Universidad Nacional Autónoma de México
Facultad de Contaduría y Administración
Sistema Universidad Abierta y Educación a Distancia

Licenciatura en Informática

Matemáticas IV
(Estadística Descriptiva e Inferencial)

Apunte electrónico

SUAyED

COLABORADORES

DIRECTOR DE LA FCA
Dr. Juan Alberto Adam Siade

SECRETARIO GENERAL
L.C. y E.F. Leonel Sebastián Chavarría

– – – –

COORDINACIÓN GENERAL
Mtra. Gabriela Montero Montiel
Jefe de la División SUAyED-FCA-UNAM

COORDINACIÓN ACADÉMICA
Mtro. Francisco Hernández Mendoza
FCA-UNAM

– – – –

COAUTORES
Mtro. Antonio Camargo Martínez
Mtro. Jorge García Castro
Lic. Eliseo Flores Alamilla

ACTUALIZACIÓN
Mtro. Luis Fernando Zúñiga López
Mtra. Adriana Rodríguez Domínguez

DISEÑO INSTRUCCIONAL
Lic. María Cristina Rico León

CORRECCIÓN DE ESTILO
Mtro. Francisco Vladimir Aceves Gaytán

DISEÑO DE PORTADAS
L.CG. Ricardo Alberto Báez Caballero
Mtra. Marlene Olga Ramírez Chavero
L.DP. Ethel Alejandra Butrón Gutiérrez

DISEÑO EDITORIAL
Mtra. Marlene Olga Ramírez Chavero

OBJETIVO GENERAL

El alumno aplicará las herramientas estadísticas que le permitan sintetizar grandes volúmenes de información para presentar informes ejecutivos que describan el comportamiento de datos, derivados del análisis e interpretación y la aplicación de modelos estadísticos.

TEMARIO OFICIAL

(64 horas)

		Horas
1. Estadística descriptiva		8
2. Teoría de la probabilidad		12
3. Distribuciones de probabilidad		12
4. Distribuciones muestrales		8
5. Pruebas de hipótesis con la distribución ji cuadrada		8
6. Análisis de regresión lineal simple		8
7. Análisis de series de tiempo		8
	Total	**64**

INTRODUCCIÓN

En esta asignatura el estudiante estudiará lo relativo a la estadística descriptiva e inferencial.

En la **unidad 1** 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 la **unidad 2** se estudiarán 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 y probabilidad condicional y además se conocerá y aplicará el teorema de Bayes.

La **unidad 3** comprenderá 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 la distribución binomial, la distribución de Poisson, la distribución hipergeométrica, la distribución multinomial, la distribución normal y la distribución exponencial. Por último, se enunciará la ley de los grandes números y su interpretación.

En la **unidad 4** se estudiarán las distribuciones muestrales y el teorema central del límite, los cuales pueden ayudar a la posterior elaboración de los intervalos de confianza.

En la **unidad 5** analizaremos las pruebas de hipótesis con la distribución ji cuadrada y su aplicación.

En la **unidad 6** se investigará el análisis de regresión lineal simple para averiguar el comportamiento de las variables y sus diferentes relaciones.

En la **unidad 7** analizaremos las series de tiempo para observar su aplicación a diferentes problemas de la vida diaria de las empresas.

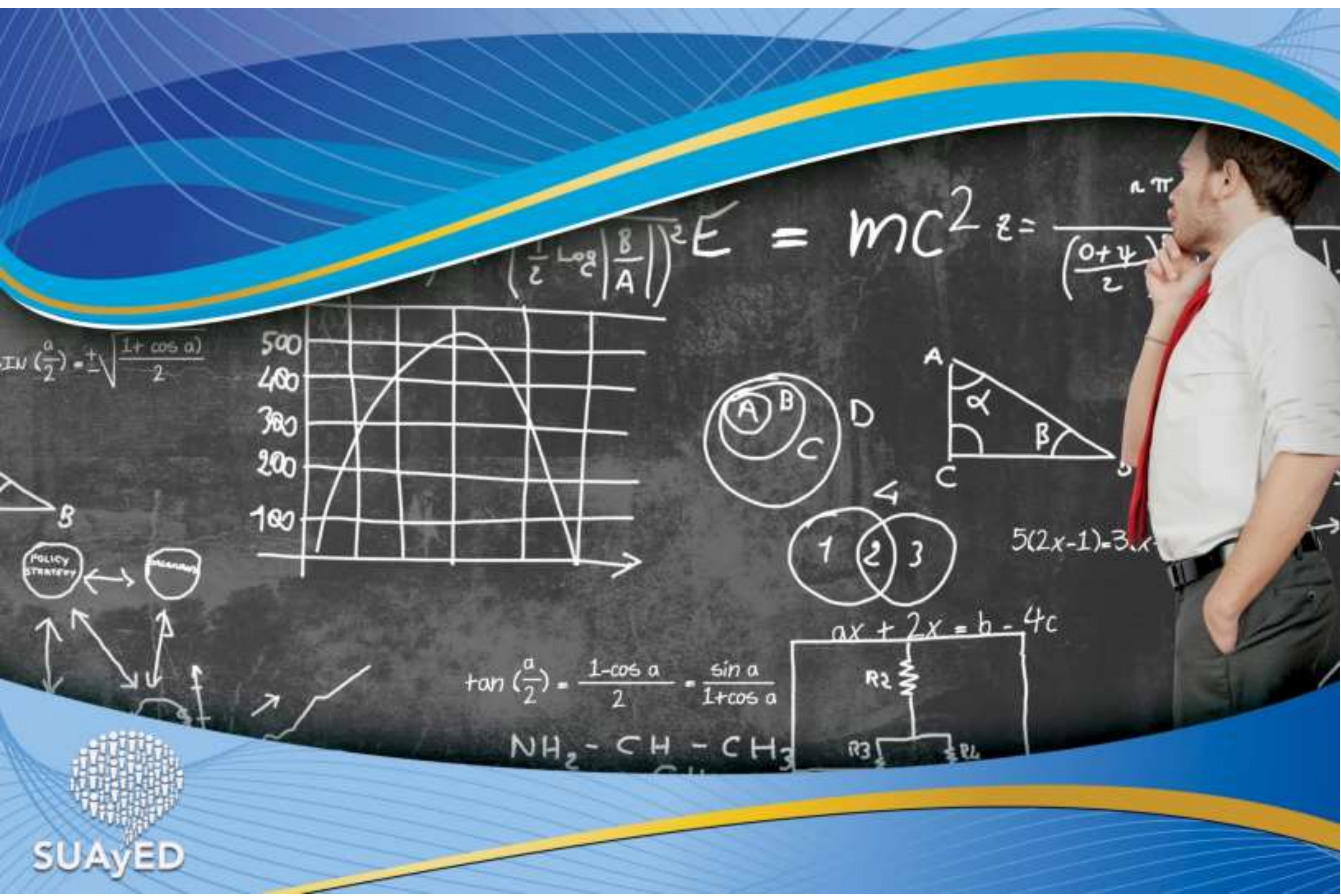
La estadística descriptiva e inferencial es un elemento imprescindible en la toma de decisiones, tanto en el nivel de las organizaciones privadas y gubernamentales como en el individual. En particular, los estudiantes de informática encontrarán campo fértil para aplicar métodos estadísticos en las áreas de programación y desarrollo de sistemas, entre muchas otras.

La estadística es una rama de las matemáticas, por lo que su tratamiento es formal. Esto no significa, sin embargo, que en el curso se requiera realizar demostraciones rigurosas. El enfoque que se ha adoptado es más bien pragmático, por cuanto está orientado a la aplicación de conceptos, de modo que el requisito fundamental es contar con conocimientos básicos de álgebra y manejo de hoja de cálculo.

ESTRUCTURA CONCEPTUAL

Unidad 1.

Estadística descriptiva



OBJETIVO PARTICULAR

El alumno aprenderá y aplicará el proceso estadístico para transformar datos en información útil para la toma de decisiones.

TEMARIO DETALLADO

(8 horas)

1. Estadística descriptiva
1.1. Tabulación de datos
1.2. Distribuciones de frecuencia
1.3. Presentación gráfica de datos
1.4. Medidas de tendencia central
1.5. Medidas de dispersión
1.6. Teorema de Tchebysheff y regla empírica

INTRODUCCIÓN

Para que la información estadística sea relevante, útil y confiable es necesario prestar atención a todas las etapas del proceso de manejo de los datos. Desde el punto de vista de la Estadística Descriptiva es importante, entonces, atender a los diferentes tipos de escalas con que pueden medirse los atributos o variables que nos interesan de un conjunto de observaciones y la forma de agrupar los datos correctamente para, a partir de aquí, aplicar los métodos estadísticos de representación gráfica, así como determinar las medidas de localización y de dispersión que nos permiten dar pasos firmes al interior de la estructura de los datos. 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. 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, 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.

Técnicas de resumen
Nos indican la mejor manera para ordenar y agrupar la información, de forma tal que ésta tenga mayor 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.

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 *pie* o “pay”, por su pronunciación en inglés).

Técnicas de obtención de parámetros

Nos llevan a calcular indicadores numéricos 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 inclinaremos a pensar que es un buen estudiante. Los parámetros son números que nos sirven para representar (bosquejar una idea) de las principales características de las poblaciones.

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, esto es, variará. Por ello decimos que la estatura es una **variable o atributo.**

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. Podemos trabajar con cualidades de las observaciones, como por ejemplo el estado civil de una persona, o con características cuantificables, como por ejemplo la edad.

No todos los atributos se miden igual, lo que da lugar a tener diferentes escalas de medición.

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 escoge 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 exhaustivo; es decir, tiene que contener a todos los casos posibles. Además, las categorías a que pertenecen los datos no cuentan con un orden lógico.

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, en donde las preguntas incluyen 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 categorías. No podemos decir, por ejemplo, que "frecuentemente" es el doble que "algunas veces" o que "nunca" es tres puntos más bajo 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).

Escalas numéricas

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

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.

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, puede ser que en el equipo los candidatos a jugador se admitan a partir de determinada estatura para arriba, en tanto que de esa estatura para abajo no serían admitidos. En este caso, la variable estatura tendría solo dos valores, a saber, “aceptado” y “no aceptado” y sería 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, una estatura de 173 cm., de modo que si mide unos milímetros más o menos se redondeará al centímetro más cercano; el registro llevaría a una variable continua si el investigador anota 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.**

Escalas de tipo numérico	
Escalas de intervalo	***Escalas de razón***
Son aquellas en las que el **cero es convencional o arbitrario.**	Son aquellas en las que el **cero absoluto sí existe.**
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	Tal es el caso de los grados Kelvin, para medir temperaturas, o algunas otras medidas que utilizamos en nuestra vida cotidiana. Encontramos un ejemplo de esta escala cuando medimos la estatura

temperaturas más frías que se miden mediante números negativos. En esta escala se pueden 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 porcentajes ya que no hay lugar a dividir en las escalas de intervalo. Si la temperatura ambiente el día de hoy es de diez grados, y el día de ayer fue de veinte grados, no podemos decir que hoy hace el doble de frío que ayer. Sólo podríamos decir que hoy hace más frío y que la temperatura es 10 grados menor que ayer.	de las personas, expresada en centímetros, por ejemplo, ya que sí existe el cero absoluto, además de que sí se pueden formar cocientes que nos permiten afirmar que alguien mide el doble.

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.

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 si bien no son difíciles desde el punto de vista conceptual, 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.

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.

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

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

Título

Encabezado

Cuerpo de la tabla

Fuente de información

Estudiantes de la FCA que trabajan

Porcentajes por semestre de estudio*

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

En él se encuentran los datos propiamente dichos

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.

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

Editorial Académica, México, 19XX

Tabla sencilla de datos

Independientemente de los elementos que pudieren tener las tablas, existen diversas maneras de presentar la información en ellas. No existe una clasificación absoluta de presentación de las diferentes tablas, dado que se pueden inventar varias maneras de presentar la información estadística. Empero, se puede intentar una clasificación que nos permita entender las principales presentaciones.

Tablas simples

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

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

Tablas de frecuencias

Es un arreglo rectangular de información en el que las columnas representan diversos conceptos, dependiendo de las intenciones de quien 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 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), por lo que algunos autores llaman a la frecuencia relativa "frecuencia porcentual".

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

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 simultáneamente por medio de dos criterios en lugar de utilizar solamente uno. Las columnas están relacionadas con un criterio y los renglones con el otro criterio.

Deportes Batista, S.A. de C.V.					
Bicicletas vendidas por modelo y tienda **Primer trimestre de 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**

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 en la tabla de frecuencias no teníamos.

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 por el turno (matutino, vespertino o nocturno) en el que se produjo el defecto. Lo que se desea estudiar es si la evidencia de los datos (la contingencia y de allí el nombre) apoya la hipótesis 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 la que 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									

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.

1.2. Distribuciones de frecuencia

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

Distribución de frecuencias para variables cuantitativas

Las variables cuantitativas o métricas pueden ser de dos tipos.

Continua

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

Discretas

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 requiere sólo del conteo del número de elementos o individuos que se encuentran dentro de cierta cualidad o, bien, dentro de determinada característica.

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**.**

Cuadros de trabajo

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.

Cuadros de referencia

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.

Cuadro 1.1

2. **Título.** 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.

Cuadro 1.1 Distribución de alumnos por días de ausencia

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.

Cuadro 1.1 Distribución de alumnos por días de ausencia

Mes base enero

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

Cuadro 1.1 Distribución de alumnos por días de ausencia

Mes base enero

En algunos casos se especifica el nombre del atributo

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

Cuadro 1.1 Distribución de alumnos por días de ausencia

Mes base enero

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

Cuadro 1.1 Distribución de alumnos por días de ausencia

Mes base enero

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

Cuadro 1.1 Distribución de alumnos por días de ausencia

Mes base enero

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

Cuadro 1.1 Distribución de alumnos por días de ausencia

Mes base enero

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

Cuadro 1.1 Distribución de alumnos por días de ausencia

Mes base enero

Ausencia **(valores de variable)**	**Número de alumnos** **(Frecuencia)**

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.

Cuadro 1.1 Distribución de alumnos por días de ausencia

Mes base enero

Ausencia (valores de variable)	**Número de alumnos (Frecuencia)**
	CASILLA

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

Cuadro 1.1 Distribución de alumnos por días de ausencia

Mes base enero

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

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

Cuadro 1.1 Distribución de alumnos por días de ausencia

Mes base enero

Ausencia (valores de variable)	Número de alumnos (Frecuencia)
0	11 4
1	4
2	2
3	2
4	1
5	1
Total	21

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.

Cuadro 1.1 Distribución de alumnos por días de ausencia

Mes base enero

Ausencia (valores de variable)	**Número de alumnos (Frecuencia)**
0	11 4
1	4
2	2
3	2
4	1
5	1
Total	21

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.

Cuadro 1.1 Distribución de alumnos por días de ausencia

Mes base enero

Ausencia (valores de variable)	**Número de alumnos (Frecuencia)**
0	11 4
1	4
2	2
3	2
4	1
5	1
Total	21

Nota: No se tiene registrado ningún caso con más de 5 ausencias.

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

Cuadro 1.1 Distribución de alumnos por días de ausencia

Mes base enero

Ausencia (valores de variable)	**Número de alumnos (Frecuencia)**
0	11 4
1	4
2	2
3	2
4	1
5	1
Total	21

Nota: No se tiene registrado ningún caso con más de 5 ausencias.

Fuente: Informe mensual de actividades. Mes enero 2007

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

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.

1.3 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, etc. 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.

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 continúa. 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, 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.

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

Considera el siguiente conjunto de datos:

8.98.39.28.49.1				
8.68.99.18.88.8				
8.89.18.98.78.8				
8.99.08.68.78.4				
8.69.08.88.99.1				
9.49.09.29.18.8				
9.19.39.09.28.8				
9.78.99.78.39.3				
8.9	8.8	9.3	8.5	8.9
8.39.28.28.98.7				
8.98.88.58.48.0				
8.58.78.78.88.8				
8.38.68.79.08.7				

	8.48.88.48.69.0	
	9.38.88.58.79.6	
	8.59.19.08.89.1	
	8.68.68.49.18.5	
	9.19.28.88.58.3	
	9.38.68.78.79.1	
	8.88.79.09.08.5	
	8.58.88.98.29.0	
	9.08.78.78.99.4	
	8.38.69.28.78.7	
	8.79.78.99.28.8	
	8.38.68.58.69.7	
Máximo	**9.79.79.79.29.2**	**máximo = 9.7**
Mínimo	**8.38.38.28.28.0**	**mínimo = 8.0**

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

Paso 2. Calcula 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 tanto el valor más grande como el más pequeño por columna. Se anotan los resultados en dos renglones, uno para los valores máximos y otro para los mínimos y de entre estos números se determina nuevamente el mayor y el menor, mismos que serán identificados como el *máximo y mínimo* de las lecturas en la tabla. En este caso: MÁX = 9.7 y MÍN = 8.0. El rango (R) es la diferencia entre éstos valores, por lo que R = MÁX – MÍN = 9.7 – 8.0 = 1.7.

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

En la construcción de un diagrama de frecuencias o de un histograma es necesario encasillar las lecturas. Si bien existe una expresión matemática para el cálculo del número de clases que debe tener la distribución de frecuencias, hay un camino más práctico, el cual señala que el número de clases no debe ser menor que 6 ni mayor que 15. En este sentido, si “Q” es 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**

Paso 4. Determina el ancho “c” del 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 números más cómodos. En esta ocasión daremos un valor de c=0.20 unidades el cual debe mantenerse constante a lo largo del rango, que en este caso es de R=1.7

Paso 5. Establece 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 instancias, la resolución en los instrumentos de medición es la que determina el ancho de clase, aun cuando se siga una regla general para la normalización del histograma. En el ejemplo, la lectura menor fue de 8.0 por lo que se podría tomar éste como el límite inferior de la primera clase, y al sumar al valor de 8.0 el ancho de clase “c” se tendría el límite inferior del segundo intervalo y así sucesivamente hasta que todos los valores de la tabla queden contenidos.

Paso 6. Construye la distribución de frecuencias:

Clase	Límite de clase	Marca de clase	*Frecuencia*	Total
1	8.00-819	8.1		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

Tabla de distribución de frecuencias

Al graficar los datos anteriores obtenemos la siguiente figura:

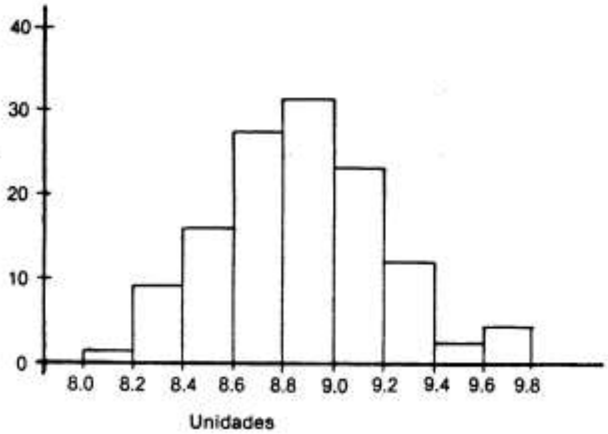


Figura 1-3

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 características no cuantitativas 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**

Gráficos para distribuciones de frecuencias no agrupadas

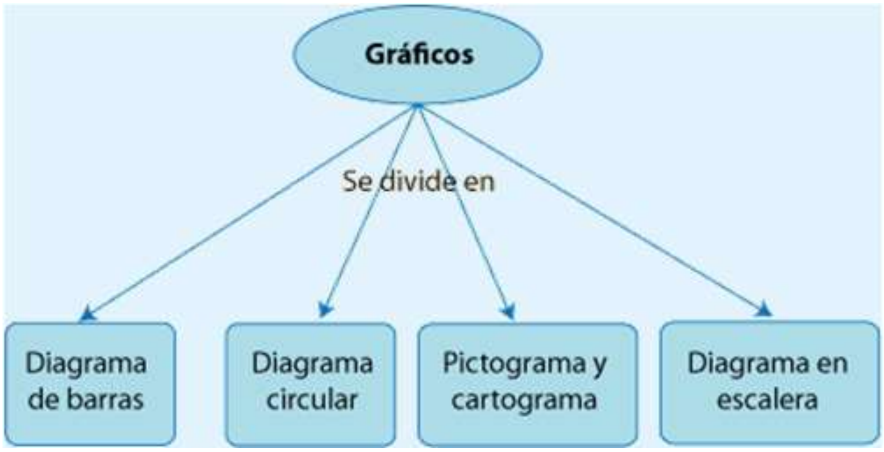
Para representar este tipo de distribuciones, los gráficos más utilizados son:

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; representa los datos mediante una barra o columna, que puede colocarse horizontal o verticalmente. 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.

