener problemas.

$\Rightarrow$ **Distribución de Poisson**

Es otra distribución teórica de probabilidad que es discreta y tiene muchos usos en economía y comercio. Se debe al teórico francés Simeón Poisson quien la derivó en 1837como un caso especial (límite) de la distribución binomial.

Se puede utilizar para determinar la probabilidad de un número designado de éxitos cuando los eventos ocurren en un espectro continuo de tiempo y espacio. Es semejante al proceso de Bernoulli excepto que los eventos ocurren en un espectro continuo en vez de ocurrir en ensayos u observaciones.

Como ejemplo tenemos el número de llamadas de entrada a un conmutador en un tiempo determinado, o el número de defectos en 10 m^2. de tela.

Sólo se requiere conocer el número promedio de éxitos para la dimensión especifica de tiempo o espacio de interés.

Este número promedio se representa generalmente por λ (lambda) y la fórmula de una distribución de Poisson es la siguiente:

$$P(x/\lambda) = \frac{\lambda^x \cdot e^{-\lambda}}{x!}$$

En esta fórmula, x representa el número de éxitos cuya probabilidad deseamos calcular; λ. es el promedio de éxitos en un tiempo o espacio; "e" es la base de los logaritmos naturales; y el símbolo de admiración representa la operación de factorial que ya vimos al estudiar las reglas de conteo.

Ejemplo 1: El manuscrito de un texto de estudio tiene un total de 40 errores en las 400 páginas de material. Los errores están distribuidos aleatoriamente a lo largo del texto. Calcular la probabilidad de que:

a) Un capítulo de 25 páginas tenga dos errores exactamente.
b) Un capítulo de 40 páginas tenga más de dos errores.
c) Una página seleccionada aleatoriamente no tenga errores.

Solución:

a) dos errores exactamente:

Datos

$40 - 400$

$\lambda - \quad 25$

$\therefore \ \lambda = 2.5$

$x = 2$

$e = 2.71828$

$$p(2/2.5) = \frac{2.5^2 \cdot 2.71828^{-2.5}}{2!} = 0.256 = 25.6\%$$

Existe un 25.6% de probabilidad de que un capítulo de 25 páginas tenga exactamente dos errores.

b) Más de dos errores

$$40 - 400 \qquad p(0/4) = \frac{4^0 \cdot 2.71828^{-4}}{0!} = 0.018 = 1.8\%$$

$$\lambda - \quad 40$$

$$\therefore \ \lambda = 4 \qquad P(1/4) = \frac{4^1 \cdot 2.71828^{-4}}{1!} = 0.073 = 7.3\%$$

$$x = 2$$

$$e = 2.71828 \qquad P(2/4) = \frac{4^2 \cdot 2.71828^{-4}}{2!} = 0.146 = 14.6\%$$

$$\therefore P(> 2/4) = 1 - (0.018 + 7.3 + 14.6) = 0.762 = 76.2\%$$

Existe un 76.2% de probabilidad de que un capítulo de 40 páginas tenga más de dos errores.

c) Una página no tenga errores:

Datos

$$40 - 400$$

$$\lambda - \quad 1$$

$$\therefore \ \lambda = 0.10 \qquad p(0/0.10) = \frac{0.10^0 \cdot 2.71828^{-0.10}}{0!} = 0.905 = 90.5\%$$

$$x = 2$$

$$e = 2.71828$$

Existe un 90.5% de probabilidad de que una sola página seleccionada aleatoriamente no tenga errores.

En un experimento de Bernoulli, tal como los que acabamos de estudiar en la distribución binomial, puede suceder que el número de ensayos sea muy grande y/o que la probabilidad de acierto sea muy pequeña y los cálculos se vuelven muy laboriosos. En estas circunstancias, podemos usar la distribución de Poisson como una aproximación a la distribución binomial.

Ejemplo 2: Una fábrica recibe un embarque de 1,000,000 de rondanas. Se sabe que la probabilidad de tener una rondana defectuosa es de .001. Indique usted, si obtenemos una muestra de 3000 rondanas, ¿cuál será la probabilidad de encontrar un máximo de tres defectuosas?

Solución*:*

Este ejemplo, desde el punto de vista de su estructura, corresponde a una distribución binomial. Dados los volúmenes y probabilidades que se manejan es conveniente trabajar con Poisson, tal como se realiza a continuación. Debemos recordar que un máximo de tres defectuosas incluye la probabilidad de encontrar una, dos y tres piezas defectuosas o ninguna

Media: $\mu = np = 3,000 \times 0.001 = 3$

$$P(x/\mu) = \frac{\mu^x \cdot e^{-u}}{x!}$$

$$P(0/3) = \frac{3^0 \cdot 2.71828^{-3}}{0!} = 0.0498 = 5.0\%$$

$$P(1/3) = \frac{3^1 \cdot 2.71828^{-3}}{1!} = 0.149 = 14.9\%$$

$$P(2/3) = \frac{3^2 \cdot 2.71828^{-3}}{3!} = 0.224 = 22.4\%$$

$$P(3/3) = \frac{3^3 \cdot 2.71828^{-3}}{3!} = 0.224 = 22.4\%$$

La probabilidad de encontrar un máximo de tres piezas defectuosas está dado por la suma de las probabilidades arriba calculadas, es decir: 0.647 ó 64.7% aproximadamente.

5.4 Distribuciones de probabilidad de variable continúa

Para comprender la diferencia entre las variables aleatorias discretas y las continuas recordemos, primero, que para una variable aleatoria discreta podemos calcular la probabilidad de que la variable aleatoria tome un determinado valor. Para las variables aleatorias continuas el caso es muy distinto porque puede asumir cualquier valor dentro de un intervalo de la recta numérica o de un conjunto de intervalos.

Como cualquier intervalo contiene una cantidad infinita de valores, no es posible hablar de la probabilidad de que la variable aleatoria tome un determinado valor; en lugar de ello, debemos pensar en términos de la probabilidad de que una variable aleatoria continua tome un valor dentro de un intervalo dado.

Para describir las distribuciones discretas de probabilidades presentamos el concepto de una función de probabilidad **f(x).** Recordemos que esta función da la probabilidad de que la variable aleatoria "**x**" tome un valor específico. En el caso continuo, la contraparte de la función de probabilidad es **la función de densidad de probabilidad** que también se representa por **f(x).** Para una variable aleatoria continua, la función de densidad de probabilidad especifica el valor de la función en cualquier valor particular de "**x**" sin dar como resultado directo la probabilidad de que la variable aleatoria tome un valor específico. Sin embargo, el área bajo la gráfica de **f(x)** que corresponde a un intervalo determinado sí determina la probabilidad de que la variable aleatoria continua tome un valor en ese intervalo.

En los ejemplos anteriores los valores que podía tomar la variable aleatoria eran cero o un numero natural: 1,2,3. Existen situaciones en las que la variable aleatoria puede tomar cualquier valor de número real. Para esos casos se utilizan las distribuciones de probabilidad continuas, en las que la curva normal es la más importante y tiene numerosas aplicaciones en la contaduría y la administración.

5.5 Distribución normal y distribución exponencial

$\Rightarrow$ Distribución normal

Esta distribución de probabilidad también es conocida como "Campana de Gauss" por la forma que tiene su gráfica y por el matemático que la desarrolló. Tal vez dé la impresión al alumno de ser un tanto complicada. No debe preocuparse por ello, pues para efectos de los dos cursos de estadística que cursa en nuestra Facultad, no es necesario usarla de manera analítica, sino comprender intuitivamente su significado.

De cualquier manera se dará una breve explicación de la misma para efectos de una mejor comprensión del tema. Su función aparece a continuación.

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

En esta función todos los términos son conocidos por el estudiante. La "y" es la ordenada de las coordenadas rectangulares cartesianas que el alumno ya conoce y que representa la altura sobre el eje "x"; x es la abscisa en este sistema de coordenadas; $\pi =$ 3.14159...; "e" corresponde a la base de los logaritmos naturales que el estudiante ya tuvo ocasión de utilizar en la distribución de Poisson.

Los símbolos σ y μ corresponden a la desviación estándar y la media que ya se vieron en el primer tema de esta materia. Podemos decir que ésta es la expresión de la ecuación normal, de la misma manera que y=mx+b es la expresión de la ecuación de la recta (en su forma cartesiana). Así como podemos sustituir m por cualquier pendiente (por ejemplo, 4) y b por cualquier ordenada al origen (por ejemplo, 2) para obtener una ecuación particular (en este caso y=4x+2); de la misma manera podemos sustituir σ y μ por cualquier par de valores para obtener un caso particular de la función normal. Si lo hacemos de esa manera (por ejemplo, dándole a la media un valor de cero y a la desviación estándar un valor de 1) podemos ir asignando valores a “x” (por ejemplo de –4 a 4) para calcular los valores de “y”. Una vez teniendo los valores de cada punto es fácil graficar en el plano cartesiano toda la colección de los mismos. Obtendremos una curva de forma acampanada. A continuación se muestran tanto los puntos como la gráfica para estos valores.