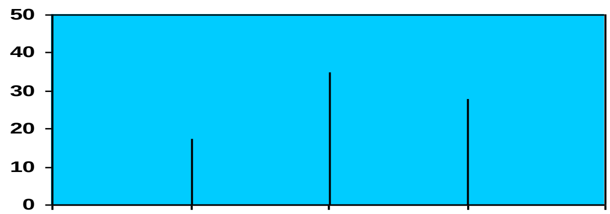


Diagrama de barras

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

Una variante de este diagrama, más utilizada quizá por ser más ilustrativa, es el *diagrama de rectángulos*, que representa 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.

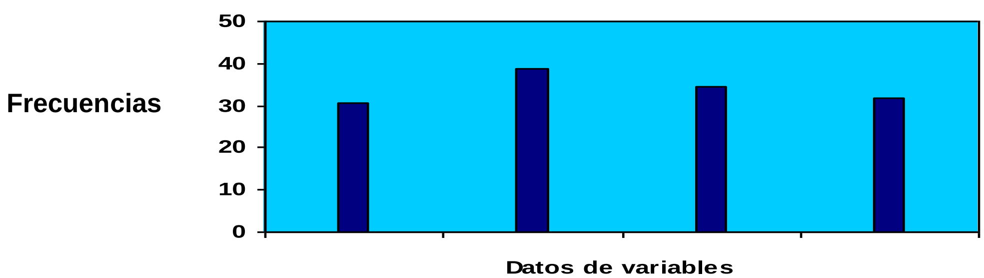


Diagrama de rectángulos

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

Además, el diagrama de rectángulos es especialmente útil cuando se desea comparar, en un mismo gráfico, el comportamiento del fenómeno en dos o más situaciones o ámbitos distintos, para lo cual podemos usar colores, uno por ámbito, y con ello obtener una visión simplificada y conjunta de lo que ocurre en ambos casos por tratar. Ejemplo de análisis comparativo que puede ser representado con rectángulos en dos tonos.

b) Diagrama circular

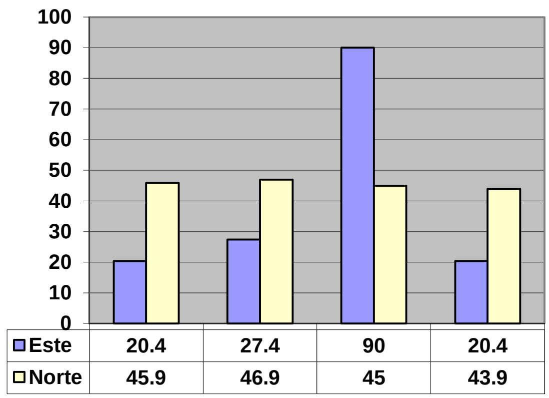


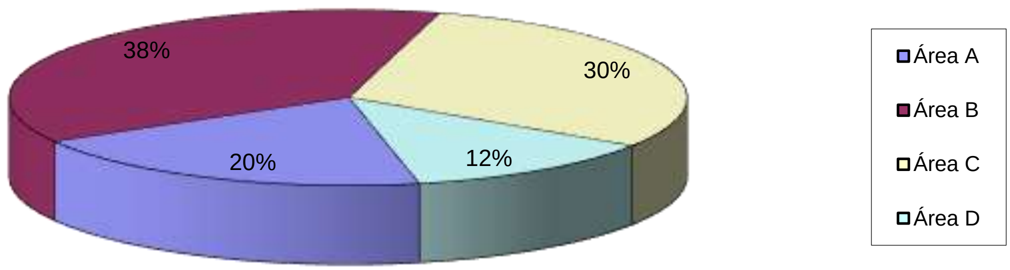
Diagrama de rectángulos

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 traza un 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. En otras palabras, como el círculo tiene 360º, éstos se reparten proporcionalmente a las frecuencias absolutas de cada componente.



Gráfica circular o pastel

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.

c1) 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:

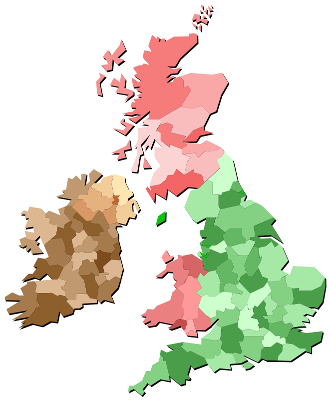
 = 100 viviendas

Y para representar 300 viviendas

 = 300 viviendas

c2) 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.



Fuente: Revista *Expansión*, núm. 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.

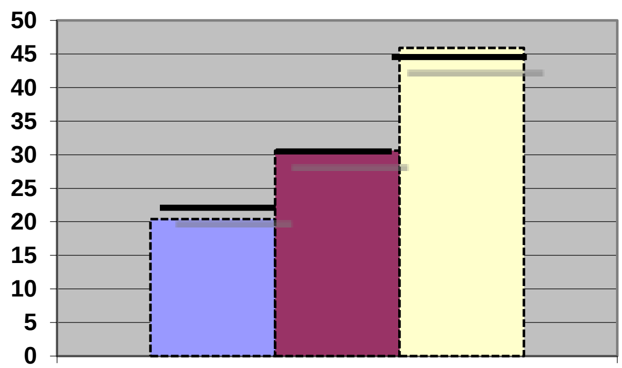


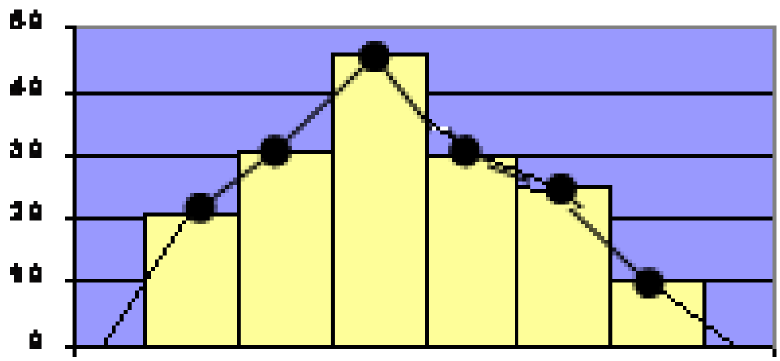
Diagrama en escalera

Gráficas para distribuciones de frecuencias agrupadas en clases

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.



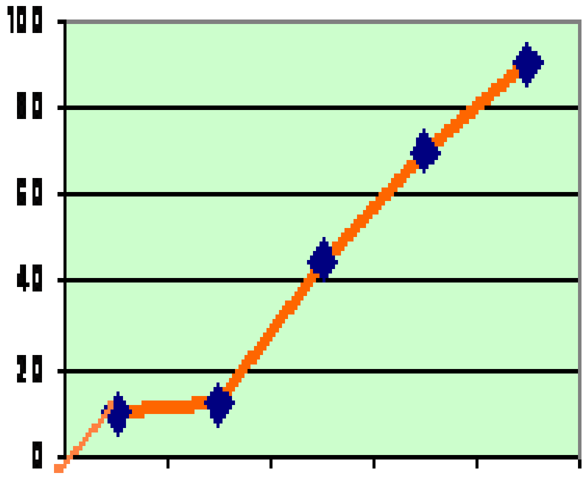
Polígono de frecuencias

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 se acumulan hasta diferentes valores de la variable, esto es, cuántas hay en la zona 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.

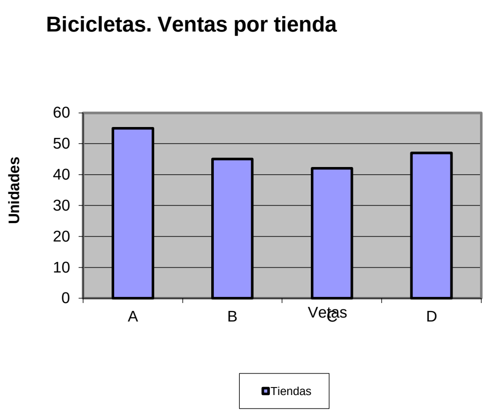


Ojivas Fuente: Revista *Expansión*, núm. 852, (octubre 30 del 2002), p. 14.

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.



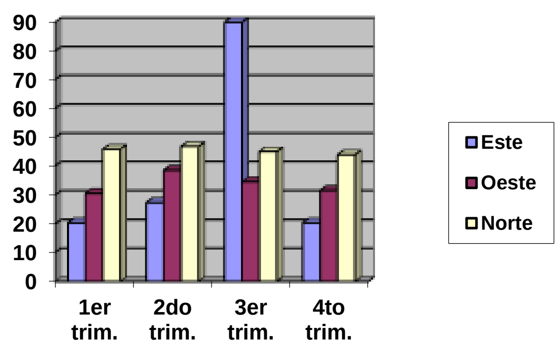


Diagrama de columnas

Barras

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

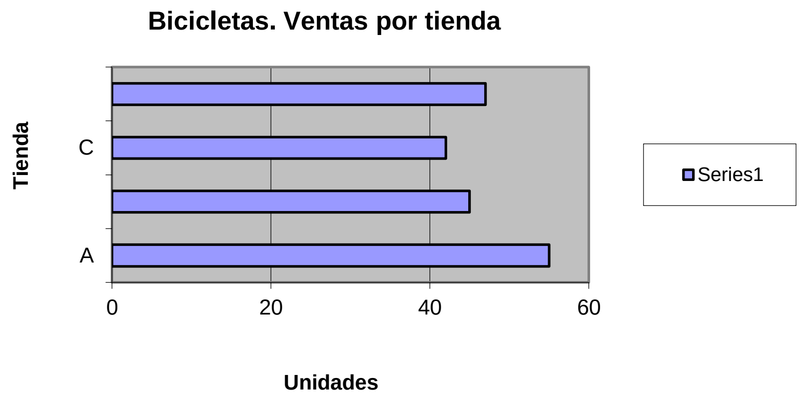


Diagrama de barras

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.

Diagrama circular

1.4 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 nuestros datos de diversas maneras para obtener información. 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**.

Medidas de posición

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

Las **medidas de tendencia central** son las que nos indican datos representativos de una distribución y que tienden a ubicarse en el centro de la misma.

A su vez, las medidas de posición tienen el objetivo de localizar diversos puntos de interés ubicados en diversas partes de la distribución; por ejemplo, el punto que divide la distribución en dos partes: a la izquierda (datos más pequeños), 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 dividiendo el total entre el número de datos. Podemos decir entonces que la media aritmética determina cómo repartir un total entre N observaciones si el reparto es a partes iguales.

La manera formal de expresar este concepto es la siguiente:

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

Esta expresión 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.

Considera 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 y se le conoce como Md. 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.

Si en lugar de un número impar 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.

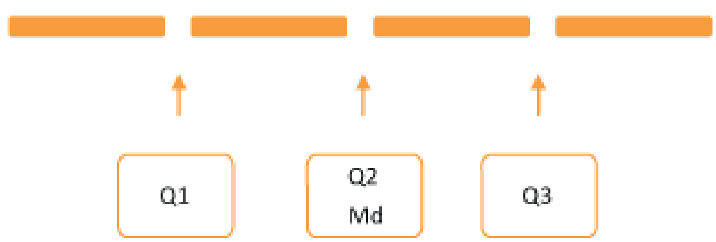
6, 7, 7, **8, 8, 8, 8**, 9, 9, 10 ,10.

En el caso de las ventas de la tienda, se puede ver que no 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**:



Posición de cuartiles

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

Hay otras dos medidas de posición que se asemejan al concepto de cuartiles. Se trata de los “**deciles**” y los “**percentiles**”, sólo que éstas son medidas que en lugar de separar los datos en grupos de 25% lo hacen en grupos de 10% y de 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.

1.5 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; $10,500; $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:

$$R = X_M - X_m$$

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 considera los datos que están en medio**. Para tomar en cuenta todos los datos se inventaron las medidas de dispersión que son la varianza y la desviación estándar.

Varianza y desviación estándar

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, qué 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**
5,000	5,000 – 10,000 = –5,000	5,000	5,000 – 10,000 = –5,000
10,000	10,000 – 10,000 = 0	6,000	6,000 – 10,000 = –4,000
10,000	10,000 – 10,000 = 0	10,000	10,000 – 10,000 = 0
10,000	10,000 – 10,000 = 0	14,000	14,000 – 10,000 = 4,000
15,000	15,000 – 10,000 = 5,000	15,000	15,000 – 10,000 = 5,000
	Suma = 0		**Suma = 0**

Tabla de desviaciones de datos

Como se puede apreciar la suma de las diferencias entre la media y cada dato tiene como resultado el valor cero, 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**
5,000	5,000	25,000,000	5,000	–5,000	25,000,000
10,000	0	0	6,000	–4,000	16,000,000
10,000	0	0	10,000	0	0
10,000	0	0	14,000	4,000	16,000,000
15,000	5,000	25,000,000	15,000	5,000	25,000,000
SUMA	**0**	**50,000,000**	**SUMA**	**0**	**82,000,000**

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 número total de datos lo conocemos como varianza (o variancia, según el libro que se consulte).

De acuerdo con lo anterior, 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). En resumen:

La varianza es la medida de dispersión que corresponde al promedio aritmético de las desviaciones cuadráticas de cada valor de la variable, con respecto a la media de los datos.

La expresión algebraica que corresponde a este concepto 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 también se elevan al cuadrado y no es nada fácil captar lo que significan, por ejemplo, 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.

La raíz cuadrada de la varianza recibe el nombre de **desviación estándar o desviación típica**.

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 elevada 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 (5,000 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 (como ejercicio, el alumno puede comprobar las medias y las desviaciones estándar calculándolas él mismo):

Peso de los costales	Peso de los camiones
40 Kg	4,990 Kg
50 Kg	5,000 Kg
60 Kg	5,010 Kg

Tabla de datos

- Media de los costales 50 kg.
- Media de los camiones 5,000 kg.
- Desviación estándar de los costales 8.165 kg.
- Desviación estándar de los camiones 8.165 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:

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

donde:

$C.V.$ coeficiente de variación.

σ desviación estándar.

μ media de la población.

En el caso de los costales tendíamos que C.V.= (8.165/50)100=16.33, lo que nos indica que la desviación estándar del peso de los costales es de 16.33% del peso medio (una desviación significativamente grande).

Por otra parte, en el caso de los camiones, el coeficiente de variación nos arroja:

C.V.= (8.165/5000)100= 0.1633, lo que 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**", ya que así es más cómodo manipularlos aunque se pierde alguna precisión.

Imagine que se tienen 400 datos y el trabajo que representaría ordenarlos uno por uno para obtener la mediana. Por ello se han desarrollado técnicas que permiten 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.

Hay varios criterios para establecer el número de clases. Entre ellos, que el número de clases es aproximadamente…

- la raíz cuadrada del número de datos.

- el logaritmo del número de datos entre el logaritmo de 2.

Normalmente se afirma que las clases no deben ser menores que cinco ni mayores que veinte. De cualquier manera, el responsable de trabajar con los datos puede utilizar su criterio.

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

Tabla de Número de clases según número de datos

Supongamos que tenemos 44 datos —como en el caso de la tabla que se presenta a continuación—, que corresponden a las ventas diarias de una pequeña miscelánea. Si seguimos el criterio de los logaritmos, el número de clases será: logaritmo de 44 entre logaritmo de 2, esto es, 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	12	532	23	763	34	603
2	918	13	628	24	829	35	890
3	911	14	935	25	671	36	772
4	639	15	606	26	965	37	951
5	615	16	680	27	816	38	667

6	906	17	993	28	525	39	897
7	638	18	693	29	846	40	742
8	955	19	586	30	773	41	1000
9	549	20	508	31	547	42	800
10	603	21	885	32	624	43	747
11	767	22	590	33	524	44	500

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 cuál de ellas clasificarlos. Si un dato es exactamente 700, no sabríamos si 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 la siguiente manera:

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

Tabla de clases

Como vemos, los intervalos de cada clase están cerrados por la izquierda y abiertos por la derecha. 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

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 esto no tendrá impacto en nuestro trabajo, o bien podemos ajustar 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

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. En nuestro ejemplo se tendría:

Clase	Incluye datos iguales o mayores a:	Incluye datos menores a:	Marca de clase (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**

Marcas de clase

Éstas serían las marcas si las clases se construyen como en la primera tabla de clases.

Si se aplica el criterio de la tercera tabla, las marcas quedarían como sigue:

Clase	Incluye datos iguales o mayores a:	Incluye datos menores a:	Marca e clase (Xi)
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

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 clasificarlos. Para esto usaremos una de las clasificaciones ya especificadas:

Clase	Incluye datos mayores a:	Incluye datos menores o iguales a:	Conteo de casos	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**

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:

$$\overline{x} = \frac{\sum_{x=1}^{N} x_i f_i}{n}$$

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:

$$Mo = L_{Mo} + \frac{d_1}{d_1 + d_2} \cdot i \quad \begin{matrix} d_1 = f_{Mo} - f_1 \\ d_2 = f_{Mo} - f_2 \end{matrix}$$

En donde:

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

d_1 es la diferencia entre la frecuencia de clase $\left(f_{Mo} \right)$ del intervalo que contiene a la moda y la frecuencia de la clase inmediata anterior $\left(f_1 \right)$.

d_2 es la diferencia entre la frecuencia de clase $\left(f_{Mo} \right)$ del intervalo que contiene a la moda y la frecuencia de la clase inmediata posterior $\left(f_2 \right)$.

Medidas de dispersión

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

b) Varianza:

$$\sigma^2 = \frac{\sum \left(x_i - \overline{x} \right)^2 f_i}{n}$$

En donde:

x_i es la marca de clase.

f_i es la frecuencia de clase.

$\overline{x}$ es la media.

n es el número de datos.

c) Desviación estándar:

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

d) Coeficiente de variación:

$$C.V. = \frac{\sigma}{x}$$

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 sí 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	$f_i acum$	$x_i\, f_i$	$\left(x_i - \overline{x}\right)^2\ f_i$
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
SUMA		110		8,390.0	14,109.32

Medidas de tendencia central

a) Media:

$$\overline{x} = \frac{\sum_{x=1}^{N} x_i f_i}{n} = \frac{8,390.0}{110} = 76.27$$

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

b) Mediana:

$$Md = L_M + \frac{n/2 - F_M}{f_M} \cdot i = 75 + \frac{55 - 47}{18} \cdot 5 = 77.22$$

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:

$$Mo = L_{Mo} + \frac{d_1}{d_1 + d_2} \cdot i \quad \begin{array}{l} d_1 = f_{Mo} - f_1 \\ d_2 = f_{Mo} - f_2 \end{array}$$

$Mo = 80 + \frac{2}{2+7} \cdot 5 = 80.11$,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:

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

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.

1.6 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. Ambas herramientas se utilizan principalmente para estimar el número aproximado de datos que se encuentran en determinadas áreas de la distribución de datos.

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.

Ejemplo 1. 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 2 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, el cociente (58-70)/5=2.4 indica que 50 está a 2.4 desviaciones estándar abajo del promedio, en tanto 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 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, en cuyo caso se dice 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.

La regla empírica

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

- aproximadamente 68% de los elementos están a menos de una desviación estándar de la media.
- aproximadamente 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 2: 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 68% de los envases llenos tienen entre 15.75 y 16.25 onzas (esto es, a menos de una desviación estándar del promedio).
- aproximadamente 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.

RESUMEN

La estadística descriptiva es una herramienta matemática que conjuga una serie de indicadores numéricos y gráficos, así como los procedimientos con que éstos se construyen, para descubrir y describir, en forma abreviada y a través de símbolos precisos, la estructura inmersa en el conjunto de datos. Se dice que se conoce la estructura cuando se sabe:

a) Lo que ocurre en ciertos puntos específicos de la distribución de los datos.

b) En qué medida los valores de las observaciones difieren.

c) La forma general de la distribución de los datos.

La confiabilidad y relevancia de los indicadores depende en buena medida de una adecuada definición del objeto bajo estudio y de la medición correcta de sus atributos. De hecho, se puede decir que según la manera en que se midan los atributos dependerá el tipo de indicador que se puede construir.

BIBLIOGRAFÍA

SUGERIDA

Autor	Capítulo	Páginas
Berenson, Levine, Krehbiel (2001)	1. Introducción y recopilación de datos. Secciones: 1.7 Tipos de datos.	9-11
	2.1 Organización de datos numéricos.	40-45
	2. Presentación de datos en tablas y gráficas. Secciones: 2.2 Tablas y gráficas para datos numéricos.	45-57
	2.3 Tablas y gráficas para datos categóricos.	57-65
	2.4 Tablas y gráficas para datos bivariados.	65-70
	2.5 Excelencia gráfica.	70-78
	3. Resumen y descripción de datos numéricos. Secciones: 3.1 Exploración de datos numéricos y sus propiedades.	102-103
	3.2 Medidas de tendencia central, variación y	103-127

	forma.	
	3.4 Obtención de medidas descriptivas de resumen a partir de una población.	133-139
Levin y Rubin (2004)	2. Agrupación y presentación de datos para expresar significados: tablas y gráficas. Secciones: 2.1 ¿Cómo podemos ordenar los datos?	8-11
	2.3 Ordenamiento de datos en arreglos de datos y distribuciones de frecuencias.	12-20
	2.4 Construcción de una distribución de frecuencias.	20-9
	2.5 Representación gráfica de distribuciones de frecuencias.	29-41
	3. Medidas de tendencia central y dispersión en distribuciones de frecuencias. Secciones: 3.2 Representación gráfica de distribuciones de frecuencias.	29-41
	3.5 Una cuarta medida de tendencia central: la mediana.	77-83
	3.6 Una medida final de tendencia central: la moda.	84-89
	3.7 Dispersión: ¿por qué es importante?	89-91
	3.8 Rangos: medidas de dispersión útiles.	91-95
	3.9 Dispersión: medidas de dispersión promedio.	96-107

	3.10 Dispersión relativa: el coeficiente de variación.	107-112
Lind, Marchal, Wathen (2008)	1. ¿Qué es la estadística? Sección: Tipos de variables.	8-9
	Niveles de medición.	9-13
	2. Descripción de datos: tablas de frecuencias, distribuciones de frecuencias y su representación. Secciones: Construcción de una tabla de frecuencias.	22-27
	Construcción de distribuciones de frecuencias: datos cuantitativos.	28-32
	Representación gráfica de una distribución de frecuencias.	36-39
	3. Descripción de datos: medidas numéricas Secciones: La media poblacional.	57-58
	Media de una muestra.	58-59
	Propiedades de la media aritmética.	59-61
	Mediana.	62-64
	Moda.	64-65
	Posiciones relativas de la media, la mediana y la moda.	67-68
	¿Por qué estudiar la dispersión?	71-73
	Medidas de dispersión.	73-80

	Interpretación y usos de la desviación estándar.	81-83
	La media y la desviación estándar de datos agrupados.	84-87

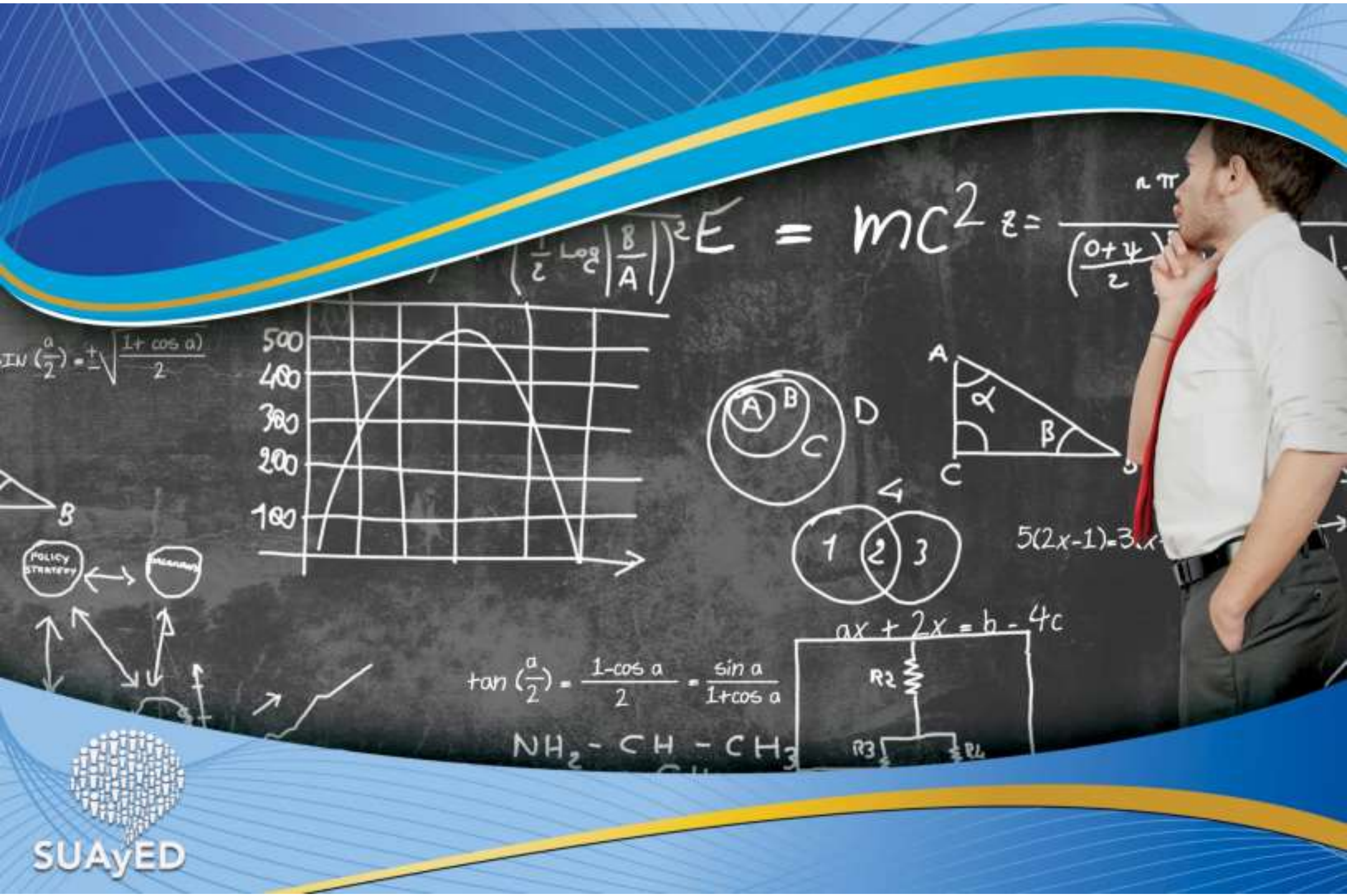
Berenson, Mark L., David M. Levine, y Timothy C Krehbiel. (2001), *Estadística para administración.* 2ª edición, México: Prentice Hall, 734 pp.

Levin, Richard I. y David S. Rubin. (2004), *Estadística para administración y economía*. 7ª edición, México: Pearson Educación Prentice Hall, 826 pp.

Lind, Douglas A., Marchal, William G., Wathen, Samuel, A. (2008), *Estadística aplicada a los negocios y la economía*. 13ª edición, México: McGraw-Hill Interamericana, 859 pp.

Unidad 2.

Teoría de la probabilidad



OBJETIVO PARTICULAR

El alumno identificará los diferentes enfoques de probabilidad y su interpretación para la toma de decisiones.

TEMARIO DETALLADO

(12 horas)

2. Teoría de la probabilidad

2.1. Interpretaciones de la probabilidad
2.1.1. Teórica o clásica
2.1.2. La probabilidad como frecuencia relativa
2.1.3. Interpretación subjetiva de la probabilidad
2.2. Espacio muestral y eventos
2.3. Los axiomas de la probabilidad
2.4. La regla de la suma de probabilidades
2.5. Tablas de contingencias y probabilidad condicional
2.6. Independencia estadística
2.7. La regla de multiplicación de probabilidades
2.8. Teorema de Bayes

INTRODUCCIÓN

Dicen que solamente existen dos cosas en la vida que con toda seguridad habremos de enfrentar: 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 concentraremos en lo que se conoce como probabilidad básica.

En ella no existen muchas fórmulas a las cuales recurrir, aunque sí existen desde luego algunas expresiones algebraicas. 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, advertir sobre la presentación de datos. Al ser la probabilidad un número entre cero y uno *es frecuente expresarla en porcentaje.* A la mayoría 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 de 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.

2.1. Interpretaciones de la probabilidad

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.

2.1.1. Teórica o clásica

En el enfoque **teórico**, **clásico** o ***a priori*** (es decir, antes de que ocurran las cosas) 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.

2.1.2. La probabilidad como frecuencia relativa

También se le conoce como enfoque ***a posteriori*** (es decir, a la luz de lo ocurrido) y al igual que el enfoque anterior es un paradigma objetivo.

Para asignarle probabilidad a un suceso se experimenta antes y a partir de los resultados se determinan las frecuencias con que ocurren los diversos resultados. En el caso de la moneda, este enfoque nos recomendaría hacer un número muy grande de “volados”, por ejemplo diez mil, y con base en ellos definir la probabilidad de una y otra cara. Si decimos, por ejemplo, que la probabilidad de que salga águila es de 4888/10000, damos a entender que lanzamos la moneda diez mil veces y que en 4888 ocasiones el resultado fue águila. Estamos entonces 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 de aplicación del enfoque *a posteriori*.

2.1.3. Interpretación subjetiva de la probabilidad

La **probabilidad subjetiva** es una cuestión de opinión. 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 del campeonato mundial de fútbol del 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.

2.2. Espacio muestral y eventos

Para trabajar con comodidad la probabilidad, vale la pena expresar algunos conceptos básicos que necesitaremos para el desarrollo del tema.

Conceptos estadísticos

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.

De estos tres conceptos podemos desprender un cuarto, el concepto de **evento aleatorio** que no es sino el resultado de un experimento aleatorio. Por ejemplo, si el experimento es arrojar un dado, por el sólo hecho de que no podemos anticipar qué cara mostrará éste al detenerse, podemos decir que el experimento es aleatorio. Uno de los resultados posibles es que salga un número par. Tal resultado es un evento aleatorio.

Para referirnos a los eventos aleatorios usaremos letras mayúsculas. De este modo podemos decir que:

A es el evento de que al arrojar un dado salga un número non.

B es el evento de que al arrojar un dado salga un número par.

Como es claro, podemos definir varios eventos aleatorios respecto del mismo experimento. Algunos de ellos tendrían la característica de que encierran a su vez varias posibilidades (en el evento A quedan incluidas las posibilidades “que salga 1”, “que salga 3” o “que salga 5”).

En este contexto, conviene distinguir eventos simples de eventos compuestos:

Los **eventos simples** son aquéllos que ya no pueden descomponerse en otros más sencillos. 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.

Los **eventos compuestos** incluyen varias posibilidades por lo que pueden descomponerse en eventos sencillos.

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.

A su vez, E1, E2 y E3 son eventos sencillos.

Ante la interrogante de qué eventos consideraremos en un experimento aleatorio dado debemos contestar que esto depende de la perspectiva que tengamos respecto del experimento aleatorio. Si estamos jugando a los dados y las apuestas sólo consideran el obtener un número par o un número impar o non, entonces los únicos resultados que nos interesarán serán precisamente esos dos: obtener número par o número impar.

Con esto damos lugar a un concepto adicional básico.

Espacio muestral: es el conjunto de todos los resultados posibles, en función de nuestra perspectiva del experimento aleatorio. También se le conoce como evento universo.

En suma, ante un experimento aleatorio cualquiera tenemos varias alternativas para definir eventos cuya probabilidad pueda sernos de interés.

Por ejemplo, si tenemos una colectividad de 47 estudiantes egresados, entre Contadores, Administradores e Informáticos de ambos sexos, y de esa colectividad seleccionamos al azar a una persona, puede ser que nos interesen las probabilidades de los siguientes eventos:

- **a)** Que la persona seleccionada haya estudiado contaduría.
- **b)** Que la persona seleccionada haya estudiado administración o contaduría.
- **c)** Que la persona seleccionada no haya estudiado administración.
- **d)** Que la persona seleccionada sea mujer y haya estudiado informática.
- **e)** Que la persona seleccionada sea hombre pero que no haya estudiado administración.

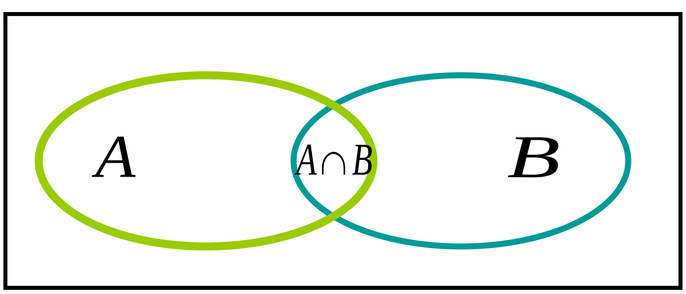
Como puede verse, en los incisos anteriores no sólo estamos manejando diversos eventos sino que además estamos incorporando relaciones entre ellos. Tales relaciones se pueden establecer de manera más eficiente recurriendo a la estructura formal de la teoría de conjuntos, esto es, incorporando los diagramas de Venn-Euler, la terminología de conjuntos, así como las operaciones que has aprendido a realizar con ellos en cursos anteriores —como la unión, la intersección, el complemento, la diferencia, entre otras— son por entero aplicables al caso de los eventos, en el contexto de la teoría de la probabilidad.

Estos elementos junto con algunas definiciones que se detallan a continuación nos permitirán trabajar adecuadamente los problemas de probabilidad que enfrentaremos.

Si definimos a los eventos A y B como resultados de un experimento aleatorio y recordamos que todos los **eventos posibles** (el conjunto universal) constituyen el **espacio muestral** y representamos éste como S, tenemos que la unión de A con B es un evento que contiene todos los puntos muestrales que pertenecen al evento A y/o que pertenecen al evento B. Podemos hacer uso de la notación de conjuntos para escribir: $A \cup B$.

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, esto es en forma simultánea. La intersección se denota en la simbología de conjuntos como $A \cap B$.

A manera de resumen en la siguiente tabla te mostramos cuatro operaciones que



Eventos simultáneos.

serán muy útiles para manejar eventos aleatorios y su equivalencia con operaciones lógicas.

Operación Lógica	Operación en conjuntos
o	Unión (U)
y	Intersección (∩)
no	Complemento (‘) Diferencia (–)

Si en nuestro ejemplo de los egresados incorporamos estas operaciones y llamamos C al evento “egresado de contaduría”, A al evento “egresado de administración”, I al evento “egresado de informática”, M al evento “mujer” y H al evento “hombre”, tendríamos que nuestro interés es conocer las siguientes probabilidades:

- **a)** Probabilidad de C
- **b)** Probabilidad de A U C
- **c)** Probabilidad de A^c
- **d)** Probabilidad de M ∩ I
- **e)** Probabilidad de H – A^c

Si además, adoptamos la convención de usar la letra P para no escribir todo el texto “probabilidad de“, y encerramos entre paréntesis el evento de interés, nuestras preguntas quedarían de la siguiente manera:

- **a)** P(C)
- **b)** P(A U C)
- **c)** P(A^c)
- **d)** P(M ∩ I)
- **e)** P(H – A^c)

Esta es la forma en que manejaremos relaciones entre eventos y denotaremos probabilidades.

2.3. Los axiomas de la probabilidad

Los elementos hasta ahora expuestos nos permiten dar ya una definición más formal de probabilidad en el contexto frecuentista:

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

Entonces tenemos que **P(A)=** $\lim_{N \to \infty} \left(\frac{n_A}{N} \right)$

Es decir, que la probabilidad de que ocurra el evento A, resulta de dividir el número de veces que A efectivamente apareció entre el número de veces que se intentó el experimento. (La expresión N → ∞ se lee «N tiende a infinito» y 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 el evento en cuestión 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.

Consideremos el siguiente ejemplo.

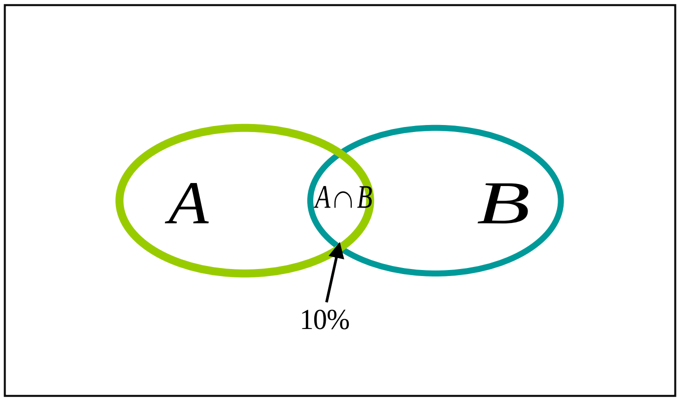
Ejemplo 1. En una investigación de mercado se encontró que entre los integrantes de un club, 30% de los hombres usan loción para después de afeitarse, en tanto que 40% de ellos utiliza desodorante y 10% utiliza ambos productos. Si elegimos al azar a un varón de ese club, ¿qué probabilidades existen de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

Solución:

Es evidente que la probabilidad que buscamos es un número positivo ya que de entre los integrantes del club sí hay varones que usan desodorante además de que también hay varones que usan loción. Es evidente además que la probabilidad que buscamos será menor a uno porque no todos usan loción y no todos usan desodorante.

Por otro lado, si hacemos que A represente el evento «El sujeto usa loción para después de afeitarse», y que B represente el evento «El sujeto usa desodorante», podemos intentar una representación gráfica empleando diagramas de Venn-Euler como sigue.

Cuando nos preguntan por la probabilidad de que la persona seleccionada al azar utilice desodorante o de que use loción para después de afeitar, sabemos que tal pregunta en lenguaje probabilístico se transforma en:



P(AUB)

Intrínsecamente la pregunta se refiere a aquellos elementos que se encuentran en A o se encuentran en B, esto es, en el interior del óvalo verde o en el interior del óvalo azul. De acuerdo con los datos, 30% de los casos se encuentran en A y 40% en B, por lo que al sumar tendríamos que aparentemente hay 70% de integrantes del club que se encuentran en la unión de ambos eventos, sólo que de ese 70% hay un 10% que es común, precisamente el porcentaje de casos que se encuentra en la intersección. Este 10% ya ha sido contado una vez al considerar el porcentaje de casos en A y fue incluido otra vez al considerar el porcentaje de casos en B, de manera que se le ha contado dos veces. Por lo tanto, para determinar el número de casos que están en la unión de A con B, debemos efectivamente considerar el 30% que está en A, el 40% que está en B, pero además debemos descontar el 10% que está en la intersección para que los elementos que están en dicha intersección sean contados sólo una vez.

De esta manera, P (AUB) = 30% + 40% – 10%.

P (AUB) = 60%

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

Las situaciones que hemos discutido dentro de este tema ilustran tres postulados básicos de la probabilidad, a los que se conoce como **Axiomas de probabilidad**, lo que en lenguaje matemático significa que son proposiciones que por su carácter evidente no requieren demostración. Constituyen, por decirlo de alguna manera, "las reglas del juego", sin importar si estamos trabajando una probabilidad subjetiva o empírica, o si seguimos los postulados de la probabilidad clásica.

Estos axiomas, que constituyen el cimiento de la teoría moderna de probabilidades y fueron propuestos por el matemático ruso Kolmogorov, se expresan de manera formal en los siguientes términos:

1) Para todo evento A, $P(A) \geq 0$
2) Si Ω representa el evento universo, entonces $P(\Omega) = 1$
3) Dados dos eventos A y B, ocurre que $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Claramente, el primer axioma nos indica que no hay probabilidades negativas y, el segundo, que ningún evento tiene una probabilidad mayor a uno.

A partir de ellos se tienen otros resultados de suyo importantes, tales como:

a) $P(\varphi) = 0$, donde φ representa el conjunto vacío.
b) $P(A^c) = 1 - P(A)$

En el segundo de estos resultados estamos haciendo referencia a **eventos complementarios.** Si Ω es el evento universo, entonces para todo evento A existe un evento complemento constituido por todos aquellos resultados del espacio muestral que no están en A, con la propiedad de que $A \cup A^c = \Omega$, por lo que $P(A \cup A^c) = P(\Omega)$, de modo que $P(A \cup A^c) = 1$.