X	Y
-4.0	0.00013
-3.6	0.00061
-3.2	0.00238
-2.8	0.00792
-2.4	0.02239
-2.0	0.05399
-1.6	0.11092
-1.2	0.19419
-0.8	0.28969
-0.4	0.36827
0.0	0.39894
0.4	0.36827
0.8	0.28969
1.2	0.19419
1.6	0.11092
2.0	0.05399
2.4	0.02239
2.8	0.00792
3.2	0.00238
3.6	0.00061
4.0	0.00013

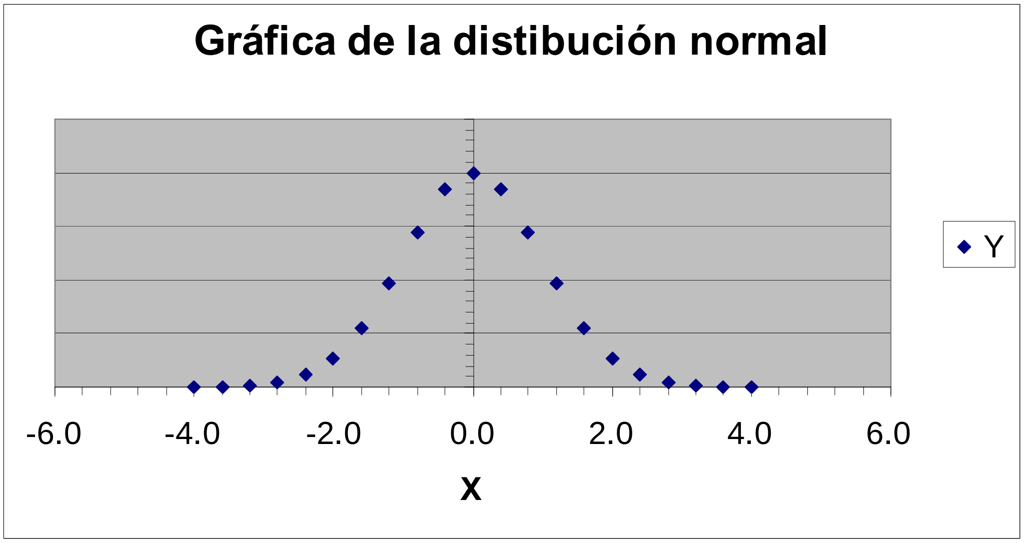


Figura 5.1. Curva normal

Es importante mencionar que el área que se encuentra entre la curva y el eje de las abscisas es igual a la unidad o 100%. La curva normal es simétrica en relación con la media. Esto quiere decir que la parte de la curva que se encuentra a la derecha de la curva es como una imagen reflejada en un espejo de la parte que se encuentra a la izquierda de la misma. Esto es importante, pues el área que se encuentra a la izquierda de la media es igual a la que se encuentra a la derecha de la misma y ambas son iguales a 0.5 o el 50%.

Para trabajar con la distribución normal debemos unir los conceptos de área bajo la curva y de probabilidad. La probabilidad de un evento es proporcional al área bajo la curva normal que cubre ese mismo evento. Un ejemplo nos ayudará a entender estos conceptos.

Con base en la Figura 5.1 vamos a suponer que el rendimiento de las acciones en la Bolsa de Valores en un mes determinado tuvo una media de 0% con una desviación estándar de 1%. (Esto se asimila a lo dicho sobre nuestra gráfica de una distribución normal con una media de cero y una desviación estándar de 1). De acuerdo con esta información, es mucho más probable encontrar acciones que ganen entre –2% y 2%, que acciones que ganen más o menos (ver las siguientes gráficas).

El valor exacto de las áreas y, por tanto, de las probabilidades de ocurrencia requeriría de herramientas de cálculo integral, pues se trata de calcular áreas bajo las curvas. Dado que se necesitaría calcular una integral distinta para cada media y cada desviación estándar diferente con la que nos topáramos, esto volvería nuestro trabajo muy laborioso. Afortunadamente existe una alternativa, que consiste en el uso de una tabla que contiene las áreas bajo la curva normal y que se encuentra en la mayoría de los libros de estadística. También se puede generar por medio de una hoja de cálculo electrónica. Al final de este tema se encuentra una tabla de la distribución normal, que se conoce como distribución normal estándar y que tiene una media igual a 0 y una desviación estándar igual a 1. En los próximos párrafos aprenderemos a utilizar la tabla de la distribución normal estándar.

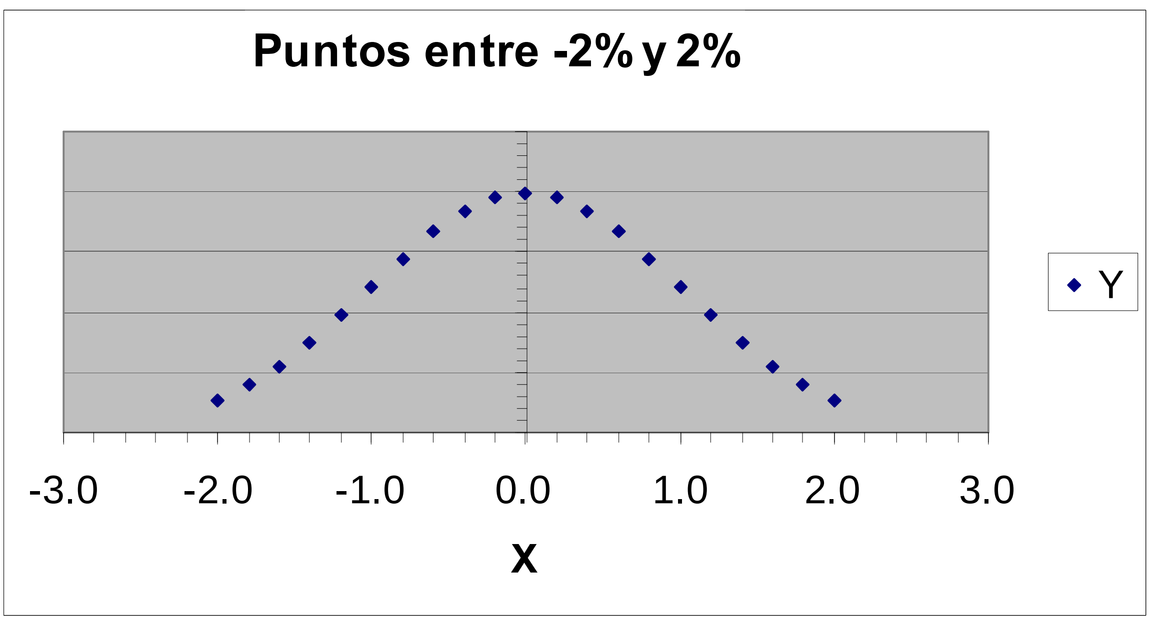


Figura 5.2 Distribución normal

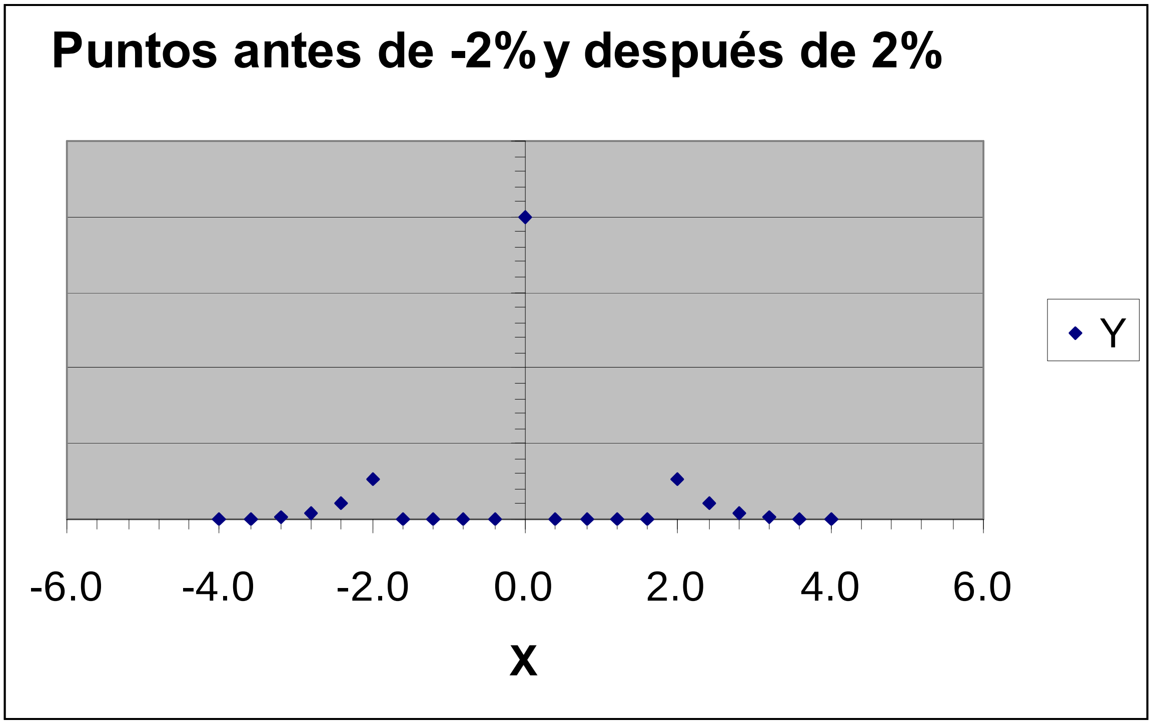


Figura 5.3 Distribución normal

La figura 5.3 nos muestra el **área** que hay entre la media y un conjunto de valores seleccionados de una variable a la que llamamos "**Z**". Al examinar la tabla de curva normal, (el alumno puede consultar la que aparece en el apéndice de esta unidad o la de cualquier libro de estadística), podemos observar que la columna de la extrema izquierda tiene, precisamente el encabezado de "Z". Los valores de la misma se van incrementando de un décimo en un décimo a partir de 0.0 y hasta 4.2 (en nuestra tabla, en otras puede variar). El primer renglón de la tabla también tiene valores de "Z" que se incrementan de un centésimo en un centésimo de .00 a .09. Este arreglo nos permite encontrar los valores del área bajo la curva para valores de "Z" de 0.0 a 4.29.

Así podemos ver que para Z=1 (primera columna del cuerpo de la tabla y renglón de 1.0), el área es de .34134. Esto quiere decir, que entre la media y una unidad de z a la derecha tenemos el 34.134% del área de toda la curva.

Por el mismo procedimiento podemos ver que para un valor de Z=1.96 (renglón de Z=1.9 y columna de Z=.06), tenemos el .47500 del área. Esto quiere decir que entre la media y una Z de 1.96 se encuentra el 47.5% del área bajo la curva normal. De esta manera, para cualquier valor de Z se puede encontrar el área bajo la curva.

La manera en que este conocimiento de la tabla de la distribución normal puede aplicarse a situaciones más relacionadas con nuestras profesiones se puede ver en el siguiente ejemplo.

Ejemplo 1*:* Una empresa tiene 2000 clientes. Cada cliente debe en promedio $7000 con una desviación estándar de $1000. La distribución de los adeudos de los clientes es aproximadamente normal. Diga usted cuantos clientes esperamos que tengan un adeudo entre $7000 y $8,500.

Solución:

Nos damos cuenta de que valores como 7000 o 1000 no aparecen en la tabla de la distribución normal. Es allí donde interviene la variable Z porque nos permite convertir los datos de nuestro problema en números que podemos utilizar en la tabla. Lo anterior lo podemos hacer con la siguiente fórmula:

$$z = \frac{x_i - \overline{x}}{\sigma}$$

En nuestro caso, nos damos cuenta de que buscamos el área bajo la curva normal entre la media, 7000, y el valor de 8500. Sustituyendo los valores en la fórmula obtenemos lo siguiente:

$$z = \frac{8,500 - 7,000}{1,000} = 1.5$$

Buscamos en la tabla de la normal el área bajo la curva para Z=1.5 y encontramos 0.43319. Esto quiere decir que aproximadamente el 43.3% de los saldos de clientes están entre los dos valores señalados.

En caso de que el cálculo de Z arroje un número negativo significa que estamos trabajando a la izquierda de la media. El siguiente ejemplo ilustra esta situación.

Ejemplo 2: En la misma empresa del ejemplo anterior deseamos saber que proporción de la población estará entre $6,500.00 y $7,000.00.

Solución:

Como en el caso anterior, nos damos cuenta de que nos piden el valor de un área entre la media y otro número. Volvemos a calcular el valor de Z

$$z = \frac{6,500 - 7,000}{1,000} = -0.5$$

Este valor de Z no significa un área negativa; lo único que indica es que el área buscada se encuentra a la izquierda de la media.

Buscamos el área bajo la curva en la tabla para Z=0.5 (positivo, la tabla no maneja números negativos) y encontramos que el área es de 0.19146. Es decir que la proporción de saldos entre los dos valores considerados es de aproximadamente el 19.1%.

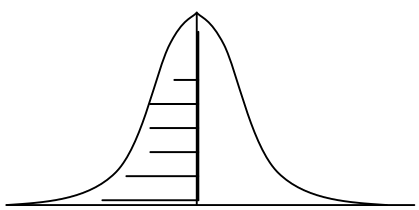
No siempre el área que se necesita bajo la curva normal se encuentra entre la media y cualquier otro valor. Frecuentemente son valores a lo largo de toda la curva. Por ello, es buena idea hacer un pequeño dibujo de la curva de distribución normal para localizar el o las áreas que se buscan. Esto facilita mucho la visualización del problema y, por lo mismo, su solución. A continuación se presenta un problema en el que se ilustra esta técnica.

Ejemplo 3: Una pequeña población recibe, durante la época de sequía, la dotación de agua potable mediante pipas que surten del líquido a la cisterna del pueblo una vez a la semana. El consumo semanal medio es de 160 metros cúbicos con una desviación estándar de 20 metros cúbicos. Indique cuál será la probabilidad de que el suministro sea suficiente en una semana cualquiera si se surten:

a) 160 metros cúbicos
b) 180 metros cúbicos
c) 200 metros cúbicos
d) Indique asimismo cual será la probabilidad de que se acabe el agua si una semana cualquiera surten 190 metros cúbicos.

Solución*:*

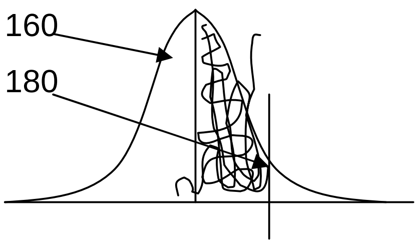
a) 160 metros cúbicos.



El valor de Z en este caso sería: $z = \frac{160 - 160}{20} = 0.0$ Esto nos puede desconcertar un poco; sin embargo, nos podemos dar cuenta de que si se surten 160 metros cúbicos el agua alcanzará si el consumo es menor que esa cifra. La media está en 160. Por ello el agua alcanzará en toda el área de la curva que se muestra rayada. Es decir, toda la mitad izquierda de la curva. El área de cada una de las mitades de la curva es de 0.5, por tanto la probabilidad buscada es también de 0 .5.

b) 180 metros cúbicos

El valor de Z es $z = \dfrac{180 - 160}{20} = 1.0$

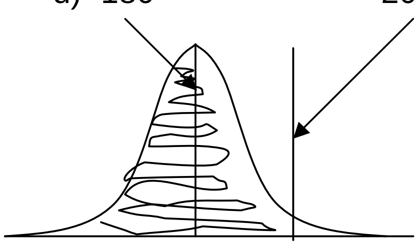


El área que se busca es la que está entre la media, 160, y 180. Se marca con una curva en el diagrama Si buscamos el área bajo la curva en la tabla de la normal, encontraremos el valor de 0.34134. Sin embargo, debemos agregarle toda la mitad izquierda dela curva (que por el diseño de la tabla no aparece). Ese valor, como ya se comentó es de .5. Por tanto, el valor buscado es de .5 más 0.34134. Por ello la probabilidad de que el agua alcance si se surten 180 metros cúbicos es de 0.84134, es decir, aproximadamente el 84%.

z	**0**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
1.0	**0.34134**	0.34375	0.34614	0.34849	0.35083	0.35314	0.35543	0.35769	0.35993	0.36214

c) 200 metros cúbicos

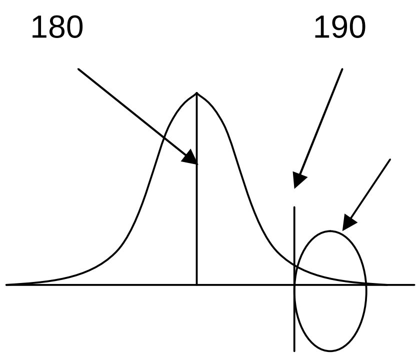
d) 180 200



El área buscada se señala en el dibujo. Incluye la primera mitad de la curva y la parte de la segunda mitad (la derecha) que se encuentra entre la media y 200. Ya sabemos que la primera mitad de la curva tiene un área de 0.5. Para la otra parte tenemos que encontrar el valor de Z y buscar el área correspondiente en la tabla. Z=(200-160)/20=2 es de 0.47725. Al sumar las dos partes nos queda .97725. Es decir, si se surten 200 metros cúbicos hay una probabilidad de casi 98% de que el agua alcance.

e) Indique asimismo cual será la probabilidad de que se acabe el agua si una semana cualquiera surten 190 metros cúbicos.

La probabilidad de que se termine el agua en estas condiciones se encuentra representada en la siguiente figura. El cálculo de la probabilidad de que ocurra se encuentra a continuación.



La probabilidad de que falte el agua está representada por el área en la cola de la distribución, después del 190. La tabla no nos da directamente ese valor. Para obtenerlo debemos de calcular Z para 190 y el valor del área entre la media y 190 restársela a .5 que es el área total de la parte derecha de la curva. Z=(190-160)/20= 1.5. El área para Z= 1.5 es de 0.43319. Por tanto, la probabilidad buscada es 0.5000 – 0.43319 = 0.06681 o aproximadamente el 6.7%.

⇒ **Búsqueda de Z cuando el área bajo la curva es conocida**

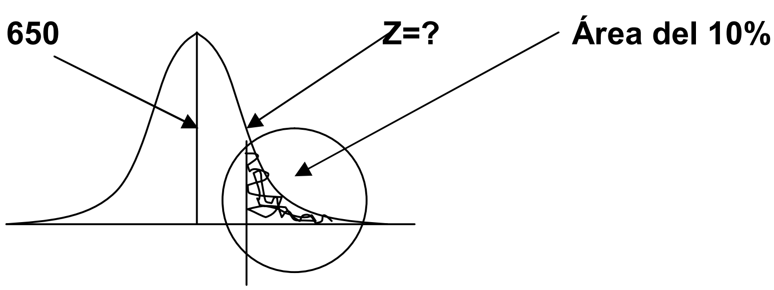
Frecuentemente el problema no es encontrar el área bajo la curva normal mediante el cálculo de Z y el acceso a la tabla para buscar el área ya mencionada. Efectivamente, a veces debemos enfrentar el problema inverso. Conocemos dicha área y deseamos conocer el valor de la variable que lo verifica. El siguiente problema ilustra esta situación.

Ejemplo 4: Una universidad realiza un examen de admisión a 10,000 aspirantes para asignar los lugares disponibles. La calificación media de los estudiantes es de 650 puntos sobre 1000 y la desviación estándar es de 100 puntos; las calificaciones siguen una distribución normal. Indique usted qué calificación mínima deberá de tener un aspirante para ser admitido si:

a) Se aceptará al 10% de los aspirantes con mejor calificación.
b) Se aceptará al 5% de aspirantes con mejor calificación.

Solución:

a) Si hacemos un pequeño esquema de la curva normal, los aspirantes aceptados representan el 10% del área que se acumula en la cola derecha de la distribución. El siguiente esquema nos dará una mejor idea.



El razonamiento que se hace es el siguiente: Si el área que se busca es el 10% de la cola izquierda, por lo mismo el área que debemos de buscar en la tabla es lo más cercano posible al 40%, esto es 0.4000 (esto se busca en el cuerpo de la tabla, no en los encabezados que representan el valor de Z. Este es el valor de **0.39973** y se encuentra en el renglón donde aparece un valor para Z de 1.2 y en la columna de 0.08. Eso quiere decir que el valor de Z que más se aproxima es el de 1.28. No importa si al valor de la tabla le falta un poco o se pasa un poco; la idea es que sea el más cercano posible.

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.2	0.38493	0.38686	0.38877	0.39065	0.39251	0.39435	0.39617	0.39796	**0.39973**	0.40147

Si ya sabemos el valor de Z, calcular el valor de la calificación (es decir “x”) es un problema de álgebra elemental y se trabaja despejando la fórmula de Z, tal como a continuación se indica.

Z= (x-650)/100. El alumno puede observar que ya sustituimos los valores de la media y de la desviación estándar. Ahora sustituimos el valor de Z y nos queda:

1.28 =(x-650)/100. A continuación despejamos el valor de x.
1.28(100)=128=x-650
128+650= x
Por lo mismo x=650+128; x=778.

En estas condiciones los aspirantes comenzarán a ser admitidos a partir de la calificación de 778 puntos en su examen de admisión.

b) El razonamiento es análogo al del inciso a. Solamente que ahora no buscamos que el área de la cola derecha sea el 10% del total sino solamente el 5% del mismo. Esto quiere decir que debemos buscar en la tabla en complemento del 5%, es decir 45% 0.45000. Vemos que el valor más cercano se encuentra en el renglón de Z de 1.6. y en la centésima 0.04

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.6	0.44520	0.44630	0.44738	0.44845	**0.44950**	0.45053	0.45154	0.45254	0.45352	0.45449

Esto nos indica que el valor de z que buscamos es el de 1.64. El despeje de x se lleva a efecto de manera análoga al inciso anterior, tal como a continuación se muestra.