En consecuencia, de acuerdo con el axioma (3),

$$P(A \cup A^c) = P(A) + P(A^c) - P(A \cap A^c),$$

$$\rightarrow 1 = P(A) + P(A^c) - P(A \cap A^c),$$

Sin embargo, $P(A \cap A^c) = P(\phi)$ y de acuerdo con el resultado (a), esta probabilidad es cero. Por lo tanto,

$$1 = P(A) + P(A^c),$$

de donde al despejar queda:

$$P(A^c) = 1 - P(A)$$

Ejemplo 2. Sea el experimento aleatorio que consiste en arrojar dos dados y sea Ω el espacio muestral que contiene todos los resultados posibles de sumar los puntos obtenidos. Se definen además los eventos A como el hecho de que el tiro sume menos de cuatro y B como el hecho de que la suma sea número par. Se desea determinar las probabilidades siguientes:

- **a)** $P(A^c)$
- **b)** $P(B)$
- **c)** $P(A \cup B)$

Solución:

Claramente,

$\Omega = \{2,3,4,5,6,7,8,9,10,11,12\}$,

$A = \{2,3\}$;

$B = \{2,4,6,8,10,12\}$.

Entonces,

a) De acuerdo con lo anterior, A^c= {4,5,6,7,8,9,10,11,12}, de donde se sigue que $P(A^c)$ = 9/11. Alternativamente, $P(A^c)$ = 1 – P(A), donde P(A) = 2/11, por lo que $P(A^c)$ = (11–2)/11 = 9/11, lo que confirma el resultado.

b) Es inmediato que P(B) = 6/11

c) Aplicando el axioma 3, se tiene que:

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

donde $A \cap B = \{2\}$ por lo que $P(A \cap B\}$ = 1/11.

Finalmente,

P(A U B) = 2/11 + 6/11 – 1/11

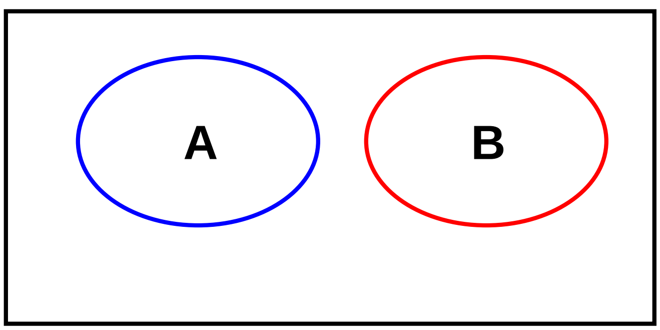
P(A U B) = 7/11

2.4. La regla de la suma de probabilidades

En el tema anterior se introdujo el axioma tres de probabilidad aplicable a cualquier pareja de eventos probabilísticos. Ahora, consideraremos un caso particular. Para ello, incorporamos primero un concepto adicional.

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.



Eventos mutuamente excluyentes.

Ejemplo 1: Sea Ω el espacio de resultados que resulta de considerar la suma de los puntos que se obtienen al arrojar dos dados.

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.

Se desea determinar las siguientes probabilidades:

a) P(A ∩ B)

b) P(A U B)

Solución:

Vemos que es imposible que ocurran A y B simultáneamente, pues para que la suma de los puntos sea doce debe ocurrir que en ambos dados salga ”seis”, pero si uno de los dos dados tiene “uno” como resultado, la suma máxima que se puede lograr es de “siete”. Los eventos son mutuamente excluyentes y, por lo tanto, P(A ∩ B) = 0.

Al aplicar el axioma 3 tenemos,

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

$P(A \cup B) = 1/36 + 11/36 - 0$

$P(A \cup B) = 12/36$

Como puede verse, el impacto de que A y B sean mutuamente excluyentes es tal que para determinar la probabilidad de la unión de dos eventos sólo debemos sumar las probabilidades de cada evento individualmente considerado.

En el caso en que A y B sean mutuamente excluyentes, esto es, cuando su intersección es vacía, la probabilidad de la unión de dos eventos es la suma de las probabilidades de los eventos tomados individualmente.

$$P(A \cup B) = P(A) + P(B) \qquad \text{si } A \cap B = \varphi$$

Si tenemos varios eventos mutuamente excluyentes en el espacio de eventos Ω y queremos saber cuál es la probabilidad de que ocurra cualquiera de ellos, la pregunta que estaríamos planteando se refiere a la probabilidad de la unión de los mismos. Al ser eventos mutuamente excluyentes, la intersección está vacía y 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**.

El siguiente ejemplo nos ayudará a dejar en claro estos conceptos.

Ejemplo 2: En un club deportivo, 20% de los socios pertenece al equipo de natación y 10% al equipo de waterpolo. Ningún socio pertenece a ambos equipos simultáneamente. Diga cuál es la probabilidad, si elegimos al azar un socio del club, de que sea integrante de alguno de los dos equipos.

Solución:

El cálculo de probabilidades aparece a continuación. El estudiante debe tener en mente que, dado que ningún socio pertenece a los dos equipos simultáneamente, la intersección está vacía y por lo mismo su probabilidad es cero.

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

2.5. Tablas de contingencias y 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.

En este tema nos abocaremos a analizar este tipo de situaciones. Para ello es conveniente introducir dos conceptos preliminares.

Probabilidad simple (marginal)

En un experimento cualquiera, 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**.

Repasemos a continuación el procedimiento para calcular la probabilidad simple o marginal de un evento.

1. Definimos el experimento.
2. Hacemos la lista de todos los eventos simples asociados con el experimento que definió (es decir, haga la lista de todos los puntos muestrales).
3. Asignamos 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. Definimos el evento que le interesa como un conjunto de puntos muestrales.

5. Encontramos 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.

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

2. 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

3. 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.

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

   a. Evento A: que salga un número menor a cuatro. Se compone de los eventos E1, E2 y E3.
   b. Evento B: que salga un número par. Se compone de los eventos E2, E4, E6.

c. Evento C: que salga un número mayor que seis. Ningún evento lo compone.

5. Calculamos las probabilidades solicitadas:

- 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.
- La probabilidad de B es la suma de las probabilidades de E2, E4, E6: 1/6+1/6+1/6 = 3/6 = 1/2.
- 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 durante 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 directivo está compuesto por: M1, M2, M3, H1 y H2, donde M1 es la primera mujer, M2 la segunda, H1 el primer hombre y así sucesivamente.

Los eventos simples factibles 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.

Por otro lado, las parejas que están constituidas por un hombre y 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 de 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. Nos interesa saber cuál es la probabilidad de que acepte el embarque.

Solución:

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

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.

A continuación listamos todos los eventos posibles (es decir, todos los pares diferentes que se pueden elegir). Los eventos simples de interés (aquellos en que los dos refrigeradores están en buen estado) están resaltados.

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), tendremos que la probabilidad de aceptar el embarque es 6/15.

Probabilidad conjunta

En muchas ocasiones estaremos enfrentando problemas en los que nuestros eventos de interés estarán definidos por la ocurrencia de dos o más eventos simples.

Tomemos el caso del siguiente ejemplo.

Ejemplo 4. Consideremos el caso de una pareja que tiene dos hijos, situación respecto de la cual definimos los siguientes eventos de interés:

Evento A: La pareja tiene por lo menos un varón.

Evento B: La pareja tiene por lo menos una niña.

Nuestros eventos de interés se pueden expresar de la siguiente manera:

Evento A: Ocurre si se tiene varón y varón, varón y mujer en ese orden, o mujer y varón en ese orden.

Evento B: Ocurre si se tiene mujer y mujer, varón y mujer en ese orden o mujer y varón en ese orden.

Como puede verse, para que ocurra el evento A deben ocurrir dos cosas simultáneamente. Bien sea:

Varón **y** varón, o

Varón **y** mujer, o

Mujer **y** varón.

Si definimos los eventos simples V: varón y M: mujer, tendríamos que cada una de las posibilidades que se tienen para que ocurra el evento A implica la ocurrencia de dos o más eventos simples.

Algo similar puede decirse en relación al evento B.

Cuando los eventos de interés implican la ocurrencia de dos o más eventos simples de manera simultánea, decimos que estamos en presencia de una **probabilidad conjunta**.

El lector puede confirmar que en el ejemplo 3 también estábamos en presencia de probabilidades conjuntas, aunque por la perspectiva que se adoptó aparecían como simples.

Probabilidad condicional

Dados dos eventos podemos preguntarnos por la probabilidad de uno de ellos bajo el supuesto de que el otro ya ocurrió. Al inicio de este tema, por ejemplo, se planteaba la situación respecto de la probabilidad de pasar un examen si el estudiante realmente estudió para dicho examen. Este tipo de situaciones dan lugar a la **probabilidad condicional**.

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

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

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

El siguiente ejemplo nos ayudará a clarificar estas ideas.

Ejemplo 5. Sea el evento A: Amanece nublado en la región X

De acuerdo con información meteorológica recopilada a lo largo de muchos días, se sabe que:

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, 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.

Solución:

En el inciso “a” deseamos saber la probabilidad de B dado que A. Con la información que tenemos podemos sustituir directamente en la expresión para la probabilidad condicional.

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 de 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 marginal de que llueva (sin tener antecedentes) es de 50%.

En el inciso (b) deseamos conoce la probabilidad de que llueva en la tarde dado que amaneció despejado, esto es, buscamos B dado que A^c ya ocurrió. Como amanece nublado 60% de los días y despejado 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 de 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.

Tablas de contingencia

Una **tabla de probabilidad conjunta** es aquella 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**.

Su elaboración incluye formar una tabla de contingencia cuyos valores de cada celda se dividen entre el total de datos para obtener los valores de probabilidad correspondientes.

Ejemplo 6: *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
Menor de 30 años	35	46	81
Entre 30 y 40 años	42	59	101
Mayor de 40 años	51	67	118
Total	128	172	300

Tabla de probabilidad conjunta

Evento	Edad /sexo	Hombre H	Mujer M	Probabilidad marginal
E_1	**Menor de 30 años**	0.117	0.153	0.270
E_2	**Entre 30 y 40 años**	0.140	0.197	0.337
E_3	**Mayor de 40 años**	0.170	0.223	0.393
	Probabilidad marginal	0.427	0.573	1.000

Con esta información se desea 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) Sea mujer dado que tiene entre 30 y 40 años.

Solución:

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

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

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

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

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

f) $$P\left(M/E_2\right)=\frac{P\left(E_2\cap M\right)}{P\left(E_2\right)}=\frac{0.197}{0.337}=0.585=58.5\%$$

Las ideas que hemos presentado en esta sección nos permiten reformular la probabilidad marginal como la probabilidad incondicional de un evento particular simple, que consiste en una suma de probabilidades conjuntas. Si en el ejercicio anterior se desea calcular la probabilidad de que el siguiente cliente sea un hombre, esto podría hacerse a partir de probabilidades conjuntas, como sigue:

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

o sea:

$$P\left(H\right)=0.117+0.140+0.170=0.427=42.7\%$$

2.6. Independencia estadística

Sean dos eventos A y B del espacio de eventos Ω; decimos que **A y B son independientes en sentido probabilístico 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:

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

Es decir, el hecho de que ocurra un evento no modifica la probabilidad de que ocurra el otro, sin importar quien sea condición de quien.

Consideremos el siguiente ejemplo.

Ejemplo 1. Una tienda de departamentos ha solicitado a un despacho de consultoría que aplique un cuestionario para medir si su propaganda estática tenía impactos distintos según el grupo de edad del público. Como parte del estudio el despacho entrevistó a 150 mujeres, a las cuáles se les preguntó si recordaban haber visto dicha propaganda. Los resultados se muestran a continuación:

	Sí la recuerdan	No la recuerdan	Total
Menores de 40 años	40	30	70
40 o más años de edad	20	60	80
Total	60	90	150

Sean los eventos siguientes:

S es el evento «Sí recuerda la propaganda»

N es el evento «No recuerda la propaganda»

J es el evento «Menor de 40 años de edad»

E es el evento «40 o más años de edad»

Se desea saber

a) Si los eventos S y J son independientes en sentido probabilístico
b) Si los eventos N y E son independientes en sentido probabilístico

Solución:

a) Para saber si los eventos son independientes basta calcular P(S) y P(S│J) y comparar.
De acuerdo con los datos de la tabla,

$$P(S) = 60/150,$$

Por su parte, para determinar el valor de P(S│J) observamos que al ser J la condición, podemos modificar el universo de resultados y restringirlo sólo a aquéllos que cumplen con dicha condición. Así, el nuevo universo es de sólo 70 casos, de los cuales 40 recuerdan la propaganda. En consecuencia,

$$P(S \mid J) = 40/70$$

Es inmediato que las probabilidades no son iguales, por lo que podemos afirmar que S y J no son independientes.

b) Al igual que en el inciso anterior, para saber si los eventos son independientes basta calcular P(N) y P(N│E) y comparar.
De acuerdo con los datos de la tabla,

$$P(N) = 90/150,$$

Por su parte, para determinar el valor de P(N│E) observamos que al ser E la condición, podemos modificar el universo de resultados y restringirlo sólo a aquéllos que cumplen con dicha condición. Así, el nuevo universo es de sólo 80 casos, de los cuales 60 recuerdan la propaganda. En consecuencia,

P(N│E) = 60/80

Es inmediato que las probabilidades no son iguales, por lo que podemos afirmar que N y E no son independientes en sentido probabilístico.

El lector puede confirmar que las otras parejas de eventos tampoco son independientes.

2.7. La regla de multiplicación de probabilidades

Recordemos que en general,

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

Si A y B son independientes probabilísticamente, P(B│A) = P(B), por lo que:

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

De aquí se sigue que:

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

Podemos decir en consecuencia que **si dos eventos son estocásticamente independientes, entonces su probabilidad conjunta es igual al producto de sus probabilidades marginales, y a la inversa, si la probabilidad conjunta de dos eventos es igual al producto de sus probabilidades marginales entonces esos dos eventos son independientes probabilísticamente**.

A este resultado se le conoce como la **regla de la multiplicación de probabilidades**.

Dos eventos A y B son independientes probabilísticamente si y sólo si

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

Consideremos un ejemplo sencillo.

Ejemplo 1. Se arroja una moneda tres veces. Se desea determinar la probabilidad de obtener cara, cruz y cara en ese orden.

Solución:

Sea C el evento «sale cara»y X el evento «sale cruz».

Se desea determinar P(C, X, C). Por otro lado, nuestra experiencia —asumiendo que la moneda es legal— nos dice que la probabilidad de obtener cruz o cara en un determinado lanzamiento de la moneda no se altera por la historia de los resultados anteriores. Esto significa que podemos asumir que los eventos son independientes probabilísticamente, por lo que:

P(C, X, C) = P(C)P(X)P(C)

Como cada probabilidad marginal es igual a 0.5, el resultado final es 0.125.

2.8. 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 de estadística.

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

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

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	Equivalente a 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.

El gerente de crédito desea calcular 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.

$P_2 =$ Clientes que pagan tarde.

$P_3 =$ Clientes que no pagan.

$D =$ Clientes dueños de sus casas.

$D' =$ Clientes que no son dueños de sus casas.

Expresión general:

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

Donde,

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

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

Sustituyendo en la fórmula general:

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

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:

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

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.

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

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 sólo 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** $P(D \mid P_i)$	**Probabilidad Conjunta** $P(D \mid P_i)^{\times} P(P_i)$	**Probabilidad *a posteriori*** $P(P_i \mid D)$
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**

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.

RESUMEN

La probabilidad es una rama de las matemáticas, cuyo desarrollo tiene su génesis en el siglo XVII, cuando se buscó contar con métodos racionales de enfrentar los juegos de azar. Se puede decir que hay tres grandes enfoques, escuelas o paradigmas de probabilidad, a saber, el clásico, el empírico y el subjetivo, ninguno de los cuales escapa al tratamiento axiomático, que es lo que da la estructura al tratamiento matemático moderno de la probabilidad. Como parte de esta estructura matemática se incorporan, además, el cálculo de probabilidades a la luz de información adicional bajo el concepto de probabilidad condicional y del teorema de Bayes.

BIBLIOGRAFÍA

SUGERIDA

Autor	Capítulo	Páginas
Anderson, Sweeney, Williams (2005)	4. Introducción a la probabilidad. Sección 4.2 Eventos y sus probabilidades.	143-146
	4.3 Algunos resultados básicos de probabilidad.	148-151
	4.4 Probabilidad condicional.	153-156
	5. Teorema de Bayes.	161-165
Berenson, Levine y Krehbiel (2001)	4. Probabilidad básica y distribuciones de probabilidad. Sección: 4.1 Conceptos básicos de probabilidad.	155-165
	4.2 Probabilidad condicional.	165-175
	4.3 Teorema de Bayes.	175-179

Levin y Rubin (2004)	4. Probabilidad I: Ideas introductorias. Sección: 4.2 Terminología básica en probabilidad.	129-131
	4.3 Tres tipos de probabilidad.	131-137
	4.4 Reglas de probabilidad.	137-143
	4.5 Probabilidades bajo condiciones de independencia estadística.	143-148
	4.6 Probabilidades bajo condiciones de dependencia estadística.	151-155
	4.7 Revisión de las estimaciones anteriores de probabilidades: teorema de Bayes.	158-165
Lind, Marchal, Wathen (2008)	5. Estudio de los conceptos de la probabilidad. Secciones: ¿Qué es la probabilidad?	140–141
	Enfoques para asignar probabilidades.	142-147
	Algunas reglas para calcular probabilidades.	147-156
	Tablas de contingencias.	156-158
	Teorema de Bayes.	161-165

Anderson, David R.; Sweeney, Dennis J.; Williams, Thomas A. (2005). *Estadística para administración y economía*, 8ª edición, México: International Thomson Editores, pp. 888 más apéndices.

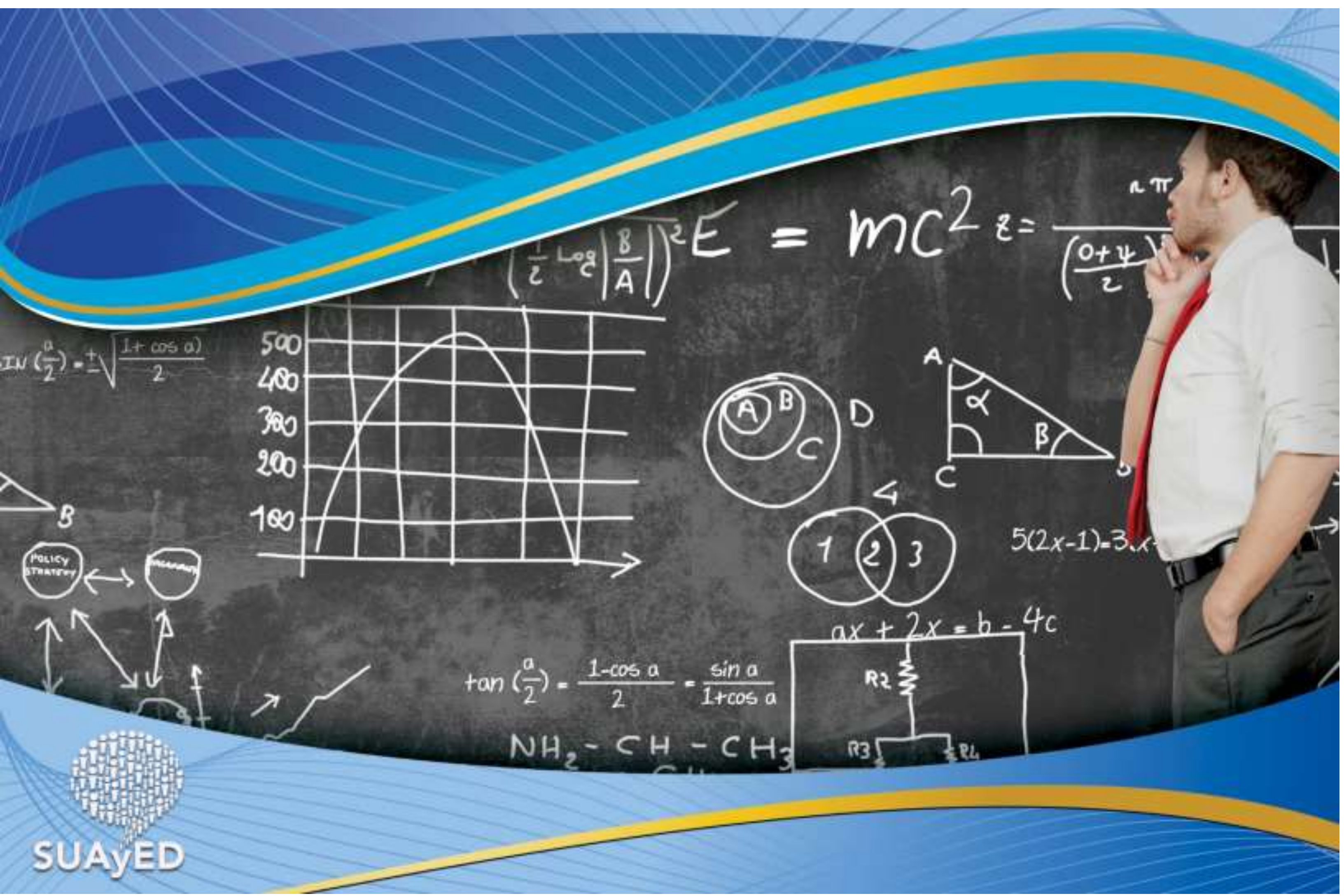
Berenson, Mark L., David M. Levine, y Timothy C. Krehbiel. (2001). *Estadística para administración*, 2ª edición, México: Prentice Hall, 734 pp.