1.64 =(x-650)/100. A continuación despejamos el valor de x.

1.64 (100)=164=x-650

164+650= x

Por lo mismo x=650+164; x=814.

En caso de que se desee mayor precisión se puede recurrir a interpolar los valores (por ejemplo, en este caso entre 1.64 y 1.65) o buscar valores más precisos en paquetes estadísticos de cómputo u hojas de cálculo electrónicas.

$\Rightarrow$ **Distribución exponencial**

En una distribución de Poisson los eventos ocurren en un espectro continuo de tiempo o espacio. Se considera entonces que son eventos sucesivos y la longitud de tiempo o espacio sigue una función exponencial.

Ésta se aplica cuando estamos interesados en el tiempo o espacio hasta el «primer evento», el tiempo entre dos eventos sucesivos o el tiempo hasta que ocurra el primer evento después de cualquier punto aleatoriamente seleccionado. Así, se presentan dos casos:

a) La probabilidad de que el primer evento ocurra dentro del intervalo de interés. Su fórmula es: $P(T \leq t) = 1 - e^{-\lambda}$

b) La probabilidad de que el primer evento *no* ocurra dentro del intervalo de interés. Su fórmula es:

$$P\left(T > t\right) = e^{-\lambda}$$

Ejemplo 1: Un departamento de reparaciones recibe un promedio de 15 llamadas por hora. A partir de este momento, cuál es la probabilidad de que:

a) En los siguientes 5 minutos no se reciba ninguna llamada.

b) Que la primera llamada ocurra dentro de esos 5 minutos.

c) En una tabla indicar las probabilidades de ocurrencia de la primera llamada en el en el minuto 1, 5, 10, 15, y 30.

Solución:

a) 1er. evento no ocurra:

$15 - 60$

$\lambda - \quad 5$

$\therefore \ \lambda = 1.25$

$x = 2$

$e = 2.71828$

$$P\left(T > t\right) = e^{-\lambda}$$

$$p\left(t > 5\right) = 2.71828^{-1.25} = 0.286 = 28.6\%$$

b) 1ª llamada en 5 minutos:

$$P\left(T \leq t\right) = 1 - e^{-\lambda}$$

$$P\left(T \leq 5\right) = 1 - 2.71828^{-1.25} = 1 - 0.287 = 0.713 = 71.3\%$$

c) Diversas probabilidades:

Espacio tiempo	λ	Probabilidad ocurra	Probabilidad No ocurra
1 minuto	0.25	0.221	0.779
5 minutos	1.25	0.713	0.287
10 minutos	2.50	0.918	0.082
15 minutos	3.75	0.976	0.024
30 minutos	7.50	0.999	0.001

5.6 Ley de los grandes números

La ley de los grandes números sugiere que **la probabilidad de una desviación significativa** de un valor de probabilidad determinado empíricamente, a partir de uno determinado teóricamente, es menor cuanto más grande sea el número de repeticiones del experimento.

El conocimiento de las diversas distribuciones de probabilidad de que dispone la estadística proporciona una importante contribución a comprender el comportamiento de ciertas variables discretas o continuas que nos permitan hacer inferencias sobre parámetros poblacionales con grados de confianza establecidos. Lo anterior tiene por consecuencia tener elementos de información estadística para una mejor toma de decisiones.

El conocimiento de las diversas distribuciones de probabilidad discreta o continua nos permite observar su comportamiento para obtener la probabilidad de ocurrencia de ciertos eventos de interés.

La distribución binomial se presenta en múltiples casos en los que se requiere conocer la probabilidad de ocurrencia de un número determinado de éxitos en una muestra aleatoriamente seleccionada. El complemento o sea la probabilidad de fallas se obtiene inmediatamente al restar a la unidad la probabilidad de éxitos.

La probabilidad de Poissón se utiliza preferentemente para casos de conocer la probabilidad de ocurrencia de eventos en un espectro de tiempo o espacio en múltiples situaciones de carácter comercial, financiero o de servicios.

La distribución normal constituye la base fundamental de la estadística inferencial ya que en una gran cantidad de situaciones de tipo aleatorio, los datos seleccionados o experimentados se distribuyen precisamente formando gráficamente una curva mesocúrtica, acampanada y simétrica, cuya área bajo esa curva constituye el 100% de probabilidad de ocurrencia de eventos. De cualquier sector de superficie de esta área se puede obtener el valor de la variable normalizada “z” y corresponde a un porcentaje de probabilidad.

Los procedimientos de cálculo para la curva normal son sumamente sencillos por lo que su utilización se prefiere incluso para aproximar a la distribución binomial, a la de Poisson y a otras distribuciones de probabilidad discreta

Bibliografía del tema 5

BERENSON L., Mark, David LEVINE M. y Timothy Krehbiel C., *Estadística para administración*, 2ª edición, Prentice Hall, 2001.

LEVIN, Richard I. y David S. Rubin, *Estadística para administradores*, 6a. Edición, México, Prentice Hall, 1996.

MASON D., Robert, Douglas LIND A. y William MARCHAL G, *Estadística para administración y economía*, 11ª edición, Colombia, Alfaomega, 2004.

Actividades de aprendizaje

Para los siguientes problemas, indica sí se trata de un problema binomial y si es así resuélvelo.

5.1 Un embarque de veinte televisores incluye tres unidades defectuosas. Si se inspeccionan tres televisores al azar, indique usted cuál es la probabilidad de que se encuentren dos defectuosos. (El alumno puede ya resolver este problema con las herramientas que adquirió en el tema de probabilidad elemental).

5.2. Una pareja de recién casados planea tener tres hijos. Diga usted cuál es la probabilidad de que los tres hijos sean varones si consideramos que la probabilidad de que el descendiente sea varón o hembra es igual.

5.3. Se sabe que el 30% de los estudiantes de secundaria en México es incapaz de localizar en un mapa el lugar donde se encuentra Afganistán. Si se entrevista a seis estudiantes de este nivel elegidos al azar:

a) ¿Cuál será la probabilidad de que exactamente dos puedan localizar este país?

b) ¿Cuál será la probabilidad de que un máximo de dos puedan localizar este país?

5.4 A partir del estudio de la bibliografía específica sugerida para el tema 5, presenta por escrito cada uno de los conceptos principales y sus aplicaciones.

5.5 Investiga el uso del valor esperado de una variable aleatoria en la toma de decisiones monetarias en condiciones de incertidumbre.

5.6 Estudia las diversas ilustraciones gráficas de la distribución binomial.
Practica el uso de tablas de la distribución binomial para facilitar cálculos.

5.7 Investiga las fórmulas de la esperanza matemática y varianza en una distribución de Poissón. Demuestre su utilización en ejemplos de aplicación.

5.8 Utilice la distribución de Poissón para aproximar a la distribución binomial.
Desarrolla ejemplos prácticos.

5.9 Utilice la distribución Normal para aproximar a la distribución binomial y a la de Poissón. Desarrolle ejemplos prácticos.

5.10 Investiga otros tipos de distribuciones discretas o continuas, y expresa sus características y fórmulas. Presente ejemplos de aplicación.

5.11 Visita la página www.aulafacil.com y compare los temas estudiados con la propuesta que se expresa e indica tus conclusiones.

Cuestionario de autoevaluación

1. Indique la diferencia entre las variables discretas y las variables continuas.
2. ¿A qué se refiere el nivel conceptual y el nivel operacional?
3. Exprese cuáles son las propiedades de una distribución binomial.
4. ¿En qué consiste una distribución de Poisson?

5. ¿En qué casos se utiliza una aproximación de la distribución de Poisson a la binomial?

6. ¿Qué es una distribución de probabilidad de variable continua?

7. Explique las características fundamentales y uso de la distribución normal.

8. Exprese la fórmula de la variable "z" parametrizada de una distribución normal.

9. Explique las características fundamentales y uso de la distribución exponencial.

10. Exprese la fórmula para obtener la probabilidad de éxito de un evento en una distribución exponencial.

Examen de autoevaluación

1. Las empresas mexicanas están aprovechando la condición de costo de mano de obra más barato para realizar trabajos en el extranjero; para ello utilizan los servicios de empresas de contratación locales para la resolución de todos los aspectos legales. Las encuestas realizadas por el Banco de Comercio Exterior indican que el 20% de las empresas mexicanas utilizan a este tipo de empresas. Si el banco selecciona al azar a un grupo de 15 empresas mexicanas. ¿Calcula la probabilidad de que exactamente cinco de ellas estén empleando a estas empresas locales?

a. 0.1032
b. 0.1058
c. 0.1028
d. 0.1035
e. 0.1038

2. De acuerdo con la situación anterior, calcula la probabilidad de que el número de empresas mexicanas que contratan empresas locales en el extranjero se ubique entre seis y nueve.

a. 0.6512
b. 0.6094
c. 0.6318

d. 0.6075

e. 0.6095

3. El Banco Nacional de México sabe por su experiencia que durante los días lunes, entre las 9.00 y las 10.00, se presentan a la ventanilla de atención a clientes un promedio de 2.8 clientes cada 4 minutos, número que la cajera puede atender con eficiencia. Con el propósito de verificar si el número de cajeras es el adecuado, calcula la probabilidad de que se presente un total de cuatro clientes en un intervalo de cuatro minutos.

a. 0.1568

b. 0.1557

c. 0.1535

d. 0.1678

e. 0.1456

4. Un auxiliar de contador puede cometer 1.2 errores por cada 200 declaraciones fiscales. ¿Calcula la probabilidad de que al seleccionar una de las declaraciones elaboradas por él no se encuentre algún error?

a. 0.596%

b. 0.58%

c. 0.585%

d. 0.599%

e. 0.589%

5. La probabilidad de que un banco cometa un error al procesar un depósito es igual a 0.0003. Si se auditan 10,000 depósitos, ¿cuál es la probabilidad de que se cometan más de seis errores? Utiliza la aproximación de Poisson a la binomial.

a. 0.0335

b. 0.0216

c. 0.0330

d. 0.0279

e. 0.0335

6. Un banco recibe en promedio a 3.2 clientes cada 4 minutos. ¿Cuál es la probabilidad de obtener exactamente 10 clientes en los próximos 8 minutos?

a. 0.0538

b. 0.0635

c. 0.0535

d. 0.0525

e. 0.0528

7. Una persona presentará el examen de conocimientos y dominio de la lengua inglesa, denominado GMAT; sus resultados tienen un valor medio de 494 puntos con desviación estándar de 100; La persona desea conocer la probabilidad de obtener 700 puntos. Considera que los resultados siguen una distribución normal estándar.

a. 0.0255

b. 0.0204

c. 0.0197

d. 0.0199

e. 0.0193

8. Una empresa de seguros está considerando incluir entre los riesgos cubiertos, una enfermedad denominada Túnel Carpiano, la cual aparece en manos y muñecas, provocada por los esfuerzos realizados con estas partes del cuerpo durante tiempos prolongados. Se estima que el costo de tratamiento de estas afecciones es alrededor de $30,000 pesos al año por trabajador lesionado, con una desviación estándar de $9,000.00. La aseguradora supone que la afección está normalmente distribuida y desea estimar los costos en que puede incurrir.

Calcula la probabilidad de que el costo de atención se encuentre entre $15,000 y $ 45,000.

a. 0.9050
b. 0.0.095
c. 0.9152
d. 0.9070
e. 0.9030

9. Una empresa de automóviles menciona en su publicidad que sustituirá por una unidad nueva los autos que presenten cualquier tipo de falla en el tren motriz durante los primeros 80 000 kilómetros. Si la empresa sabe que el valor medio del kilometraje sin fallas es de 80 000 kilómetros y la desviación estándar de 10 000 kilómetros, ¿cuál debería ser el kilometraje garantizado para no tener que reponer más del 10% de los autos?

a. 65 600
b. 68 300
c. 67 200
d. 68 450
e. 63 600

10. Una distribución normal tiene una media de 4.9 y una desviación estándar de 1.2. ¿Qué porcentaje del área bajo la curva es mayor que 6?

a. 0.168
b. 0.181
c. 0.179
d. 0.175
e. 0.175

Tema 6. Número índice

Objetivo particular

Al finalizar el tema, el alumno será capaz de aplicar los números índice en la solución de problemas reales relacionados con las áreas contable y administrativa.

Temario detallado

6.1 Tipos de números índice

- 6.1.1. Índice de cantidad
- 6.1.2. Índice de valor
- 6.1.3 Índice agregado e índice simple
- 6.1.4.Índice compuesto

6.2 Índices ponderados

6.3 Índices de precios al consumidor

Introducción

Los gobiernos y otras entidades publican diversas clases de índices. Estos están elaborados con el propósito de presentar de manera sencilla el **comportamiento** de alguna o algunas **variables de interés**. El alumno seguramente habrá escuchado mencionar el Índice Nacional de Precios al Consumidor, el Índice de Precios y Cotizaciones (IPC) de la Bolsa Mexicana de Valores o el Dow Jones que se relaciona con el mercado de valores de Nueva York.

Estos índices y otros que el alumno puede conocer son muy útiles para los profesionales de la contaduría y la administración, pues son elementos de juicio para la toma de decisiones. Es importante mencionar que un solo número índice nos arroja muy poca información. Por ejemplo, si alguien nos dice que el IPC de la Bolsa Mexicana de Valores cerro hoy a 4900 puntos, no revela una información que pueda ser usada para tomar decisiones. Lo importante es saber cómo se ha comportado el índice a lo largo de los días; es decir, ¿subió?, ¿bajó? Así, la información de los índices nos es útil en cuanto podemos ver su comportamiento en el tiempo. Podemos decir que un **índice** conforma una **serie de tiempo**. Una serie de tiempos es un conjunto de datos recopilados y utilizados en orden cronológico. De esta manera el estudio de un índice a través del tiempo nos proporciona una idea de la dinámica de los fenómenos que el propio índice contempla.

6.1 Tipos de números índice

En primer término, y como antecedente, vamos a mencionar algunos aspectos relacionados con la construcción de un índice.

Dado que lo importante de un índice es observar su comportamiento en el tiempo, la elección del periodo que va a servir de base es muy importante. Vamos a suponer que deseamos desarrollar un índice que refleje la ocupación de los hoteles y que definimos nuestro índice de ocupación como:

Porcentaje de ocupación de 1980=100.

Si el año de 1980 fue bueno, entonces la mayoría de los otros años se verá “mal” en comparación, pues tendrán valores menores a cien. En cambio, si el año de 1980 fue muy malo, la mayoría de los años se verá “bien” pues tendrá valores mayores a cien. Si deseamos construir un índice que no resulte engañoso, debemos elegir un periodo base que no sea ni exageradamente bueno ni exageradamente malo. Usualmente el índice del periodo base es igual a 100.

Otra consideración que debemos hacer al construir un numero índice es que la razón básica de su utilización es la de resumir circunstancias, a veces muy complejas, en un solo número que sea fácil de comprender y de manejar. Por ello debemos de tomar la decisión de lo que queremos reflejar en él.

Si deseamos reflejar la fluctuación en la cantidad de bienes o servicios seleccionados (cantidad de automóviles, cantidad de consultas médicas, etc.) debemos construir un índice de cantidad.

Si, en cambio, deseamos reflejar los cambios en el valor total de un grupo de bienes o servicios, crearemos un índice de valor. Por ejemplo, valor total de los automóviles vendidos en un año, valor total de los servicios médicos proporcionados en un año.

Cuando usamos un índice para reflejar un solo bien, estaremos construyendo un **índice simple**. En cambio, si conjuntamos varios bienes en el mismo índice, estaremos trabajando un índice agregado.

A veces no existe un único índice que satisfaga nuestras necesidades, pero existen dos o tres (o más) de ellos que contemplan la información que necesitemos. En ese caso podemos conjuntar estos índices para formar el que nos interesa. Esto es conocido como índice compuesto.

De acuerdo con lo comentado en la sección anterior, existen diversos tipos de índices útiles para el profesionista en contaduría y administración. A continuación se explican algunos de ellos.

6.1.1. Índice de cantidad

Este índice mide los cambios en las unidades de un bien de acuerdo con su origen, destino, utilización, etc. Todo ello a través del tiempo.

La expresión que en seguida aparece nos muestra el manejo formal de este tipo de índices.

$$IQ = \frac{Qi}{Qb}$$

En donde Qi es la cantidad del bien en el periodo que se desea obtener y Qb es la cantidad del bien que se toma como base. A continuación se presenta un ejemplo que nos permitirá una mejor comprensión de los índices de cantidad.

Ejemplo 1. Una empresa que vende autobuses de pasajeros ha decidido establecer un índice de cantidades vendidas. Se acordó como periodo base el mes de junio de 2001. En la primera columna se muestran las unidades vendidas; en la segunda, el índice propiamente dicho. Para los renglones marcados con asterisco, se ilustra debajo de la tabla la manera cómo se calculó el índice.

VENTAS	(UNIDADES)	INDICE
Marzo 01	93	109.41
Abril 01	81	95.29*
Mayo 01	78	91.76
Junio 01	85	100.00
Julio 01	90	105.88
Agosto 01	94	110.59
Septiembre 01	84	98.82
Octubre 01	89	104.71**
Noviembre 01	92	108.24

$$*\ \frac{81}{85}100 = 95.29 \qquad **\ \frac{89}{85}100 = 104.71$$

6.1.2. Índice de valor

Este índice mide en unidades monetarias (pesos, dólares, etc.) el valor (ya sea de costo o de precio de venta, según el caso) de un conjunto de bienes y/o servicios. Es importante mencionar que en este tipo de índices se toma en cuenta **el valor de cada** bien que se incluye en el índice. La siguiente expresión nos da la manera de calcular el índice de valor para cualquier periodo “i”.

$$IV = \frac{\sum PiQi}{\sum PbQb}(100)$$

Donde PiQi son respectivamente el precio de un bien (o su costo, como ya se mencionó) y PbQb son el precio y el número de unidades que se considera en el año base, tal como se muestra en el siguiente ejemplo.

Ejemplo 2. Una empresa que fabrica automóviles y camiones debe comparar su producción del último mes con la del mes base, que, para nuestro ejemplo, supondremos que fue el de abril de 1995.

Producción del mes base	**Qb Cantidad del mes base**	**Pb Costo unitario mes base**	**Pb Qb**
Auto modelo 1	200	93000	18,600,000
Auto modelo 2	250	142,000	35,500,000
Auto modelo 3	330	82,000	27,060,000
Camión modelo 1	110	240,000	26,400,000
Camión modelo 2	90	290,000	26,100,000
	Σ Pb Qb		133,660,000

En el último mes ya no se producía el modelo 1 del automóvil y se había introducido el modelo 4. Los datos se proporcionan a continuación.

	Qi	**Pi**	**Pi Qi**
Auto modelo 2	220	154,000	33’880,000
Auto modelo 3	210	86,000	18’060,000
Auto modelo 4	300	99,000	29’700,000
Camión modelo 1	130	251,000	32’630,000
Camión modelo 2	120	302,000	36’240,000
	Σ Pi Qi		150’510,000

Producción del último mes

Con los datos anteriores ya podemos obtener el índice de valor que es:

$$IV = \frac{150{,}510{,}000}{133{,}660{,}000} = 112.61$$

6.1.3 Índice agregado e índice simple

Un índice agregado es aquel que agrupa o resume el comportamiento de varios bienes o servicios. Un índice simple es aquel que se utiliza para describir el comportamiento de un solo bien o servicio. El ejemplo 1 de la sección 6.1.1 nos habla de un índice simple, en tanto que el ejemplo 2 de la sección 6.1.2 es un índice agregado; nos podemos dar cuenta que en el ejemplo 1 nos habla únicamente de autobuses en tanto que el ejemplo 2 trata de varios bienes, pues se manejan diversos modelos de coches y camiones, es decir, varios artículos.

Para reafirmar la diferencia entre un índice simple y uno agregado, a continuación se presentan los siguientes ejemplos.