Levin, Richard I. y David S Rubin. (2004). *Estadística para administración y economía*, 7ª edición, México: Pearson Educación Prentice Hall, pp. 826 más anexos.

Lind, Douglas A., Marchal, William G. y Wathen, Samuel, A. (2008). *Estadística aplicada a los negocios y la economía,* 13ª edición, México: McGraw-Hill Interamericana, 859 pp.

Unidad 3.

Distribuciones de probabilidad



OBJETIVO PARTICULAR

El alumno aplicará las diferentes distribuciones de probabilidad y su interpretación en la solución de problemas.

TEMARIO DETALLADO
(12 horas)

3. Distribuciones de probabilidad

3.1. Variables aleatorias, discretas y continuas
3.2. Media y varianza de una distribución de probabilidad
3.3. Distribuciones de probabilidad de variables discretas
3.3.1. Distribución binomial
3.3.2. Distribución de Poisson
3.3.3. La distribución de Poisson como aproximación de la distribución binomial
3.3.4. Distribución hipergeométrica
3.3.5. Distribución multinomial
3.4 Distribuciones de probabilidad de variables continuas
3.4.1 Distribución normal
3.4.2 Distribución exponencial
3.5 Ley de los grandes números

INTRODUCCIÓN

En esta unidad 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.

3.1. 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.

La variable aleatoria considera situaciones donde los resultados pueden ser de origen cuantitativo o cualitativo, asignando en cualquier caso un número a cada posible resultado.

Por ejemplo, si el experimento consiste en seleccionar a una persona de un colectivo de n de ellas, y lo que nos interesa es el sexo, la variable aleatoria podría tomar los valores 1 si resulta ser un hombre y 2 si resulta ser una mujer. Si lo que nos interesa es la edad, entonces la variable aleatoria tiene tantos posibles valores como edades haya en la población.

En esencia, lo que hace una variable aleatoria es asignar un número a cada posible resultado del experimento.

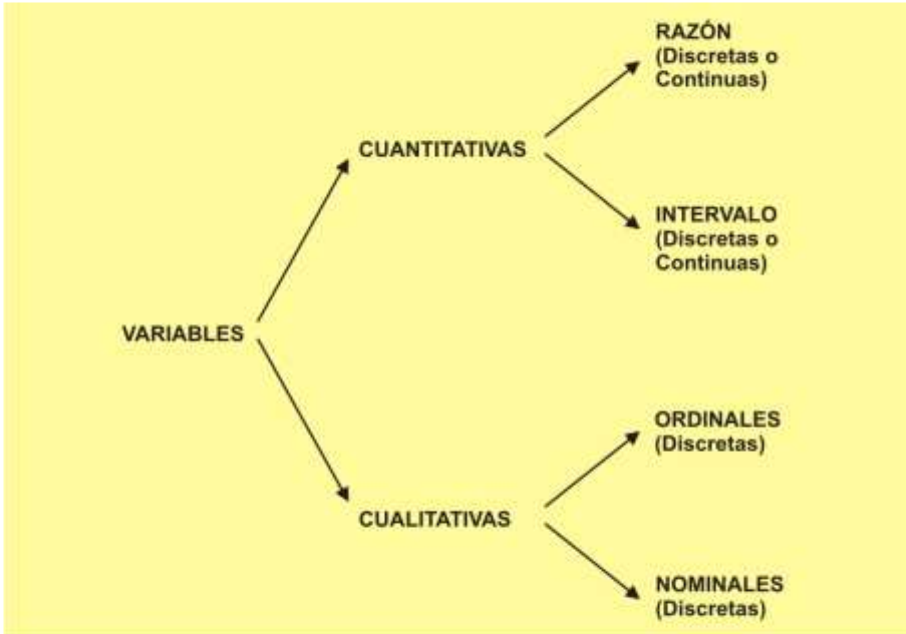
Dependiendo de esta asignación de números las variables aleatorias pueden ser discretas o continuas.

- Las **variables discretas** son aquellas que cuantifican la característica de modo tal que el número de posibles resultados se puede contar, esto es, la variable discreta toma un número finito o infinito numerable de posibles valores. 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, 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.

Desde el punto de vista de la estadística las variables aleatorias también se clasifican de acuerdo a la escala de medición inherente.

Cuando estudiaste el tema de estadística descriptiva tuviste oportunidad de aprender los conceptos de escala nominal, ordinal, de intervalo y de razón. Estas escalas generan precisamente variables aleatorias del mismo nombre. Ocurre que las variables de intervalos y de razón son cuantitativas y pueden ser discretas o continuas. Los casos nominal y ordinal se refieren a cualidades en donde la variable aleatoria al asignar un número a cada resultado asume que tales cualidades son discretas. El cuadro siguiente te proporciona un panorama general de esta situación.



VARIABLES ALEATORIAS

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, en este campo del conocimiento, las variables se clasifican, desde el punto de vista metodológico, en dependientes e independientes.

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

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 1. 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.

Un punto importante que debes tener en mente cuando trabajes con variables aleatorias es que no sólo es importante identificarlas y clasificarlas, sino que también deben definirse adecuadamente. Para algunos autores, como 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 útil, 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, estás ahora en posibilidad de acotar adecuadamente las variables para un manejo estadístico, de acuerdo con el interés que tengas en ellas, para la realización de un estudio o investigación. Mostraremos a continuación un par de ejemplos de ello.

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".

Finalmente, es importante mencionar que a la par que defines una variable aleatoria es importante que le asignes un nombre. Por lo general, éste es una letra mayúscula.

3.2. Media y varianza de una distribución de probabilidad

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, a la que también se conoce como **función de densidad**, representada por **f(X)**, que define la probabilidad de cada valor de la variable aleatoria.

Como la probabilidad del universo (o evento universal) debe ser igual a 100%, y además cualquier evento que se defina debe estar contenido en el evento universal, cuando hablamos de cómo distribuir las probabilidades nos referimos a cómo es que se reparte este 100% de probabilidad en los diferentes eventos.

Ejemplo 1. Considera el experimento aleatorio que consiste en arrojar un dado dos veces y sumar los resultados de ambas caras. Se desea conocer cuál es la probabilidad de que la suma sea 7.

Solución:

La variable X puede tomar los valores del 2 al 12, inclusive, por lo que se trata de una variable aleatoria discreta. La siguiente tabla nos permitirá calcular las probabilidades de todos los eventos simples.

Resultado	Segundo dado					
Primer dado	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

En ella vemos que las diagonales, a las que se ha dado diferente color, determinan el mismo valor de la suma para diferentes combinaciones de resultados de cada uno de los dos dados. Por ejemplo, si queremos saber la probabilidad de que la suma sea 7, nos fijaríamos en la diagonal amarilla y observaríamos que hay 6 formas distintas de obtener tal valor, de un total de 36, por lo que la probabilidad es 7/36.

El ejemplo nos permite darnos cuenta, además, que también podemos calcular fácilmente la probabilidad de que la suma sea menor o igual a 7 y que para ello debemos contar el número de casos que se acumulan desde la diagonal superior izquierda hasta la diagonal amarilla, que corresponde a los valores 2, 3, 4, 5 ,6 y 7. Esto es, se estaría considerando que:

$$P\ (X\leq 7) = P\ (2) + P\ (3) + P\ (4) + P\ (5) + P\ (6) + P\ (7)$$

Para cualquier otro resultado también estaríamos acumulando probabilidades desde la que corresponde al resultado 2 hasta el resultado tope considerado.

De este modo se construye, a partir de la función de probabilidades, otra función, a la que se denomina **función de distribución acumulativa** y que se denota como **F(*x*)**, donde la x indica el valor hasta el cual se acumulan las respectivas probabilidades. Por ejemplo, P (X≤7) corresponde a F(7).

La tabla siguiente resume la función de probabilidades y la función de distribución acumulativa para el caso del ejemplo:

i	Función de probabilidad P(X = i)	Función de distribución acumulativa P(X ≤ i)
2	1/36	1/36
3	2/36	1/36 + 2/36 = 3/36
4	3/36	3/36 + 3/36 = 6/36
5	4/36	6/36 + 4/36 = 10/36
6	5/36	10/36 + 5/36 = 15/36
7	6/36	15/36 + 6/36 = 21/36
8	5/36	21/36 + 5/36 = 26/36
9	4/36	26/36 + 4/36 = 30/36
10	3/36	30/36 + 3/36 = 33/36
11	2/36	33/36 + 2/36 = 35/36
12	1/36	35/36 +1/36 = 36/36 = 1

Obsérvese que el valor de la función de distribución acumulativa para el último valor de la variable aleatoria acumula precisamente 100%.

Esperanza y varianza

Cuando se trabaja con variables aleatorias, no basta con conocer su distribución de probabilidades. También será importante obtener algunos valores típicos que resuman, de alguna forma, la información contenida en el comportamiento de la variable. De esos valores importan fundamentalmente dos: la esperanza y la varianza.

Esperanza.

Corresponde al valor promedio, considerando que la variable aleatoria toma los distintos valores posibles con probabilidades que no son necesariamente iguales. Por ello se calcula como la suma de los productos de cada posible valor de la variable aleatoria por la probabilidad del respectivo valor. Se le denota como μ

$$\text{Esperanza} = \mu = \sum x\left[P(X = x)\right]$$

Donde la suma corre para todos los valores x de la variable aleatoria.

Varianza

Es el valor esperado o esperanza de las desviaciones cuadráticas con respecto a la media μ. Se denota como σ^2 y se calcula como la suma del producto de cada desviación cuadrática por la probabilidad del respectivo valor.

$$\text{Varianza} = \sigma^2 = \sum (x - \mu)^2\left[P(X = x)\right]$$

Donde la suma corre para todos los valores x de la variable aleatoria.

La raíz cuadrada de la varianza es, desde luego, la desviación estándar.

Ejemplo 2. Considerando el mismo experimento del ejemplo anterior, determinar la esperanza y varianza de la variable aleatoria respectiva.

X	Función de probabilidad P(X = *x*)	*x* P(X = *x*)	$(x-7)^2$	$(x-7)^2$ P(X = x)
2	1/36	2/36	25	25/36
3	2/36	6/36	16	32/36
4	3/36	12/36	9	27/36
5	4/36	20/36	4	16/36
6	5/36	30/36	1	5/36
7	6/36	42/36	0	0
8	5/36	40/36	1	5/36
9	4/36	36/36	4	16/36
10	3/36	30/36	9	27/36
11	2/36	22/36	16	32/36
12	1/36	12/36	25	25/36
	Suma	**252/36=7**		**260/36**

Podemos decir entonces que al arrojar dos dados y considerar la suma de los puntos que cada uno muestra, el valor promedio será 7 con una desviación estándar de 2.69.

3.3. Distribuciones de probabilidad de variables discretas

Las distribuciones binomial, de Poisson, hipergeométrico y multinomial son cuatro casos de distribuciones de probabilidad de variables aleatorias discretas.

3.3.1. Distribución binomial

La distribución binomial se relaciona con un experimento aleatorio conocido como **experimento de Bernoulli** el cual tiene las siguientes características:

- El experimento está constituido por un número finito, n, de pruebas idénticas.
- Cada prueba tiene exactamente dos resultados posibles. A uno de ellos se le llama arbitrariamente éxito y al otro, fracaso.
- La probabilidad de éxito de cada prueba aislada es constante para todas las pruebas y recibe la denominación de "p".
- Por medio de la distribución binomial tratamos de encontrar un número dado de éxitos en un número igual o mayor de pruebas.

Puesto que sólo hay dos resultados posibles, la probabilidad de fracaso, a la que podemos denominar q, está dada por la diferencia 1 – p, esto es, corresponde al complemento de la probabilidad de éxito, y como esta última es constante, entonces también lo es la probabilidad de fracaso.

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

$$P(x) = {}_{n}C_{x}\, p^{x} q^{n-x}$$

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 en grupos de x por la probabilidad de éxito elevada al número de éxitos deseado y por la probabilidad de fracaso elevada al número de fracasos deseados.

Con el término ***combinaciones*** *nos referimos al número de formas en que podemos extraer grupos de k objetos tomados de una colección de n de ellos (n ≥ k), considerando que el orden en que se toman o seleccionan no establece diferencia alguna. El símbolo nCk denota el número de tales combinaciones y se lee combinaciones de n objetos tomados en grupos de k. Operativamente,*

nCk = n! / [x! (n-x)!]

El símbolo ! indica el factorial del número, de modo que

n! = (1)(2)(3)…(n)

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

Ejemplo 1. Un embarque de veinte televisores incluye tres unidades defectuosas. Si se inspeccionan tres televisores al azar, indica cuál es la probabilidad de que se encuentren dos defectuosos.

Solución:

Podemos verificar si se trata de una distribución binomial mediante una lista de chequeo de cada uno de los puntos que caracterizan a esta distribución.

Característica	Estatus	Observación
Hay un número finito de ensayos	SI	Cada televisor es un ensayo y hay 3 de ellos
Cada ensayo tiene sólo dos resultados	SI	Cada televisor puede estar defectuoso o no
La probabilidad de éxito es constante	SI	La probabilidad de que la unidad esté defectuosa es 3 / 20
Se desea saber la probabilidad de un cierto número de éxitos	SI	Se desea saber la probabilidad de que X=2

Una vez que hemos confirmado que se trata de una distribución binomial aplicamos la expresión

$P(x)= nCx\ p^x q^{(n-x)}$, de modo que:

$$P(2) = {}_3C_2\ (3/20)^2(17/20)^1 = 3\ (0.0225)\ (0.85) = 0.057375$$

Ejemplo 2. Una pareja de recién casados planea tener tres hijos. Di cuál es la probabilidad de que los tres hijos sean varones si consideramos que la probabilidad de que el descendiente sea hombre o mujer es igual.

Solución:

Verificamos primero si se cumplen los puntos que caracterizan la distribución binomial.

Claramente, es un experimento aleatorio con tres ensayos, y en todos ellos sólo hay dos resultados posibles, cada uno con probabilidad de 0.5 en cada ensayo. Si se define como éxito que el sexo sea masculino, entonces podemos decir que se desea saber la probabilidad de que haya tres éxitos.

Entonces, el experimento lleva a una distribución binomial y,

$$P(3) = {}_3C_3\,(1/2)^3(1/2)^0 = (1/2)^3 = 0.0125$$

Ejemplo 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?

Solución:

Al igual que en los casos anteriores verificamos si se cumple o no que el experimento lleva a una distribución binomial.

Se trata de un experimento con seis ensayos, en cada uno de los cuales puede ocurrir que el estudiante sepa o no sepa localizar Afganistán en el mapa. Si se define como éxito que sí sepa la localización podemos decir que la probabilidad de éxito es de 0.30. Además, las probabilidades que se desea calcular se refieren al número de éxitos. Concluimos que el experimento es Bernoulli y, por lo tanto,

$P(2) = {}_6C_2\ (0.30)^2(0.70)^4 = 15\ (0.09)\ (0.2401) = 0.324135$

Por cuanto hace al inciso (b), la frase «un máximo de dos» significa que X toma los valores cero, uno o dos. Entonces,

$P(X \leq 2) = P(2) + P(1) + P(0) =$

$= {}_6C_2\ (0.30)^2(0.70)^4 + {}_6C_1\ (0.30)^1(0.70)^5 + {}_6C_0\ (0.30)^0(0.70)^6 =$

$= 15(0.09)(0.2401) + 6(0.30)(0.16807) + 0.1176 =$

$= 0.661941$

Esperanza y varianza de una variable aleatoria binomial

Consideremos de nueva cuenta el ejemplo 1.

¿Qué pasa con las probabilidades de los otros valores posibles para la variable aleatoria? Si hacemos los cálculos respectivos tendríamos:

$P(0) = {}_3C_0\ (3/20)^0(17/20)^3 = 0.614125$

$P(1) = {}_3C_1\ (3/20)^1(17/20)^2 = 3\ (0.15)(0.7225) = 0.325125$

$P(3) = {}_3C_3\ (3/20)^3(17/20)^0 = (0.003375) = 0.003375$

Si recordamos que P(2) = 0.057375, entonces podemos confirmar que

$$P(0) + P(1) + P(2) + P(3) = 1.00,$$

lo que era de esperarse puesto que los valores 0, 1, 2 y 3 constituyen el universo en el experimento en cuestión.

Con estos valores podemos determinar la esperanza y varianza de la variable aleatoria considerada. Para ello nos es útil acomodar los datos en una tabla recordando que:

$$\mu = \Sigma\, x\, [P(X=x)], \text{ y que, } \sigma^2 = \Sigma\, (x - \mu)^2\, [P(X=x)]$$

x	Función de probabilidad P(X = *x*)	*x* P(X = *x*)	$(x - 0.45)^2$	$(x - 0.45)^2$ P(X = x)
0	0.614125	0.000000	0.2025	0.124360
1	0.325125	0.325125	0.3025	0.098350
2	0.057375	0.114750	2.4025	0.137843
3	0.003375	0.010125	6.5025	0.021946
	Suma	0.450000		0.382500

Entonces, la esperanza es 0.45 y la varianza 0.3825.

Si interpretamos las probabilidades anteriores en un sentido frecuentista, diríamos que si consideramos un número grande de realizaciones del experimento, por ejemplo un millón de veces, en aproximadamente 614 125 realizaciones tendremos refrigeradores sin defecto, en 325 125 veces encontraremos un refrigerador con defecto, en otras 57 375 ocasiones encontraremos dos refrigeradores con defecto y en 3 375 veces los tres refrigeradores estarían defectuosos.

Con estos datos podemos elaborar una tabla de distribución de frecuencias y calcular el promedio de refrigeradores defectuosos.

Número de refrigeradores defectuosos (x)	Frecuencia (f)	fm
0	614 125	0
1	325 125	325 125
2	57 375	114 750
3	3 375	10 125
Total	1 000 000	450 000

Luego,

μ = 450 000 / 1 000 000 = 0.45

Asimismo, podemos calcular la varianza:

σ^2= [614 125 $(0\text{-}0.45)^2$ + 325 125 $(1\text{-}0.45)^2$ + 57 375 $(2\text{-}0.45)^2$ + 3 375 (3–0.45)2] / 100

= (124 360.313 + 98 350.3125 + 137 843.438 + 29 945.9375) / 100

= 0.3825

Observa que hemos seguido fielmente las lecciones de estadística descriptiva en el cálculo de μ y σ y que hemos llegado a los mismos valores que ya habíamos obtenido. Esto nos proporciona por lo menos un esquema con el cual podemos interpretar la esperanza y varianza, haciendo uso del concepto de frecuencias.

Es importante además, darse cuenta que podemos llegar a estos mismos valores de un modo más sencillo si nos percatamos que

- μ = 0.45 es precisamente el resultado que se obtiene al multiplicar el número de ensayos por la probabilidad de éxito, esto es, 3(0.15)

- $\sigma^2 = 0.3825$ es precisamente el resultado que se obtiene al multiplicar el número de ensayos por la probabilidad de éxito y por la de fracaso, esto es, 3(0.15)(0.85)

En otras palabras,

Media y varianza de una variable aleatoria binomial
$\mu = np \qquad \sigma^2 = npq$

Puede ocurrir, como en el caso del ejemplo anterior, que la esperanza da un valor que no coincide con los valores posibles de la variable aleatoria. Por eso se dice que la esperanza es un valor ideal.