Ejemplo 1. Índice simple

El precio de la tonelada de maíz fluctúa de acuerdo con las condiciones del mercado internacional. A continuación se listan los precios por tonelada al cierre de algunos meses, se señala el mes base y se indican los índices correspondientes.}

Mes	Año	S / tonelada	Índice
Febrero	X 2	1210	100.83
Marzo	X 2	1180	98.33
Abril	X 2	1040	86.67
Mayo	X 2	1314	109.50
Junio	X 2	1200*	100.00
Julio	X 2	1190	99.17
Agosto	X 2	1220	101.67

* mes base

Ejemplo 2. Índice agregado

Un estudio indicó que para transportarse de su casa a la escuela, a la biblioteca u otros lugares, el estudiante universitario promedio utiliza los transportes que abajo se detallan y deseamos crear un índice del costo de transporte.

	Precios unitarios (periodo base)	**Costo total**
8 viajes en metro	$1.50	$12.00
6 viajes en pesera	$2.50	$15.00
4 viajes en autobús	$2.00	$ 8.00
2 viajes en taxi	$10.00	$20.00

El costo de transporte en la semana fue (en el periodo base) de $55.00 esto corresponde al 100 en nuestro índice de costo de transporte (ICT).

Seis meses después se ratifican los precios unitarios y se obtiene la siguiente información.

		Viajes semana*	**Costo total**
Metro	$2.00	8	$16.00
Pesera	$2.00	6	$12.00
Autobús	$3.00	4	$12.00
Taxi	$9.00	9	$18.00

El costo del transporte en la semana fue $ 58.00

* De la tabla anterior

El índice de este periodo será: $ICT = \frac{58.00}{55.00}100$

$$ICT = 1.055$$

6.1.4.Índice compuesto

Se puede presentar la posibilidad de que no exista un índice ya publicado por alguna agencia de gobierno o institución de investigación privada que satisfaga las necesidades particulares de una empresa; sin embargo, existen dos o mas índices que si están publicados y que satisfacen, por partes, las necesidades de información de esta empresa.

Vamos a suponer que una empresa tiene el 70% de sus negocios en la Ciudad de México y el 30% restante en la ciudad de Puebla. Para tomar decisiones acertadas, el gerente de esta empresa necesita un índice de precios que combine ponderadamente los índices de precios de ambas ciudades. A continuación se da un ejemplo de cómo lograrlo.

Índice de precios al consumidor de la ciudad de México en abril de 20x2 214.382
Índice de precios al consumidor de la ciudad de Puebla en abril de 20x2 208.214

Índice compuesto = (0. 70) 214.382 + (0.30) 208.214 ← Índice de la Cd. de Puebla

Peso del índice De la Cd. de Puebla

Peso del Índice de la Cd. de México

Índice de la Cd. de México

Índice compuesto = 150.067 + 62.464 = 212.531

Este último valor es el índice compuesto lo podemos generar mediante la siguiente formula:

$$IC = \sum IiPi$$

En donde:

IC Es el índice compuesto

Ii Es cada uno de los índices que queremos que tomen parte del índice compuesto.

Pi Es el peso o proporción que se da a ese índice en el propio índice compuesto.

Σ es el signo de sumatoria que ya conocemos

Debemos hacer mención al hecho de que la suma de todos los pesos debe ser igual a la unidad, es decir Σ Pi =1

Para comprender mejor estos conceptos, a continuación se presenta el siguiente ejemplo:

Ejemplo 1. El gerente de la empresa que tenia negocios en Puebla y en la Ciudad de México, abrió recientemente una sucursal en Querétaro. Ahora sus negocios se reparten de la siguiente manera: 50% en la Cd. de México; 30% en Puebla y 20% en Querétaro. Los índices de precios de las tres ciudades se detallan a continuación para el mes de junio de 20X2.

	Ii	**Pi**	**Ii Pi**
I.P.C. Cd México	221.310	0.50	110.655
I.P.C Puebla	215.240	0.30	64.572
I.P.C Querétaro	218.700	0.20	43.740
		Σ Ii Pi = 1	218.967

El índice compuesto es ahora: IC =218.967

6.2 Índice ponderados

El valor de un índice agregado simple no se usa frecuentemente porque puede estar influido por las unidades de medida; por ello, se necesita un medio de «ponderar» adecuadamente los artículos según su importancia relativa.

Entre los índices compuestos ponderados que más se utilizan se encuentran los que se refieren a las variaciones de precios. Los más importantes son los de Laspeyres y de Paasche.

La característica común a estos índices y a la mayoría de los índices de precios es que utilizan valores como coeficientes de ponderación; es decir, datos que se pueden expresar como producto de un precio por una cantidad.

Estas ponderaciones se establecen para tener en cuenta las cantidades vendidas de cada producto. Lo anterior proporciona un reflejo más exacto del costo verdadero de la canasta típica del consumidor.

$\Rightarrow$ **Índice de Laspeyres.**

Utiliza las cantidades vendidas en el año base como ponderaciones y permite comparaciones más significativas con el tiempo.

Su fórmula es la siguiente: $$L = \frac{\sum p_n q_o}{\sum p_o q_o} \times 100$$

en donde: L es el índice de precios de Laspeyres.

p_n es el precio actual.

p_o es el precio base.

q_o es la cantidad vendida en el periodo base.

En siguiente ejemplo se ilustra este índice:

Una empresa posee una planta empacadora de carne y sus 3 principales productos tienen las siguientes cantidades de precio y venta en los últimos 3 años:

		Precio ($) / unidad			**Cantidad vendida**		
Artículo	**Unidad**	2003	2004	2005	2003	2004	2005
Res	kilo	30	33	45	250	320	350
Cerdo	kilo	20	22	21	150	200	225
Ternera	kilo	40	45	36.4	80	90	70

	Precio ($) / unidad			**Cantidad**	$p_n q_o$		
Artículo	2003	2004	2005	2003	2003	2004	2005
Res	30	33	45	250	7,500	8,250	11,250
Cerdo	20	22	21	150	3,000	3,300	3,150
Ternera	40	45	36.4	80	3,200	3,600	2,912
$\sum$					13,700	15,150	17,312

El índice para el año 2003 será:

$$L_{2003} = \frac{\sum p_{2003} q_{2003}}{\sum p_{2003} q_{2003}} \times 100 = \frac{13,700}{13,700} = 100$$

El índice para el año 2004 utiliza los precios del año de referencia (2004) y las cantidades en el año base (2003) para el numerador:

$$L_{2004} = \frac{\sum p_{2004} q_{2003}}{\sum p_{2003} q_{2003}} \times 100 = \frac{15,150}{13,700} = 110.58$$

El índice para el año 2005 utiliza los precios del año de referencia (2005) y las cantidades en el año base (2003) para el numerador:

$$L_{2005} = \frac{\sum p_{2005}q_{2003}}{\sum p_{2003}q_{2003}} \times 100 = \frac{17,312}{13,700} = 126.36$$

Este índice indica que del año 2003 al 2004, el precio de la canasta para estos 3 artículos se incrementó en un 10.58%. Se gastaría $110.58 en 2004 lo que con $100.00 compraba en 2003.

También indica que del año 2003 al 2005, el precio de la canasta para estos 3 artículos se incrementó en un 26.36%. Se gastaría $126.36 en 2005 lo que con $100.00 compraba en 2003.

Se hace notar que el denominador es el mismo para cualquier año, ya que el índice de Laspeyres siempre utiliza cantidades del periodo base.

$\Rightarrow$ **Índice de Paasche**

Utiliza como ponderaciones las cantidades vendidas en cada uno de los años de referencia y tiene la ventaja de de que el índice se basa en los patrones de comportamiento del consumidor común. A medida que los consumidores cambian sus hábitos de compra, los gastos que efectúan se reflejan directamente en el índice.

Su fórmula es la siguiente: $$P = \frac{\sum p_n q_n}{\sum p_o q_n} \times 100$$

en donde: p es el índice de precios de Paasche.

p_n es el precio actual.

p_n es el precio base.

q_n es la cantidad vendida en el periodo base.

En el ejemplo de la empresa de carne, utilizamos sus datos para calcular el índice de Paasche, mediante la siguiente tabla:

	2003		2004		2005	
Artículo	*p*	*q*	*p*	*q*	*p*	*q*
Carne de res	*30*	*250*	*33*	*320*	*45*	*350*
Cerdo	*20*	*150*	*22*	*200*	*21*	*225*
Ternera	*40*	*80*	*45*	*90*	*36.4*	*70*

$p_{03}q_{03}$	$p_{04}q_{04}$	$p_{05}q_{05}$	$p_{03}q_{04}$	$p_{03}q_{05}$
7,500	*10,560*	*15,750*	*9,600*	*10,500*
3,000	*4,400*	*4,725*	*4,000*	*4,500*
3,200	*4,050*	*2,548*	*3,600*	*2,800*
13,700	*19,010*	*23,023*	*17,200*	*17,800*

El índice para el año 2003 será:

$$P_{2003} = \frac{\sum p_{03}q_{03}}{\sum p_{03}q_{03}} \times 100 = \frac{13,700}{13,700} = 100$$

El índice para el año 2004 será:

$$P_{2004} = \frac{\sum p_{04}q_{04}}{\sum p_{03}q_{04}} \times 100 = \frac{19,010}{117,200} = 110.5$$

El índice para el año 2005 será:

$$P_{2005} = \frac{\sum p_{05}q_{05}}{\sum p_{03}q_{05}} \times 100 = \frac{23,023}{17,800} = 129.3$$

Este índice indica que del año 2003 al 2004, el precio de la canasta para estos 3 artículos se incrementó en un 10.5%. Se gastaría $110.50 en 2004 lo que con $100.00 compraba en 2003. También indica que del año 2003 al 2005, el precio de la canasta para estos 3 artículos se incrementó en un 29.3%. Se gastaría $129.30 en 2005 lo que con $100.00 compraba en 2003.

El índice de Laspeyres tiende a sobreponderar los bienes cuyos precios se incrementan. Lo anterior ocurre porque el incremento en el precio tiende a reducir las cantidades vendidas, pero la cantidad menor no se refleja en el índice de Laspeyres debido a que utiliza las cantidades del año base.

6.3 Índice de precios al consumidor

Este índice es publicado mensualmente en Estados Unidos por la Dirección de Estadísticas Laborales (Bureau of Labor Statistics). A esto es lo que generalmente se conoce en los medios noticiosos como el índice del costo de la vida.

En México se utiliza el Índice nacional de precios al consumidor (INPC) el cual estima la evolución de los precios de una canasta de bienes y servicios que se consumen en un periodo determinado. Es elaborado por el banco de México desde 1927 donde se investigaban 16 artículos alimenticios en la ciudad de México. Donde se investiga 4 veces al mes los precios de 27,140 productos alimenticios. Dos veces al mes los precios de productos no alimenticios, investigándose en total 170,000 precios en 50 ciudades. Además del índice general se elaboran 73,000 índices intermedios. 313 bienes y servicios genéricos que consumen las familias

Es común encontrar cláusulas de “**aumentos graduales**” en los contratos colectivos de trabajo que ligan los aumentos salariales al **índice nacional de precios al consumidor** (**INPC**). Más aún, las cuotas al Seguro Social y algunas otras incluyen cláusulas relativas a cambios que toman en consideración el nivel del INPC.

Los valores del índice de precios al consumidor se expresan como promedios anuales y mensuales.

Este índice se puede usar de múltiples maneras. Un uso común es para medir “el poder adquisitivo del consumidor” o el de la moneda.

El INPC se utiliza también para medir el ingreso "real", que es el ingreso ajustado para cambios en los precios. De este modo, dividir el salario neto entre el valor corriente del INPC en cualquier año revelará el ingreso real para ese año. También es posible hacer una comparación entre años.

Todos los índices de precios al consumidor son índices agregados. Para entender como se trabaja un índice de precios al consumidor, haremos uso de un ejemplo sencillo simplificado.

Ejemplo 1: Supongamos que vivimos en una pequeña población que tiene un solo almacén general para surtir a los pobladores de todos los bienes, tanto de primera necesidad como suntuarios. En esta población deseamos comenzar a llevar un índice de precios al consumidor. Para hacerlo, lo primero que necesitamos hacer es definir una canasta de bienes y servicios que la población en general consume o compra con regularidad. La conformación de esta canasta y el peso de cada artículo es fuente de polémica en todos los casos. En nuestro caso supondremos que ya nos pusimos de acuerdo y que la canasta de consumo mensual familiar aparece en la siguiente tabla.

ARTÍCULO	PESO
Fríjol	10 Kg
Maíz	10 Kg
Jitomate	4 Kg
Carne de res (bistec)	2 Kg
Pollo (entero sin cabeza)	3 Kg
Zapatos	1/3 par *
Pantalón	¼ unidad *
Televisión color 21"	1/60 *
Refrigerador mediano	1/20*
Automóvil compacto	1/60*

Gasolina	100 litros

Nota: Los (*) son los bienes que llamamos "de consumo duradero" no se acaban en un mes, como seria el caso del maíz o del pollo. En este caso estaremos indicando que una televisión dura 5 años (60 meses) y que un refrigerador dura 10 años (120 meses), por lo que le asignamos a cada mes una parte proporcional de su valor.

El siguiente paso para construir nuestro índice será investigar en el almacén general el precio de cada uno de los artículos del índice que deseamos construir en la siguiente tabla. En la misma tabla aparecen los precios mensuales de los artículos; después, el cálculo del índice correspondiente.

Enseguida de la tabla aparece la explicación de cómo construimos el índice.

		MES BASE		**MES SIGUIENTE**	
ARTÍCULO	PESO Ó UNIDAD	PRECIO UNITARIO $	COSTO POR ARTÍCULO	PRECIO UNITARIO $	COSTO POR ARTÍCULO
Frijol	10 Kg	2.20	22.00	2.30	23.00
Maíz	10 Kg	16.00	160.00	16.00	160.00
Jitomate	4 Kg	8.00	32.00	8.50	34.00
Carne de res (bistec)	2 Kg	42.00	84.00	4.00	88.00
Pollo (entero sin cabeza)	3 Kg.	30.00	90.00	28.00	84.00
Zapatos	1/3 par *	300.00	100.00	300.00	100.00
Pantalón	¼ unidad *	200.00	50.00	210.00	52.50
Televisión color 21"	1/60 *	2,100.00	35.00	2,050.00	34.17
Refrigerador mediano	1/20*	3,000	25.00	3,000.00	25.00
Automóvil compacto	1/60*	40,000	666.67	40,000.00	666.67
Gasolina	100 litros	6.00	600.00	6.05	605.00
		Total	$1,864.67		$ 1,872.34

Nota: El (*) es el precio estimado del automóvil menos el valor de rescate, es decir, la pérdida de valor que sufre mientras su dueño lo utiliza.

El valor de nuestros artículos (con su peso correspondiente) en el mes de base fue de $1,864.67 y en el mes siguiente fue de $ 1,872.34 de acuerdo con la metodología de construcción del índice que ya vimos, los $ 1,864.67 conforman el 100% de nuestro periodo base. Entonces nuestro índice del mes siguiente es:

$$INPC = \frac{1872.34}{1864.67} \times 100 = 100.41$$

Podemos criticar razonadamente nuestro índice; el peso del medio de transporte (automóvil y gasolina) es de dos tercios del total, faltan muchos alimentos y el costo de la vivienda y la educación están completamente ausentes.

Con este sencillo ejemplo se pude percatar de lo complicado que es construir un índice de esta naturaleza. Esta complicación se ve positivamente incrementada si consideramos que existen diversas calidades (por ejemplo, pantalones de trabajo y de vestir o diversas calidades de carne o jitomate)[φ].

El conocimiento continuo de los principales **indicadores económicos** es de gran utilidad para las empresas o negocios que experimentan ciclos de prosperidad o de depresión económica, se ha demostrado en su uso como la principal herramienta de predicción económica de los gobiernos del mundo.

Sin embargo, debe tomarse el índice como una herramienta que proporciona información más con un carácter cualitativo que cuantitativo. Un análisis de los factores económicos subyacentes al utilizar el índice de indicadores económicos principales en conjunción con otros dispositivos de predicción, es de gran beneficio para el sector financiero ya que le proporciona una amplia panorámica de la actividad económica que puede ser interpretada y utilizada para la toma de decisiones.

[φ] Si el estudiante tiene interés en ver cómo se construye el índice nacional de precios al consumidor, le recomendamos consultar la página de internet del Banco de México, en la que viene una explicación muy completa al respecto.

Bibliografía del tema 6

LEVIN, Richard I. y David S. Rubin, *Estadística para administradores*, 6a. Edición, México, Prentice Hall, 1996.

MASON D., Robert, Douglas LIND A. y William MARCHAL G, *Estadística para administración y economía*, 11ª edición, Colombia, Alfaomega, 2004.

WEBSTER L., Allen, *Estadística aplicada a los negocios y la economía*, 3ª edición, McGraw-Hill,Colombia, 2002.

Actividades de aprendizaje

6.1 A partir del estudio de la bibliografía específica sugerida para este tema, presenta en un cuadro sinóptico cada uno de los conceptos principales y sus aplicaciones.

6.2 Formula y realiza por lo menos tres ejercicios o aplicaciones de números índices de acuerdo con los conceptos estudiados utilizando la bibliografía sugerida.

6.3 Investiga el método del promedio ponderado de precios relativos y su relación con el índice de precios agregados de Laspeyres. Redacta por lo menos tres ejemplos específicos.

6.4 Investiga el denominado “Índice ideal de Fisher”, sus características, fórmulas y su relación con los índices de Laspeyres y de Paasche. Realiza ejemplos de aplicación.

6.5 Examina la relación que existe entre el INPC y la determinación del índice de inflación en nuestro País. Expresa las fórmulas para el cálculo de la inflación oficial quincenal en una tabla.

6.6 Estudia los denominados índices de cantidad e índices de valor y presenta ejemplos de aplicación.

6.7 Investiga el cálculo del valor diario de las Unidades de Inversión (UDI’S) y su relación con el índice de inflación oficial.

6.8 Determina los principales problemas en la construcción de números índice y los errores más comunes en su interpretación.

6.9 Elabora un diagrama de flujo de números índice.

6.10 Visite la página www.aulafacil.com o la página www.solocursos.net/estadistica para una información complementaria.

Cuestionario de autoevaluación

1. ¿Cuáles son los principales elementos para construir un número índice?
2. ¿Cuál es la utilidad de trabajar con números índice?
3. Haga referencia de la utilidad de conocer y seguir el comportamiento del índice de una bolsa de valores.
4. ¿Cuáles son las características de un índice compuesto?
5. Indique la diferencia entre un índice de cantidad y un índice de valor.
6. Explique la diferencia que existe entre un índice agregado y un índice simple.
7. ¿A qué se refiere el nivel conceptual y el nivel operacional?
8. ¿Qué es un índice ponderado y su utilización?
9. ¿Cuál es la utilidad del índice de Laspeyres?
10. ¿En qué consiste el Índice Nacional de Precios al Consumidor y qué beneficios proporciona su conocimiento?

Examen de autoevaluación

1. Para la elaboración de un número índice se selecciona un número base, esto significa la elección de:
   a. La información del año anterior
   b. El mismo dato, pero en otro año y país
   c. La información que se usa entre un año y el siguiente
   d. La información con la que se desea realizar la comparación
   e. El dato que se muestra en el numerador

2. Es un ejemplo de índice simple:
   a. Precios de Laspeyres
   b. Índice de Paasche
   c. Valor y precio de las tiendas departamentales
   d. Índice de precios al consumidor
   e. Número de hombres por cada 100,000 habitantes

3. De acuerdo con el registro siguiente, el precio promedio de cierta maquinaria ha tenido las variaciones siguientes, considerando el año 2000 como base calcule el índice de precios correspondiente al año 2005:

Año	**Precio**
2000	$90,400
2001	$94,600
2002	$92,800
2005	$110,500

   a. 122.4
   b. 122.3
   c. 122.2
   d. 122.5
   e. 123.2

4. Una institución hipotecaria reporta que el monto de los créditos otorgados en los últimos cuatro años ha sido como se muestra en la tabla siguiente, considerando como base el monto otorgado en el año 1998, determine el valor del índice de créditos para el año 2000.

Año	**Monto del crédito**
1997	$65,300
1998	$69,800
1999	$69,500
2000	$84,200

a. 124.6
b. 128.9
c. 120.6
d. 106.9
e. 125.