Por otra parte, si desglosamos cada uno de los elementos que integran la expresión del cálculo de probabilidades de la distribución binomial y consideramos las expresiones para el cálculo de la media y la varianza, tendremos que:

$$nCx = n! / [x! (n-x)!]$$

$$p^x = p^x$$

$$q^{n-x} = (1-p)^{n-x}$$

$$\text{Media} = np$$

$$\text{Varianza} = n\, p\, (1 - p)$$

Lo que nos revela que para poder calcular cualquier probabilidad con el modelo binomial o su esperanza o varianza debemos conocer los valores de *n*, el número total de ensayos, y de *p*, la probabilidad de éxito. El valor de *x*, el número de éxitos se establece de acuerdo con las necesidades del problema.

Lo anterior nos permite concluir que la distribución binomial queda completamente caracterizada cuando conocemos los valores de ***n*** y ***p***. Por esta razón a estos valores se les conoce como los **parámetros de la distribución**.

Un error que suele cometerse a propósito de la distribución binomial es considerar que sus parámetros son la esperanza y varianza de la variable aleatoria respectiva. En realidad estos dos valores se expresan en función de los parámetros.

El siguiente ejemplo nos ayudará a entender este concepto.

Ejemplo 4*:* 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 personas, ¿cuántos esperamos que tengan parásitos intestinales?

$$\text{Media}=np=1400(0.20)=280$$

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

Usando el teorema de Tchebyshev 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.

Teorema de Tchebyshev *El teorema de Tchebyshev señala que la probabilidad de que una variable aleatoria tome un valor contenido en k desviaciones estándar de la media es cuando menos 1 – 1 / k*^{2}

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 un intervalo de 250 a 310 personas que tienen problemas.

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

3.3.2. Distribución de Poisson

Es otra distribución teórica de probabilidad de variable aleatoria discreta y tiene muchos usos en economía y comercio. Se debe al teórico francés Simeón Poisson quien la derivó en 1837 como 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, de manera que al contrario del modelo binomial, se tiene un número infinito de ensayos.

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.

En cualquier caso, sólo se requiere conocer el **número promedio de éxitos para la dimensión específica 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:

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

En esta fórmula, x representa el número de éxitos cuya probabilidad deseamos calcular; λ es el promedio de éxitos en un periodo de tiempo o en un cierto espacio; “e” es la base de los logaritmos naturales; y el símbolo de admiración representa el factorial del número que se trate.

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:

En cada caso debemos establecer primero el número promedio de errores. En el inciso (a) nos referiremos al número promedio por cada 25 páginas, en el inciso (b) por cada 40 páginas y en el (c) por página. Esto lo podemos hacer mediante el procedimiento de proporcionalidad directa o regla de tres.

a)Dos errores en 25 páginas

Datos

$$40 - 400$$
$$\lambda - \quad 25$$
$$\therefore \quad \lambda = 2.5$$
$$x = 2$$
$$e = 2.71828$$

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

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

b)Más de dos errores en 40 páginas

Datos

$$40 - 400$$
$$\lambda - \quad 40$$
$$\therefore \quad \lambda = 4$$
$$x = 2$$
$$e = 2.71828$$

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

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

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

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

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

$$40\ -\ 400$$
$$\lambda -\quad 1$$
$$\therefore\quad \lambda=0.10$$
$$x=2$$
$$e=2.71828$$

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

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

Un aspecto importante de la distribución de Poisson es que su media y varianza son iguales. De hecho, Media y Varianza de una variable aleatoria Poisson

$\mu = \lambda \qquad \sigma^2 = \lambda$

De acuerdo con lo anterior, para determinar las probabilidades en un modelo de Poisson o calcular su esperanza o varianza debemos conocer el valor de λ, esto es, del número promedio de éxitos. Éste es el parámetro de la distribución.

3.3.3. La distribución de Poisson como una aproximación a la distribución binomial

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. 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. Sin embargo, dados los volúmenes y probabilidades que se manejan es conveniente trabajar con la distribución 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: $\mu = np = 3,000 \times 0.001 = 3$

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

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

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

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

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

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.

3.3.4. Distribución hipergeométrica

Este es otro caso de una distribución de variable aleatoria discreta y guarda aparentemente un gran parecido con la distribución binomial, por cuanto en ambas hay un número finito de ensayos, cada uno de los cuales pertenece a uno de dos grupos (el equivalente a éxito o fracaso). Sin embargo, hay otros rasgos que distinguen claramente al modelo hipergeométrico del binomial.

Para aplicar este modelo se requiere verificar los siguientes puntos:

- Hay un población constituida por N observaciones
- La población se puede dividir en dos grupos, K y L, en el primero de los cuales hay k observaciones y en el otro N-k
- De la población se seleccionan al azar n observaciones
- Se desea determinar la probabilidad de que en la muestra haya x observaciones que pertenecen al grupo K

El lector podrá observar que en este caso no se hace ninguna mención explícita en torno a que la probabilidad de éxito sea constante, como en el caso del modelo binomial. Esto se debe a que en el modelo hipergeométrico la extracción de las observaciones no sigue un esquema con reemplazo por lo que la probabilidad de éxito ya no es constante.

Si imaginamos que hacemos la extracción de la muestra elemento por elemento, la probabilidad de que el primero en ser extraído sea del grupo K es, evidentemente, k / N. A continuación, procederíamos a extraer el segundo, pero en este caso el número de casos totales ya no sería N sino N–1 y el número de casos favorables ya no sería k sino k–1, por lo que la probabilidad de que este segundo elemento provenga del grupo K sería (k–1) / (N–1).

Claramente la probabilidad de “éxito” no es constante.

La función que permite asignar las probabilidades en el modelo hipergeométrico es:

$$P(x)=\frac{\left({}_{k}C_{x}\right)\left({}_{N-k}C_{n-x}\right)}{{}_{N}C_{n}}$$

Sus parámetros son precisamente, N, n y k. Son los parámetros porque conociendo estos valores se pueden ya calcular probabilidades con el modelo hipergeométrico.

Ejemplo 1. Un juez tiene ante sí 35 actas testimoniales de las cuales sabe que 18 incluyen falso testimonio. Si extrae una muestra de tamaño 10, ¿cuál es la probabilidad de que haya 5 actas con falso testimonio?

Solución:

Los datos del problema nos permiten identificar que:

N=35

n=10

$$\Rightarrow P(5)=\frac{\left({}_{18}C_{5}\right)\left({}_{17}C_{5}\right)}{{}_{35}C_{10}}=0.2888$$

k=18

x=5

3.3.5. Distribución multinomial

Esta distribución de variable aleatoria discreta se aplica en situaciones en las que:

- Se extrae una muestra de N observaciones
- Las observaciones se pueden dividir en k grupos
- En cada extracción la probabilidad (p_k) de que el elemento seleccionado pertenezca a uno de los k diferentes grupos permanece constante

- Se desea determinar la probabilidad de que de los N elementos, x_1 pertenezcan al grupo 1, n_2 al grupo 2 y sucesivamente hasta el grupo k al que deben pertenecer x_k elementos, donde $\sum x_k = N$

Como se puede apreciar, las semejanzas con la distribución binomial son claras, ya que en ambos modelos hay un número finito de ensayos y las probabilidades se mantienen constantes. La diferencia es que en el modelo multinomial la población se divide en k grupos y en el binomial sólo en dos (“éxito” o “fracaso”). En este sentido, se puede decir que la distribución binomial es un caso particular de la multinomial.

La función que permite calcular las probabilidades es:

$$P(x_1, x_2, ..., x_{k-1}) = \frac{N!}{x_1!x_2!x_3!...x_k!} p_1^{x_1} p_2^{x_2} p_3^{x_3} ... p_k^{x_k}$$

donde,

p_k es la probabilidad (constante) de que un elemento cualquiera de la población pertenezca al grupo k, con $\sum p_k = 1$

Sus parámetros son N y un conjunto de k–1 valores de probabilidad. Son los parámetros porque conociendo estos valores se pueden ya calcular probabilidades con el modelo multinomial.

Ejemplo 1. Un perito debe presentar ante una autoridad judicial 14 peritajes. Se sabe por experiencias anteriores que dicha autoridad acepta 40% de los peritajes, desecha 35% y solicita nuevos peritajes en otro 25% de los casos. El perito desea determinar la probabilidad de que le acepten 10 peritajes y le desechen sólo uno.

Solución:

Si designamos como grupo 1 el de los peritajes aceptados y como grupo 2 el de los desechados, los datos del problema nos permiten identificar que:

N=14

p1 = 0.40

p2 = 0.35

p3 = 0.25

$$\Rightarrow P(10,1,4) = \frac{14!}{10!1!4!}(0.40)^{10}(0.35)^{1}(0.25)^{4} = 0.0001$$

x1=10

x2=1

x3=4

3.4 Distribuciones de probabilidad de variables continuas

Para comprender la diferencia entre las variables aleatorias discretas y las continuas recordemos que las variables aleatorias continuas pueden 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.

Esto significa que si X es una variable aleatoria continua, entonces por definición $P(X = x) = 0$, cualquiera que sea el valor de *x*.

Las preguntas de interés tomarán entonces alguna de las siguientes formas básicas:

- $P(X \leq a)$
- $P(X \geq b)$
- $P(c \leq X \leq d)$

donde a, b c y d son números reales

Aquí debe observarse que $P(X \leq a) = P(X < a)$, ya que como se ha hecho notar, $P(X = a) = 0$

Para describir las distribuciones discretas de probabilidades retomamos el concepto de una función de probabilidad f(x). Recordemos que en el caso discreto, 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 recibe el nombre de **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.

Para comprender esto, imagina que se tiene una variable aleatoria continua relativa a un fenómeno que puede repetirse un número muy grande de veces y que los datos se arreglan en una tabla de distribución de frecuencias con la característica especial de que los intervalos se definen de manera que sean muy finos. A continuación se graficarían los datos de la distribución formando en primera instancia un histograma, luego un polígono de frecuencias y de aquí, como paso subsecuente, una curva suavizada. Al tratar de determinar la probabilidad de que la variable tome valores en un intervalo dado se observaría que en el límite, esto es, entre más finos sean los intervalos, tal probabilidad está dada por el área bajo la curva.

La curva suavizada sería la función de densidad de probabilidad, f(x), de modo que el área entre esta curva y el eje X da la probabilidad. Esto lleva a hacer uso del cálculo integral ya que el área está dada por:

$$P(X \leq a) = \int_{-\infty}^{a} f(x)dx$$

donde el símbolo ∫ denota el proceso de integración.

Los valores que se obtienen de P(X ≤ a) para todos los valores posibles a, constituyen la **función de distribución acumulativa** de la variable aleatoria X, misma que se denota como $\mathbf{F_X}$.

Debe ocurrir, para que F_X sea realmente una función de distribución de probabilidades, que $F_X(\infty) = 1$.

En principio parece complicado el manejo de las funciones de distribución de probabilidades en el caso continuo, particularmente si no se manejan las herramientas del cálculo integral. Sin embargo, en muchas situaciones de soluciones informáticas en las que se requieran análisis de datos, las distribuciones de probabilidad pueden ser de gran ayuda, como por ejemplo la distribución normal.

3.4.1. Distribución normal

Esta distribución de probabilidad también es conocida como **“Campana de Gauss”** por la forma que tiene su gráfica y en honor del matemático que la desarrolló. Tal vez dé la impresión de ser un tanto complicada, pero no debes preocuparte por ello, pues para efectos del curso, no es necesario usarla de manera analítica, sino comprender intuitivamente su significado.

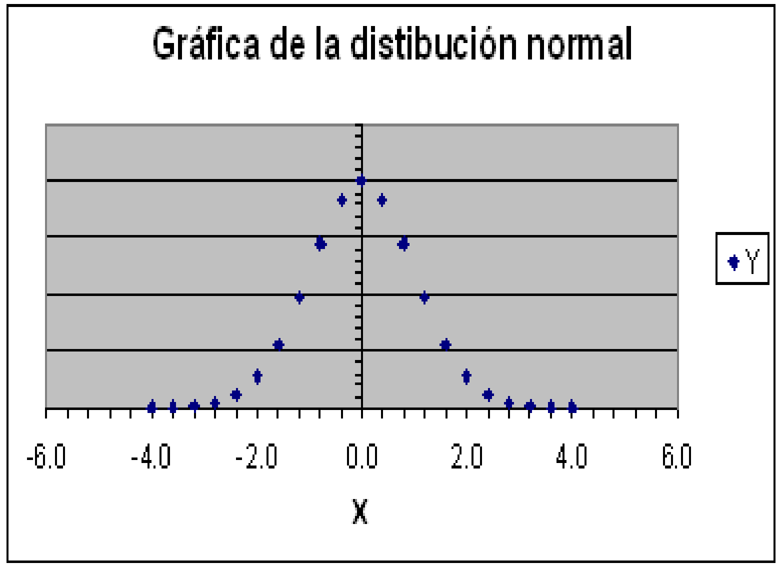
De cualquier manera se dará una breve explicación de la misma para efectos de una mejor comprensión del tema. Su función aparece a continuación.

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

En esta función la “y” es la ordenada de las coordenadas rectangulares cartesianas y representa la altura sobre el eje “x”; x es la abscisa en este sistema de coordenadas; π = 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 μ y σ corresponden a la media y a la desviación estándar.

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), por lo que así como podemos asignar distintos valores a m (la pendiente) y b (la ordenada al origen), para obtener una ecuación particular (p. ej. y=4x+2), de la misma manera podemos sustituir μ y σ 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 distintos valores a “x” (en el rango de –4 a 4, por ejemplo) para calcular los valores de “y”. Una vez que se ha completado la tabla es fácil graficar en el plano cartesiano. 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

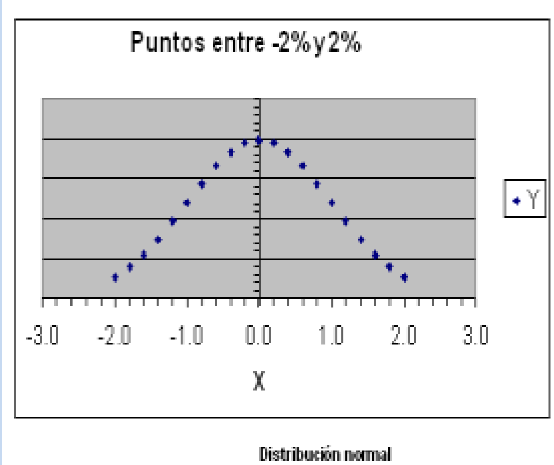


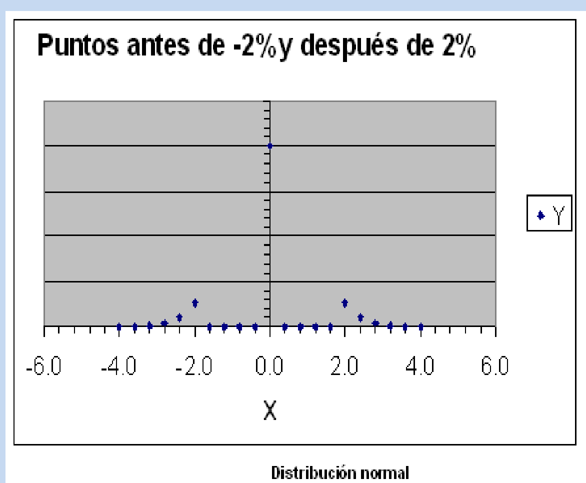
Curva normal

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, 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 cuyo precio fluctúe entre –2% y 2%, que acciones con mayor fluctuación (ver las siguientes gráficas).





Para calcular probabilidades en el caso de la distribución normal se cuenta afortunadamente con valores ya tabulados para el caso en que la distribución tiene una media igual a 0 y una desviación estándar igual a 1. A esta distribución se le conoce como **distribución normal estándar**, y se le denota como **N(0,1)**. En los próximos párrafos aprenderemos a utilizar la tabla de la distribución normal estándar.

La figura "Puntos menores a –2% y 2%", nos muestra el área que hay entre los valores –2 y 2 y además nos enseña que al ser la campana simétrica, tal área es el doble de la que hay entre los valores 0 y 2. En general, el área entre los valores –z y z es el doble de la que hay entre los valores 0 y z. ¿De qué manera nos puede ayudar la tabla a encontrar el valor de tal área?

Al examinar la tabla de la distribución 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 0.34134. 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 0.47500 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.

En el caso en z=2, la tabla nos da un valor de 0.47725, por lo que el área encerrada bajo la curva entre los valores –2 y 2 es 2(0.47725) = 0.9545

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 registrados en su base de datos 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. Di cuantos clientes esperamos que tengan un adeudo entre $7000 y $8,500.

Solución:

Nos percatamos 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:

$$Z = \frac{x_i - \mu}{\sigma}$$

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 base de datos de la empresa del ejemplo anterior deseamos saber qué 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.

Aprovechando la simetría de la curva 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. Indica 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.

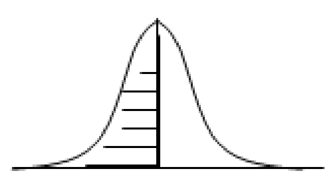
c) 200 metros cúbicos.

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

Solución*:*

a) 160 metros cúbicos.

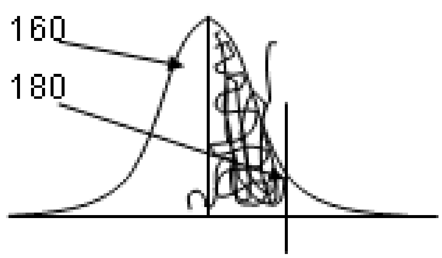
El valor de Z en este caso sería: $z = \frac{160-160}{20} = 0.0$ Esto nos puede desconcertar un poco; sin embargo, nos podemos dar cuenta de que si se surten 160 metros cúbicos el agua alcanzará si el consumo es menor que esa cifra. La media está en 160. Por ello el agua alcanzará en toda el área de 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 $z = \frac{180 - 160}{20} = 1.0$

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, para z=1.0, encontraremos el valor de 0.34134. Sin embargo, debemos agregarle toda la mitad izquierda de la 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%.



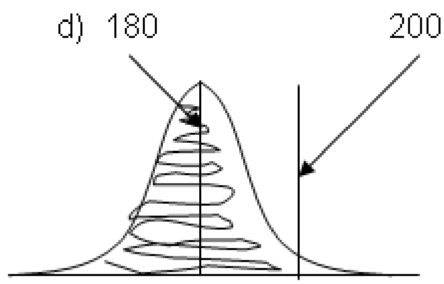
z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621

c) 200 metros cúbicos.

El área buscada se señala en el dibujo. Incluye la primera mitad de la curva y parte de la segunda mitad (la derecha), la 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.

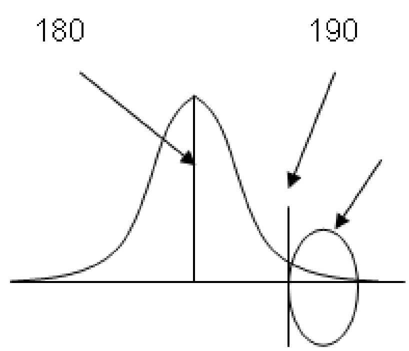
$$Z = \frac{200 - 160}{20} = 2$$

Lo que nos lleva a un valor en tablas 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.



d) Indica 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.



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 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

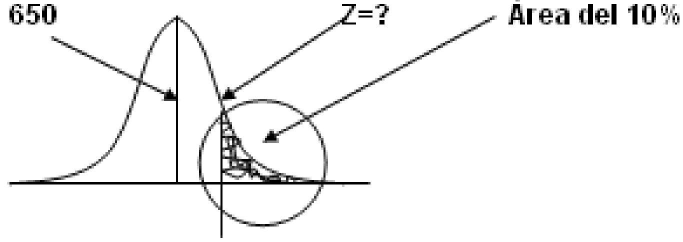
Frecuentemente el problema no es encontrar el área bajo la curva normal mediante el cálculo de Z y el acceso a la tabla para buscar el área ya mencionada. Efectivamente, a veces debemos enfrentar el problema inverso. Conocemos dicha área y deseamos conocer el valor de la variable que lo verifica. El 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. Indica 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 derecha, entonces 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). Éste 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	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	**0.3997**	0.4015

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 se indica a continuación.

Partimos de la relación

$$Z = (X - 650) / 100$$

Observa 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

$$1.28\,(100) = X - 650$$

$$128 + 650 = X$$

$$X = 778$$

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% o 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	0.4452	0.4463	0.4474	0.4485	**0.4495**	0.4505	0.4515	0.4525	0.4535	0.4545

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.

$$1.64 = (X - 650)/100$$

$$1.64\,(100) = X - 650$$

164 + 650 = X

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.