5. Si en el 2000 el Producto Interno Bruto por persona era 150% con respecto a 1990 y la medición de precios al consumidor creció al doble en ese período, ¿cuál es porcentaje del ingreso neto con respecto a 1990?

a. 150
b. 100
c. 75
d. 125
e. 95

6. El índice de precios al consumidor (IPC) mide:
   a. Los precios de los productos en cierto mercado
   b. Los productos que más subieron de precio en cierta ciudad
   c. Los cambios de precios en los artículos producidos por cierta zona
   d. Los cambios en los precios de una canasta fija de productos
   e. El cambio de precios antes y después de un año determinado

7. Uno de los elementos que es considerado en el análisis y determinación de los incrementos de salario de las grandes organizaciones sindicales en México es:
   a. La deflación de las ventas
   b. El balance de importaciones y exportaciones
   c. El índice de valor del mercado
   d. El índice de Laspeyres
   e. El índice de precios al consumidor

8. El poder adquisitivo actual de la unidad monetaria es el resultado de la relación entre el valor 100 y el valor del índice :
   a. De valor
   b. Compuesto
   c. De precios
   d. Actual de precios al consumidor
   e. Ponderado

9. La determinación del cambio del poder adquisitivo del dinero es el resultado de comparar el salario en un año determinado con el del periodo base y dividirlo entre:
   a. El factor de deflación de las ventas
   b. El balance de las exportaciones y las importaciones
   c. El índice de valor del mercado
   d. El índice de Laspeyres
   e. El índice de precios al consumidor

10. Una de las ventajas de utilizar números índices es la de convertir cantidades reales a:
   a. Imaginarias
   b. Relativas
   c. Decimales
   d. Fraccionarias
   e. Dimensionales

Bibliografía básica

ANDERSON R., David, *Estadística para administración y economía*, 8ª edición, Thopmson, 2004.

BERENSON L., Mark, David Levine M. y Timothy Krehbiel C., *Estadística para administración*, 2ª edición, Prentice Hall, 2001.

LEVIN, Richard I. y David S. Rubin, *Estadística para administradores*, 6a. Edición, México, Prentice Hall, 1996.

MASON D., Robert, Douglas Lind A. y William Marchal G, *Estadística para administración y economía*, 11ª edición, Alfaomega, 2004.

WEBSTER L., Allen, *Estadística aplicada a los negocios y la economía*, 3ª edición, McGraw-Hill, 2002.

Bibliografía complementaria

ATO, Manuel y Juan J. López, *Fundamentos de estadística con SYSTAT*, México, Addison Wesley Iberoamericana, 1996.

CHRISTENSEN, H., *Estadística paso a paso*, 2ª edición, México, Trillas, 1990.

GARZA, Tomás, *Probabilidad y estadística*, México, Iberoamericana, 1996.

HANKE, Jonh E. y Arthur G. Reitsch, *Estadística para negocios*, México, McGraw–Hill, 1997.

HILDEBRAN y Lyman, *Estadística aplicada a la administración y a la economía*, Addison Wesley, México, 1998.

KAZMIER, L. y A. Díaz Mata, *Estadística aplicada a la administración y economía*, México, McGraw–Hill, 1998.

MENDENHALL, W. y R.L. Sheaffer, *Estadística matemática con aplicaciones*, México, Iberoamérica, 1986.

MEYER, Paul L., *Probabilidad y aplicaciones estadísticas*, México, Addison Wesley Iberoamericana, 2002.

SCHEAFFER, R. y W. Mendenhall, *Elementos de muestreo*, México, Iberoamericana, 1987.

WEIMER Richard E., *Estadística*, México, CECSA, 1996.

Respuestas a los exámenes de autoevaluación

Matemáticas III

Tema 1	**Tema 2**	**Tema 3**	**Tema 4**	**Tema 5**	**Tema 6**
1. d)	1. c)	1. d)	1. a)	1. a)	1. d)
2. a)	2. c)	2. d)	2. d)	2. b)	2. e)
3. b)	3. b)	3. a)	3. d)	3. b)	3. c)
4. a)	4. d)	4. d)	4. d)	4. a)	4. c)
5. e)	5. e)	5. a)	5. e)	5. a)	5. c)
6. c)	6. b)	6. d)	6. c)	6. e)	6. d)
7. e)	7. c)	7. d)	7. e)	7. c)	7. e)
8. d)	8. b)	8. b)	8. b)	8. a)	8. d)
9. e)	9. a)	9. b)	9. e)	9. c)	9. e)
10. c)	10. b)	10. c)	10. c)	10. c)	10. b)

Apéndice (tema 5)

Distribución normal estándar

Área bajo la curva

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.00000	0.00399	0.00798	0.01197	0.01595	0.01994	0.02392	0.02790	0.03188	0.03586
0.1	0.03983	0.04380	0.04776	0.05172	0.05567	0.05962	0.06356	0.06749	0.07142	0.07535
0.2	0.07926	0.08317	0.08706	0.09095	0.09483	0.09871	0.10257	0.10642	0.11026	0.11409
0.3	0.11791	0.12172	0.12552	0.12930	0.13307	0.13683	0.14058	0.14431	0.14803	0.15173
0.4	0.15542	0.15910	0.16276	0.16640	0.17003	0.17364	0.17724	0.18082	0.18439	0.18793
0.5	0.19146	0.19497	0.19847	0.20194	0.20540	0.20884	0.21226	0.21566	0.21904	0.22240
0.6	0.22575	0.22907	0.23237	0.23565	0.23891	0.24215	0.24537	0.24857	0.25175	0.25490
0.7	0.25804	0.26115	0.26424	0.26730	0.27035	0.27337	0.27637	0.27935	0.28230	0.28524
0.8	0.28814	0.29103	0.29389	0.29673	0.29955	0.30234	0.30511	0.30785	0.31057	0.31327
0.9	0.31594	0.31859	0.32121	0.32381	0.32639	0.32894	0.33147	0.33398	0.33646	0.33891
1.0	0.34134	0.34375	0.34614	0.34849	0.35083	0.35314	0.35543	0.35769	0.35993	0.36214
1.1	0.36433	0.36650	0.36864	0.37076	0.37286	0.37493	0.37698	0.37900	0.38100	0.38298
1.2	0.38493	0.38686	0.38877	0.39065	0.39251	0.39435	0.39617	0.39796	0.39973	0.40147
1.3	0.40320	0.40490	0.40658	0.40824	0.40988	0.41149	0.41308	0.41466	0.41621	0.41774
1.4	0.41924	0.42073	0.42220	0.42364	0.42507	0.42647	0.42785	0.42922	0.43056	0.43189
1.5	0.43319	0.43448	0.43574	0.43699	0.43822	0.43943	0.44062	0.44179	0.44295	0.44408
1.6	0.44520	0.44630	0.44738	0.44845	0.44950	0.45053	0.45154	0.45254	0.45352	0.45449
1.7	0.45543	0.45637	0.45728	0.45818	0.45907	0.45994	0.46080	0.46164	0.46246	0.46327
1.8	0.46407	0.46485	0.46562	0.46638	0.46712	0.46784	0.46856	0.46926	0.46995	0.47062
1.9	0.47128	0.47193	0.47257	0.47320	0.47381	0.47441	0.47500	0.47558	0.47615	0.47670
2.0	0.47725	0.47778	0.47831	0.47882	0.47932	0.47982	0.48030	0.48077	0.48124	0.48169
2.1	0.48214	0.48257	0.48300	0.48341	0.48382	0.48422	0.48461	0.48500	0.48537	0.48574
2.2	0.48610	0.48645	0.48679	0.48713	0.48745	0.48778	0.48809	0.48840	0.48870	0.48899
2.3	0.48928	0.48956	0.48983	0.49010	0.49036	0.49061	0.49086	0.49111	0.49134	0.49158
2.4	0.49180	0.49202	0.49224	0.49245	0.49266	0.49286	0.49305	0.49324	0.49343	0.49361
2.5	0.49379	0.49396	0.49413	0.49430	0.49446	0.49461	0.49477	0.49492	0.49506	0.49520
2.6	0.49534	0.49547	0.49560	0.49573	0.49585	0.49598	0.49609	0.49621	0.49632	0.49643
2.7	0.49653	0.49664	0.49674	0.49683	0.49693	0.49702	0.49711	0.49720	0.49728	0.49736
2.8	0.49744	0.49752	0.49760	0.49767	0.49774	0.49781	0.49788	0.49795	0.49801	0.49807
2.9	0.49813	0.49819	0.49825	0.49831	0.49836	0.49841	0.49846	0.49851	0.49856	0.49861
3.0	0.49865	0.49869	0.49874	0.49878	0.49882	0.49886	0.49889	0.49893	0.49896	0.49900
3.1	0.49903	0.49906	0.49910	0.49913	0.49916	0.49918	0.49921	0.49924	0.49926	0.49929
3.2	0.49931	0.49934	0.49936	0.49938	0.49940	0.49942	0.49944	0.49946	0.49948	0.49950
3.3	0.49952	0.49953	0.49955	0.49957	0.49958	0.49960	0.49961	0.49962	0.49964	0.49965
3.4	0.49966	0.49968	0.49969	0.49970	0.49971	0.49972	0.49973	0.49974	0.49975	0.49976
3.5	0.49977	0.49978	0.49978	0.49979	0.49980	0.49981	0.49981	0.49982	0.49983	0.49983
3.6	0.49984	0.49985	0.49985	0.49986	0.49986	0.49987	0.49987	0.49988	0.49988	0.49989
3.7	0.49989	0.49990	0.49990	0.49990	0.49991	0.49991	0.49992	0.49992	0.49992	0.49992
3.8	0.49993	0.49993	0.49993	0.49994	0.49994	0.49994	0.49994	0.49995	0.49995	0.49995
3.9	0.49995	0.49995	0.49996	0.49996	0.49996	0.49996	0.49996	0.49996	0.49997	0.49997
4.0	0.49997	0.49997	0.49997	0.49997	0.49997	0.49997	0.49998	0.49998	0.49998	0.49998
4.1	0.49998	0.49998	0.49998	0.49998	0.49998	0.49998	0.49998	0.49998	0.49999	0.49999
4.2	0.49999	0.49999	0.49999	0.49999	0.49999	0.49999	0.49999	0.49999	0.49999	0.49999