3.4.2. 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 de modo tal que la longitud o tiempo que transcurre entre cada realización del evento es una variable aleatoria, cuya distribución de probabilidades recibe precisamente el nombre de distribución 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: $P(T > t) = e^{-\lambda}$

Donde λ es la tasa promedio de eventos por unidad de tiempo o longitud, según se trate. Como en el caso de la distribución de Poisson, su parámetro es precisamente esta tasa promedio.

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 minuto 1, 5, 10, 15, y 30.

Solución*:*

a)No se reciba ninguna llamada:

Como la tasa promedio está expresada en llamadas por hora y en la pregunta se hace referencia a un periodo de 5 minutos, primero debemos hacer compatibles las unidades. Para ello, establecemos una relación de proporcionalidad directa.

$$15 - 60$$
$$\lambda - \quad 5$$
$$\therefore \quad \lambda = 1.25$$

$$P(T > t) = e^{-\lambda}$$
$$p(t > 5) = 2.71828^{-1.25} = 0.286 = 28.6\%$$

$$e = 2.71828$$

b) Primera llamada en 5 minutos:

$$P(T \le t) = 1 - e^{-\lambda}$$
$$P(T \le 5) = 1 - 2.71828^{-1.25} = 1 - 0.287 = 0.713 = 71.3\%$$

c) Primera llamada en 1, 5, 10, 15 y 30 minutos:

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

3.5. 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**.

Esta ley forma parte de lo que en la probabilidad se conoce como teoremas de límites, uno de los cuales es el teorema de De Moivre-Laplace según el cual, la distribución binomial —que 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— puede aproximarse por la distribución normal si el número de ensayos es suficientemente grande y donde el error en la aproximación disminuye en la medida en que la probabilidad de éxito se acerca a 0.5.

Desde el punto de vista de las operaciones, si lo que deseamos es calcular la probabilidad de que una variable aleatoria binomial con parámetros n y p tome valores entre a y b, entonces debemos:

- Determinar la media y desviación estándar de la variable binomial, esto es, calcular los valores de $\mu = np$, y $\sigma = (npq)^{1/2}$
- Reformular la probabilidad deseada en el contexto binomial por la probabilidad deseada en el contexto de la distribución normal, incorporando una corrección por finitud, esto es, si nuestra pregunta original es determinar el valor de $P(a \leq X \leq b)$ entonces, buscaremos aplicar la distribución normal para calcular:

$$P\left(\left[\frac{a-0.5-np}{\sqrt{npq}}\right] \leq Z \leq \left[\frac{b+0.5-np}{\sqrt{npq}}\right]\right)$$

donde los sumandos 0.5 y –0.5 constituyen la corrección por finitud.

- Emplear la tabla de la distribución normal.

Veamos un ejemplo.

Ejemplo 1. Se arroja una moneda legal 200 veces. Se desea saber la probabilidad de que aparezca sol más de 110 veces pero menos de 130.

Solución:

> El hecho de que la moneda sea legal significa que la probabilidad de que el resultado sea sol es igual a la probabilidad de que salga águila, de modo que tanto la probabilidad de éxito como de fracaso es 0.5, y esta probabilidad no cambia de ensayo a ensayo. Podemos decir entonces que estamos en presencia de un experimento Binomial, de modo que podemos

plantear el problema en los siguientes términos, donde S es la variable aleatoria que denota el número de soles:

P(110 < S < 130) = P(S= 111) + P(S= 112) + P(S= 113) + … + P(S= 129)

$$= \sum_{i=111}^{129} {}_{200}C_i \, (0.5)^i \, (0.5)^{200-i}$$

El problema es que tendríamos dificultades al hacer las operaciones incluso con una calculadora. Es aquí donde resulta útil aplicar la distribución normal como aproximación a la distribución binomial.

Como n=200 y p=0.5, entonces la media es μ = 200(0.5) = 100, en tanto que la varianza es σ^2 = 200(0.5)(0.5) = 50, de modo que la desviación estándar es σ = 7.07.

En consecuencia,

P (111 ≤ X ≤ 129) = P [(110.5 – 100) / 7.07 ≤ Z ≤ (129.5 – 100) / 7.07]

= P (10.5 / 7.07 ≤ Z ≤ 29.5 / 7.07)

= P (1.49 ≤ Z ≤ 4.17)

= 0.5 – 0.4319

= 0.0681

Cuando el número de ensayos es grande pero el valor de la probabilidad de éxito se acerca a cero o a uno, esto es se aleja de 0.5, es mejor emplear la distribución de Poisson como aproximación a la binomial.

DISTRIBUCIÓN NORMAL ESTÁNDAR (ÁREA BAJO LA CURVA)

z	0.00000	0.01000	0.02000	0.03000	0.04000	0.05000	0.06000	0.07000	0.08000	0.09000
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	0.34134	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	0.47725	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	0.49865	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	0.49997	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

RESUMEN

Se define el concepto de variable aleatoria y se señalan sus diferentes tipos. Asimismo, se presentan los rasgos que permiten distinguir algunos modelos de distribución probabilística de variables aleatorias, tipificando los mismos a través de las expresiones analíticas de la función de probabilidad y de densidad, su esperanza matemática, su varianza y sus parámetros. Además, en el caso de la distribución normal se presenta el concepto de distribución normal estándar y se muestra el manejo de las tablas respectivas, así como el uso de esta distribución por cuanto aproximación al modelo binomial.

BIBLIOGRAFÍA

SUGERIDA

Autor	Capítulo	Páginas
1. Anderson, Sweeney, Williams. (2005)	5. Distribuciones discretas de probabilidad. Sección 5.3 Valor esperado y varianza.	184-186
	5.4 Distribución de probabilidad binomial.	189-197
	5.5 Distribución de probabilidad de Poisson.	199-201
	5.6 Distribución de probabilidad hipergeométrica.	203-204
	6. Distribuciones continuas de probabilidad. Sección 6.2 Distribución de probabilidad normal.	218-229

	Sección 6.3 Distribución de probabilidad exponencial.	232-234
2. Berenson, Levine y Krehbiel. (2001)	4. Probabilidad básica y distribuciones de probabilidad. Sección 4.4 Distribución de probabilidad para una variable aleatoria.	179-186
	4.5 Distribución binomial.	186-194
	4.6 Distribución de Poisson.	194-197
	4.7 Distribución normal.	198-219
3. Hernández, Fernández, Baptista. (2006)	6. Formación de hipótesis. Sección: Definición conceptual o constitutiva.	145-146
4. Levin y Rubin. (2004)	5. Distribuciones de probabilidad. Sección 5.1 ¿Qué es una distribución de probabilidad?	178-181
	5.2 Variable aleatoria.	181-187
	5.4 La distribución binomial.	191-202
	5.5 La distribución de Poisson.	202-208
	5.6 La distribución normal: distribución de una variable aleatoria continua.	209-222
	6. Distribuciones discretas de de probabilidad.	181-183

5. Lind, Marchal, Wathen. (2008)	Secciones: ¿Qué es una distribución de probabilidad?	
	Variables aleatorias.	183-185
	Media, varianza y desviación estándar de una distribución de probabilidad.	185-187
	Distribución de probabilidad binomial.	189-199
	Distribución de probabilidad hipergeométrica.	199-203
	Distribución de probabilidad de Poisson.	203-207
	7. Distribuciones de probabilidad continua. Secciones: La familia de distribuciones de probabilidad normal.	227-229
	Distribución de probabilidad normal estándar.	229-233
	Determinación de áreas bajo la curva normal.	233-237

Anderson, David R., Sweeney, Dennis J., Williams, Thomas A. (2005). *Estadística para administración y economía* (8ª. Edición). México: International Thomson Editores, 888 pp. más apéndices.

Berenson, Mark L., David M. Levine, y Timothy C. Krehbiel (2001). *Estadística para administración* (2ª Edición). México: Prentice Hall, 734 pp.

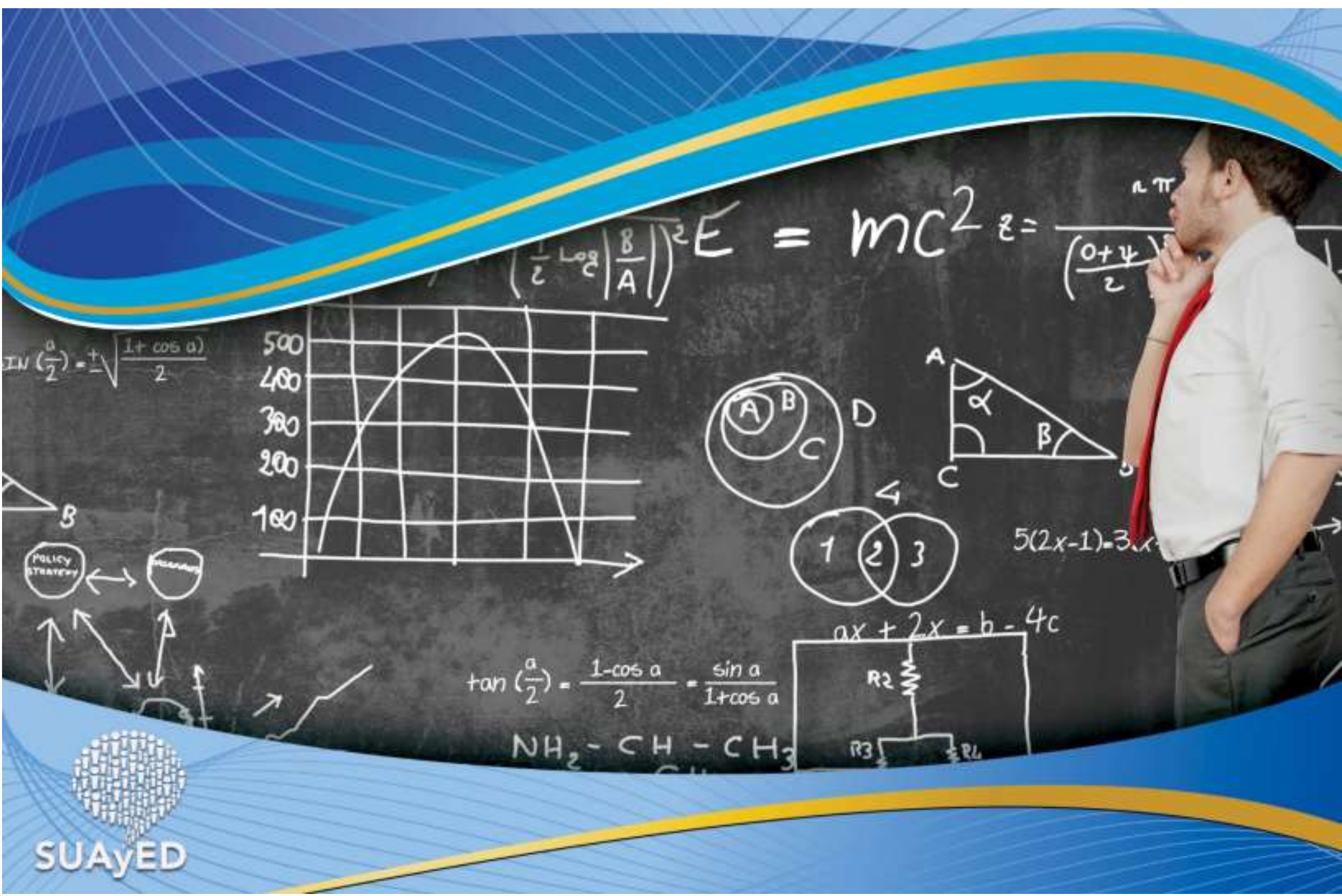
Hernández Sampieri, R., C. Fernández Collado, Lucio P Baptista (2006). *Metodología de la investigación* (4ª edición). México: McGraw-Hill Interamericana, 850 pp.

Levin, Richard I. y David S. Rubin. (2004). *Estadística para administración y economía* (7ª. Edición). México: Pearson Educación Prentice Hall, 826 pp. más anexos.

Lind, Douglas A., Marchal, William G.,Wathen, Samuel, A. (2008). *Estadística aplicada a los negocios y la economía* **(**13ª edición). México: McGraw-Hill Interamericana, 859 pp.

Unidad 4.

Distribuciones muestrales



OBJETIVO PARTICULAR

El alumno identificará e interpretará los diferentes tipos de distribuciones muestrales.

TEMARIO DETALLADO

(8 horas)

4. Distribuciones muestrales

4.1 La distribución muestral de la media
4.2 El teorema central del límite
4.3 La distribución muestral de la proporción
4.4 La distribución muestral de la varianza

INTRODUCCIÓN

La distribución de la población, de la cual extraemos la muestra con la que trabajamos en estadística, es importante para saber qué tipo de distribución debemos aplicar en cada una de las situaciones que se nos presenten en la práctica; en esta unidad veremos algunas de estas distribuciones que se encuentran relacionadas con la distribución normal, además de observar la distribución muestral para la media y para la proporción.

4.1. La distribución muestral de la media

El estudio de determinadas características de una población se efectúa a través de diversas muestras que pueden extraerse de ella.

El muestreo puede hacerse con o sin reposición, y la población de partida puede ser infinita o finita. Una población finita en la que se efectúa muestreo con reposición puede considerarse infinita teóricamente. También, a efectos prácticos, una población muy grande puede considerarse como infinita. En todo nuestro estudio vamos a limitarnos a una población de partida infinita o a muestreo con reposición.

Consideremos todas las posibles muestras de tamaño n en una población. Para cada muestra podemos calcular un *estadístico* (media, desviación típica, proporción) que variará de una a otra. Así obtenemos una distribución del estadístico que se llama distribución muestral.

Las dos medidas fundamentales de esta distribución son la media y la desviación típica, también denominada error típico.

Hay que hacer notar que si el tamaño de la muestra es lo suficientemente grande las distribuciones muestrales son normales y en esto se basarán todos los resultados que alcancemos.

Distribución muestral de medias

Cada muestra de tamaño n que podemos extraer de una población proporciona una media. Si consideramos cada una de estas medias como valores de una variable aleatoria podemos estudiar su distribución que llamaremos distribución muestral de medias.

Si tenemos una población normal N (□, □) y extraemos de ella muestras de tamaño n, la distribución muestral de medias sigue también una distribución normal

$$N(\mu, \frac{\sigma}{\sqrt{n}})$$

Si la población no sigue una distribución normal pero n>30, aplicando el llamado *Teorema central del límite*, la distribución muestral de medias se aproxima también a la normal anterior[1].

---

1 Tomado de: http://recursostic.educacion.es/descartes/web/materiales_didacticos /inferencia _estadística/distrib_muestrales.htm

4.2. El teorema central del límite[2]

El enunciado formal del teorema del límite central es el siguiente: si en cualquier población se seleccionan muestras de un tamaño específico, la distribución muestral de las medias de muestras es aproximadamente una distribución normal. Esta aproximación mejora con muestras de mayor tamaño.

Ésta es una de las conclusiones más útiles en estadística pues nos permite razonar sobre la distribución muestral de las medias de muestras sin contar con información alguna sobre la forma de la distribución original de la que se toma la muestra. En otras palabras, de acuerdo con el teorema del límite central, es válido aproximar la distribución de probabilidad normal a cualquier distribución de valores medios muestrales, siempre y cuando se trate de una muestra suficientemente grande.

El teorema central del límite o teorema del límite central se aplica a la distribución muestral de las medias de muestras que veremos a continuación y permite utilizar la distribución de probabilidad normal para crear **intervalos de confianza** para la media de la población.

[2] Douglas A. Lind, et al. *Estadística para administración y economía*, p. 234.

4.3. La distribución muestral de la proporción

Hoy es bien sabido[3] que si la investigación produce datos mensurables tales como el peso, distancia, tiempo e ingreso, la media muestral es en ocasiones el estadístico más utilizado; pero, si la investigación resulta en artículos “contables” como, por ejemplo, cuántas personas de una muestra escogen la marca “Peñafiel” como su refresco, o cuántas personas de una muestra tienen un horario flexible de trabajo, la proporción muestral es generalmente el mejor estadístico a utilizar.

Mientras que la media se calcula al promediar un conjunto de valores, la “**proporción muestral**” se calcula al dividir la frecuencia con la cual una característica dada se presenta en una muestra entre el número de elementos de la muestra. Es decir:

$$\hat{p} = \frac{x}{n}$$

Donde: x = número de elementos de una muestra que tienen la característica.

n = número de elementos de la muestra.

Ejemplo: suponga que una comercializadora pretende establecer un nuevo centro y desea saber la proporción del consumidor potencial que compraría el principal producto que vende para lo cual realiza un estudio de mercado mediante una

[3] Black, Ken. *Estadística en los negocios*, pp. 241-242.

encuesta a 30 participantes, lo cual permitirá saber quiénes lo comprarían y quiénes no; se obtuvieron los siguientes resultados:

$x_1=1$	$x_7=1$	$x_{13}=0$	$x_{19}=1$	$x_{25}=0$
$x_2=0$	$x_8=0$	$x_{14}=1$	$x_{20}=0$	$x_{26}=0$
$x_3=0$	$x_9=0$	$x_{15}=1$	$x_{21}=1$	$x_{27}=0$
$x_4=0$	$x_{10}=0$	$x_{16}=0$	$x_{22}=1$	$x_{28}=1$
$x_5=0$	$x_{11}=0$	$x_{17}=0$	$x_{23}=1$	$x_{29}=0$
$x_6=1$	$x_{12}=0$	$x_{18}=1$	$x_{24}=0$	$x_{30}=1$

Donde "1" significa que está dispuesto a comprar el producto y "0" no está dispuesto a comprarlo.

En este caso, la proporción de la población (P) que compraría el producto, se puede estimar con $\bar{p}$ (proporción de la muestra que lo compraría), cuyo valor esperado sería $E(\bar{p}) = P$, y el error de $\bar{p}$ al estimar P es:

$$\sigma_{\bar{p}} = \sqrt{\frac{N-n}{N-1}}\sqrt{\frac{P(1-P)}{n}}$$

Si la población es finita, y si la población es infinita o si el muestreo es con reposición, los resultados anteriores se reducen a:

$$\sigma_{\bar{p}} = \sqrt{\frac{P(1-P)}{n}}$$

Es decir, de acuerdo con el teorema del límite central, $\bar{p}$ muestral se comportará como una normal con media P (la verdadera proporción poblacional) y desviación estándar $\sigma_{\bar{p}}$.

En el ejemplo de la comercializadora se tiene que $\bar{p}=\frac{12}{30}=0.40$.

Pero suponiendo que el verdadero parámetro de la población es P=0.30, es decir, que sólo el 30% de la población lo compraría, entonces el promedio $\bar{p}$ estimará a P poblacional pero con un error igual a $\sigma_{\bar{p}}$ que, en este caso, es:

$$\sigma_{\bar{p}} = \sqrt{\frac{0.30(0.70)}{30}} = 0.1195$$

En este caso $\bar{p}$ muestral tendrá distribución normal con media P=0.30 y desviación estándar $\sigma_{\bar{p}}=0.1195$.

Dado que todas las muestras aleatorias que sean tomadas de una misma población en general serán distintas y tendrán por ende diferentes valores para sus estadísticos, tales como la media aritmética o la desviación estándar, entonces resulta importante estudiar la distribución de todos los valores posibles de un estadístico, lo cual significa estudiar las distribuciones muestrales para diferentes estadísticos[4] La importancia de éstas distribuciones muestrales radica en el hecho de que en estadística inferencial, las inferencias sobre poblaciones se hacen utilizando estadísticas muestrales, pues con el análisis de las distribuciones asociadas con éstos estadísticos se da la confiabilidad del estadístico muestral como instrumento para hacer inferencias sobre un parámetro poblacional desconocido.

[4] Weimer, Richard, C. *Estadística*, pp. 353.

4.4. La distribución muestral de la varianza

La varianza de las muestras sigue un proceso distinto a los de la media y proporción. La causa es que el promedio de todas las varianzas de las muestras no coincide con la varianza de la población σ^2. Se queda un poco por debajo, por tanto:

$$E(S_n^2) = \frac{n-1}{n}\sigma^2$$

Comúnmente se utiliza el subíndice $_n$ para recordar que en la varianza se divide entre n. Si deseamos que la media de la varianza coincida con la varianza de la población, tenemos que acudir a la cuasivarianza o varianza insesgada, que es similar a la varianza, pero dividiendo las sumas de cuadrados entre n-1.

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

Su raíz cuadrada es la cuasidesviación típica o desviación estándar. Si se usa esta varianza, si coinciden su media y la varianza de la población, esto indica que la cuasivarianza es un estimador insesgado, y la varianza lo es sesgado[5].

5 Tomado de http://hojamat.es/estadistica/tema7/teoria/teoria7.pdf

RESUMEN

El teorema central del límite es útil para entender que la distribución (las medias de muestras tomadas de una misma población y del mismo tamaño) es aproximadamente normal y que esta aproximación es más precisa a medida que se incrementa el tamaño de la muestra; dando pie al estudio de la distribución muestral para la media y la proporción y, además, a la elaboración de “intervalos de confianza”. La proporción muestral es el mejor estadístico a utilizar cuando en la investigación se trata de averiguar cuestiones tales como: ¿Cuántos integrantes de la población tienen una característica en particular o una tendencia similar?

Con se puede observar, la estadística nos brinda la oportunidad de estudiar el comportamiento de una población por medio de diferentes herramientas, tales como las distribuciones relacionadas con la normal, entre otras; además de diferentes teorías como la del muestreo y la de la estimación estadística, con las que los tomadores de decisiones pueden aunar estos conocimientos a su experiencia, en el medio en el que se estén desenvolviendo, y, en consecuencia, tomar decisiones más certeras en el cada vez más complejo mundo globalizado.

BIBLIOGRAFÍA

SUGERIDA

Autor	Capítulo	Páginas
Berenson y otros (2001).	7	205 - 217
Levin y otros (1996).	6	247 - 261
Christensen (1990).	5	235 - 250
Lind y otros (2004).	8	270 - 281

Berenson, L. Mark; Davis Levine y Timothy Krehbiel (2001). *Estadística para administración* (2ª edición). México: Prentice Hall, 734 pp.

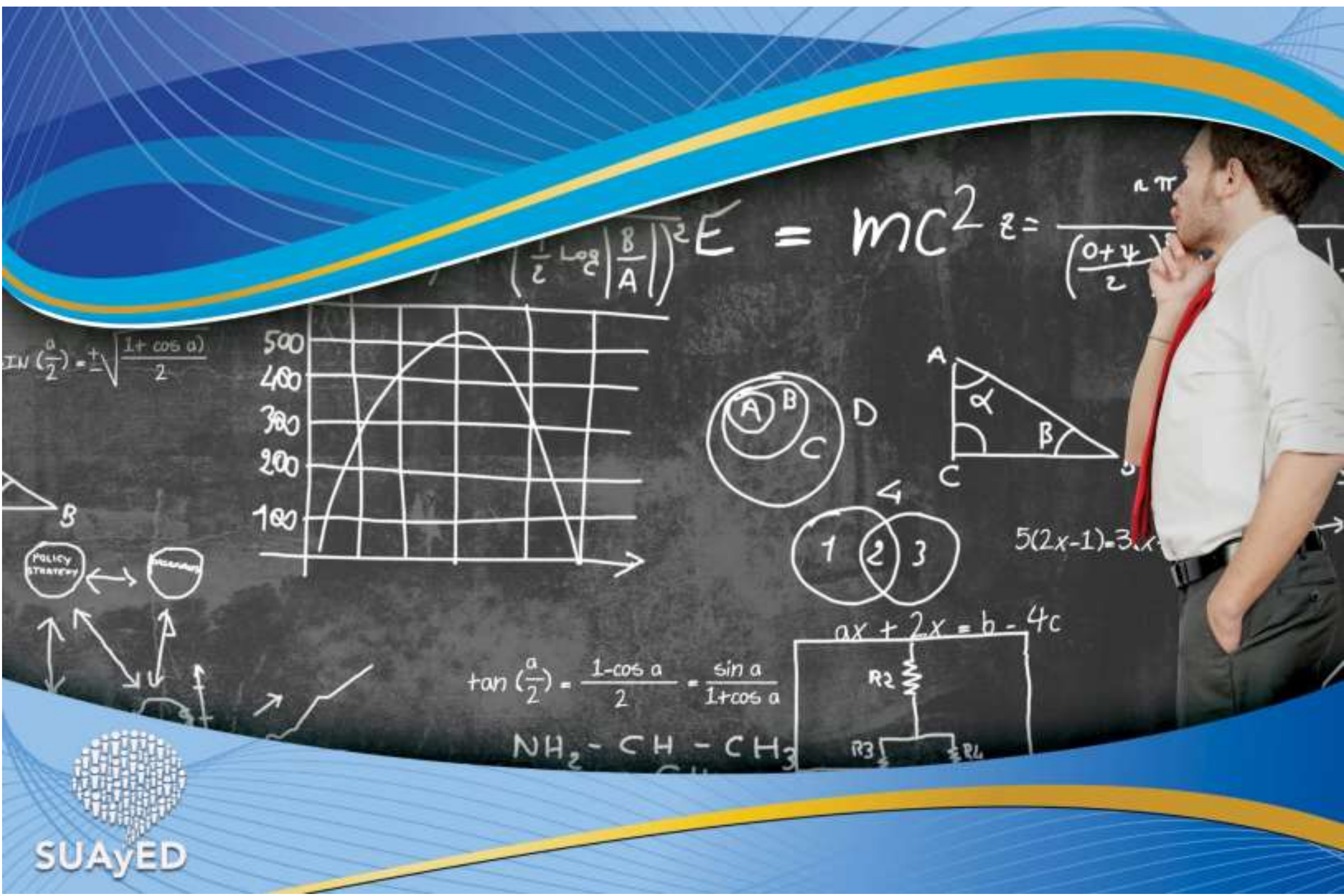
Levin Richard I. y David Rubín (1996). *Estadística para administradores*. México: Alfaomega, 1017 pp.

Christensen H. (1990). *Estadística paso a paso* (2ª ed.). México: Trillas, 682 pp.

Lind A. Douglas, William Marchal y Robert Mason (2004). *Estadística para administración y economía* (11ª edición). México: Alfaomega.

Unidad 5.

Pruebas de hipótesis con la distribución ji cuadrada



OBJETIVO PARTICULAR

El alumno relacionará los conceptos de prueba de hipótesis con la distribución ji cuadrada.

TEMARIO DETALLADO

(8 horas)

5. Pruebas de hipótesis con la distribución ji cuadrada

5.1 La distribución ji cuadrada, χ^2
5.2 Pruebas de hipótesis para la varianza de una población
5.3 Prueba para la diferencia entre n proporciones
5.4 Pruebas de bondad de ajuste a distribuciones teóricas
5.5 Pruebas sobre la independencia entre dos variables
5.6 Pruebas de homogeneidad

INTRODUCCIÓN

En esta unidad, el alumno investigará y analizará el concepto de prueba de hipótesis y lo aplicará sobre varianzas, medias, etc.; ello le permitirá percatarse de la importancia que tienen las pruebas de hipótesis para la toma de decisiones dentro de las empresas.

Actualmente, sabemos que la matemática es una herramienta importante en la toma de decisiones, y la estadística junto con todos sus procesos no es la excepción; así, es importante que el alumno desarrolle todos los conceptos y ejercicios planteados en la presente unidad, enriqueciendo su cultura para su futuro desempeño profesional.

Sabemos que cuando las personas toman decisiones, inevitablemente lo hacen con base en las creencias que tienen en relación al mundo que los rodea; llevan en la mente una cierta imagen de la realidad, piensan que algunas cosas son verdaderas y otras falsas y actúan en consecuencia, así, los ejecutivos de empresas toman todos los días decisiones de importancia crucial porque tienen ciertas creencias tales como:

- De que un tipo de máquina llenadora pone al menos un kilogramo de detergente en una bolsa.
- De que cierto cable de acero tiene una resistencia de 100 kg. o más a la rotura.
- De que la duración promedio de una batería es igual a 500 horas.

- De que en un proceso de elaboración de cápsulas, éstas contengan precisamente 250 miligramos de un medicamento.
- Que la empresa de transportes de nuestra competencia tiene tiempos de entrega más rápidos que la nuestra.
- De que la producción de las plantas del oriente contiene menos unidades defectuosas que las del occidente.

Incluso los estadistas basan su trabajo en creencias tentativas:

- Que dos poblaciones tienen varianzas iguales.
- Que esta población está normalmente distribuida.
- Que estos datos muestrales se derivan de una población uniformemente distribuida.

En todos estos casos y en muchos más, las personas actúan con base en alguna creencia sobre la realidad, la cual quizá llegó al mundo como una simple conjetura, como un poco más que una suposición informada; una proposición adelantada tentativamente como una verdad posible es llamada hipótesis.

Sin embargo, tarde o temprano, toda hipótesis se enfrenta a la evidencia que la comprueba o la rechaza y, en esta forma, la imagen de la realidad cambia de mucha a poca incertidumbre.

Por lo tanto, de una manera sencilla podemos decir que una prueba de hipótesis es un método sistemático de evaluar creencias tentativas sobre la realidad, dicho método requiere de la confrontación de tales creencias con evidencia real y decidir, en vista de esta evidencia, si dichas creencias se pueden conservar como razonables o deben desecharse por insostenibles.

A continuación estudiaremos la forma en que las creencias de las personas pueden ser probadas de manera sistemática.

5.1. La distribución ji cuadrada, χ^2

En ocasiones los investigadores muestran más interés en la varianza poblacional que en la proporción o media poblacionales y las razones llegan desde el campo de la calidad total, donde la importancia en demostrar una disminución continua en la variabilidad de las piezas que la industria de la aviación llega a solicitar es de vital importancia. Por ejemplo, el aterrizaje de un avión depende de una gran cantidad de variables, entre las que encontramos la velocidad y dirección del aire, el peso del avión, la pericia del piloto, la altitud, etc.; si en el caso de la altitud, los altímetros del avión tienen variaciones considerables, entonces podemos esperar con cierta probabilidad un aterrizaje algo abrupto, por lo tanto, la variabilidad de estos altímetros debe mostrar una disminución continua; y qué decir de los motores que impulsan al avión mismo, si las piezas que los conforman son demasiado grandes, el motor puede incluso no poder armarse y si son demasiado pequeñas, entonces los motores tendrán demasiada vibración y en ambos casos las pérdidas de la industria serían cuantiosas.

Así, la relación entre la varianza de la muestra y la varianza de la población está determinada por la distribución ji cuadrada (χ^2) siempre y cuando la población de la cual se toman los valores de la muestra se encuentre normalmente distribuida. Y aquí debemos tener especial cuidado, pues la distribución ji cuadrada es sumamente sensible a la suposición de que la población está normalmente distribuida y, por ejemplo, construir intervalos de confianza para estimar una varianza poblacional, pero puede que los resultados no sean correctos si la población no está normalmente distribuida.

La distribución ji cuadrada (χ^2) es la razón que existe entre la varianza de la muestra (s^2) multiplicada por los grados de libertad y la varianza de la población. Es decir:

$$\chi^2 = \frac{s^2(gl)}{\sigma^2}$$

El término grados de libertad[6] se refiere al número de observaciones independientes para una fuente de variación menos el número de parámetros independientes estimado al calcular la variación.

Para la distribución ji cuadrada (χ^2), los grados de libertad vienen dados por (n – 1), por lo tanto, la fórmula anterior quedaría expresada como:

$$\chi^2 = \frac{s^2(n-1)}{\sigma^2}$$

Donde podemos observar que la variación de la distribución ji cuadrada (χ^2) depende del tamaño de la muestra y de los grados de libertad que posea.

En general y debido a que la distribución ji cuadrada (χ^2) no es simétrica a medida que se incrementa el número de grados de libertad, la curva característica de la distribución se vuelve menos sesgada.

La distribución ji cuadrada (χ^2), es en sí toda una familia de distribuciones por lo que existe una distribución ji cuadrada para cada grado de libertad.

[6] Black, Ken. *Estadística en los negocios*. CECSA, p. 264.

Algebraicamente podemos manipular la fórmula anterior $\chi^2 = \frac{s^2(n-1)}{\sigma^2}$ con el objetivo de que nos sea de utilidad para construir intervalos de confianza para varianzas poblacionales, quedando de la siguiente manera:

$$\frac{s^2(n-1)}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq \frac{s^2(n-1)}{\chi^2_{1-\alpha/2}}$$

Ejemplo:

Supón que una muestra de 7 pernos especiales utilizados en el ensamblado de computadoras portátiles arrojó los siguientes resultados:

2.10 mm; 2.00 mm, 1.90 mm, 1.97 mm, 1.98 mm, 2.01 mm, 2.05 mm

Si quisiéramos una estimación puntual de la varianza de la población, sería suficiente con calcular la varianza de la muestra, de la siguiente manera:

Primero calculamos la media aritmética de los datos utilizando la siguiente fórmula:

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

por lo tanto, sustituyendo datos tenemos que:

$$\bar{X} = \frac{2.10 + 1.90 + 1.98 + 2.05 + 2.00 + 1.97 + 2.01}{7}$$

y al efectuar cálculos el resultado de la media aritmética (redondeado a 2 decimales) es de:

$$\bar{X} = 2.00$$

a continuación elaboramos una tabla para facilitar el cálculo de la varianza de los datos:

I-dato	DATOS	Dato-media	(Dato-media) elevado al cuadrado
I	x_i	$(x_i - \mu)$	$(x_i - \mu)^2$
1	2,10	0,10	0,00972
2	1,90	–0,10	0,01029
3	1,98	–0,02	0,00046
4	2,05	0,05	0,00236
5	2,00	0,00	0,00000
6	1,97	–0,03	0,00099
7	2,01	0,01	0,00007
	14,01	0,01	0,02389

Recordando ahora la fórmula correspondiente a la varianza de una muestra:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

y sustituyendo datos en esta fórmula, podemos ver que el valor obtenido en la esquina inferior derecha de la tabla anterior corresponde a:

$$\sum_{i=1}^{n}(X_i - \bar{X})^2$$

por lo tanto:

$$s^2 = \frac{1}{7-1}(0.02389)$$

de donde al efectuar cálculos vemos que:

$$s^2 = 0.003981$$

Es decir, la varianza de la muestra tiene un valor de: 0.003981, pero si consideramos que el valor de la estimación puntual puede cambiar de una muestra a otra, entonces será mejor construir un intervalo de confianza, para lo cual debemos suponer que la población de los diámetros de los pernos está normalmente distribuida, y como vemos que n=7, entonces los grados de libertad serán: gl= 7–1= 6, si queremos que el intervalo sea de 90% de confianza, entonces el nivel de significancia α será de 0.10, siendo ésta la parte del área bajo la curva de la distribución ji cuadrada que está fuera del intervalo de confianza, esta área es importante porque los valores de la tabla de distribución ji cuadrada están dados de acuerdo con el área de la cola derecha de la distribución. Además, en nuestro caso $\alpha/2$ = 0.05 es decir, 0.05 del área está en la cola derecha y 0.05 está en la cola izquierda de la distribución.

Es importante hacer notar que debido a la forma de curva de la distribución ji cuadrada, el valor para ambas colas será diferente, así, el primer valor que se debe de obtener es el de la cola derecha, mismo que se obtiene al ubicar en el primer renglón de la tabla el valor correspondiente al nivel de significancia, que en este caso es de 0.05 y, posteriormente se ubica en el lugar de las columnas los correspondientes grados de libertad ya calculados, que en este caso es de 6 grados de libertad, por lo tanto el valor de ji cuadrada obtenido es de:

$$\chi^2_{0.05,6} = 12.5916$$

Observa que en la nomenclatura se escribe la denotación de ji cuadrada teniendo como subíndice el nivel de significancia y los grados de libertad y, a continuación, se escribe el valor correspondiente.[7]

El valor de ji cuadrada para la cola izquierda se obtiene al calcular el área que se encuentra a la derecha de la cola izquierda, entonces:

A, a la derecha de la cola izquierda = 1 – 0.05

A, a la derecha de la cola izquierda = 0.95

por lo tanto, el valor de ji cuadrada para la cola izquierda será, utilizando el mismo procedimiento anterior para un área de 0.95 y 6 grados de libertad, de:

$$\chi^2_{0.95,6} = 1.63538$$

incorporando estos valores a la fórmula, tenemos que el intervalo de 90% de confianza para los 7 pernos utilizados en el ensamblado de computadoras portátiles tendrá la forma mostrada a continuación:

[7] El valor se obtuvo utilizando la tabla correspondiente a la ji cuadrada en el libro: *Estadística en los negocios* del autor: Ken Black, pp. 779.

$$\frac{s^2(n-1)}{\chi^2_{\alpha/2}} \le \sigma^2 \le \frac{s^2(n-1)}{\chi^2_{1-\alpha/2}}$$

$$\frac{0.0034122(7-1)}{12.5916} \le \sigma^2 \le \frac{0.0034122(7-1)}{1.63538}$$

$$0.0001625 \le \sigma^2 \le 0.0125189$$

Este intervalo de confianza nos dice que con 90% de confianza, la varianza de la población está entre 0.0001625 y 0.0125189.

La prueba estadística de X^2 para una muestra se emplea frecuentemente como prueba de bondad de ajuste, sin embargo, en un plan experimental, en el que se cuenta con un grupo muestra, con diversas subclases y las mediciones están en escala nominal, resulta muy útil este procedimiento.

La eficacia de la prueba está de acuerdo con el tamaño de la muestra, pues con un grado de libertad, si hay dos subclases, algunos autores consideran que la prueba es insensible, no obstante la información que aporta más de dos categorías es satisfactoria en función de la fórmula:

$$X^2 = \Sigma^{H}_{N=1} \frac{(fo - fe)^2}{fe}$$

Donde:
X^2 = valor estadístico de ji cuadrada.
fo = frecuencia observada.
fe = frecuencia esperada.

La ji cuadrada se utiliza cuando:

- Los datos puntualizan a las escalas nominal u ordinal.
- Se utiliza sólo la frecuencia.
- Las poblaciones son pequeñas.
- Se desconocen los parámetros media, moda, etcétera.
- Los datos son independientes.
- Se quiere contrastar o comparar hipótesis.
- Investigaciones de tipo social —muestras pequeñas no representativas >5.
- Se requiere de establecer el nivel de confianza o significatividad en las diferencias.
- La muestra es seleccionada no probabilísticamente.
- X^2 permite establecer diferencias entre f y se utiliza sólo en escala nominal.
- Población > a 5 y < a 20.

Pasos

1. Arreglar las categorías y las frecuencias observadas.
2. Calcular los valores teóricos esperados para el modelo experimental o tipo de distribución muestral: normal, binomial y de Poisson.
3. Calcular las diferencias de las frecuencias observadas en el experimento con respecto a las frecuencias esperadas.
4. Elevar al cuadrado las diferencias y dividirlas entre los valores esperados de cada categoría.
5. Efectuar la sumatoria de los valores calculados.
6. Calcular los grados de libertad (gl) en función de número de categorías [K]: gl = K – 1.

7. Comparar el estadístico X^2 con los valores de la distribución de ji cuadrada en la tabla.

8. Decidir si se acepta o rechaza la hipótesis $X^2c \geq X^2t$.

5.2. Pruebas de hipótesis para la varianza de una población

En ocasiones, analistas investigan la variabilidad de una población, en lugar de su media o proporción.

Esto es debido a que la uniformidad de la producción muchas veces es crítica en la práctica industrial.

La variabilidad excesiva es el peor enemigo de la alta calidad y la prueba de hipótesis está diseñada para determinar si la varianza de una población es igual a algún valor predeterminado.

La desviación estándar de una colección de datos se usa para describir la variabilidad en esa colección y se puede definir como la diferencia estándar entre los elementos de una colección de datos y su media.

La varianza de un conjunto de datos se define como el cuadrado de su desviación estándar; y la varianza muestral se utiliza para probar la hipótesis nula que se refiere a la variabilidad y es útil para entender el procedimiento de análisis de la varianza.

La hipótesis nula para la prueba de la varianza es que la varianza poblacional es igual a algún valor previamente especificado. Como el aspecto de interés, por lo general si la varianza de la población es mayor que este valor, siempre se aplica una de una cola.

Para probar la hipótesis nula, se toma una muestra aleatoria de elementos de una población que se investiga; y a partir de esos datos, se calcula el estadístico de prueba.

Por ejemplo, si se desea averiguar si la variabilidad de edades en una comunidad local es la misma o mayor que la de todo el Estado. La desviación estándar de las edades del Estado, conocida por un estudio reciente es de 12 años. Tomamos una muestra aleatoria de 25 personas de la comunidad y determinamos sus edades.

5.3. Prueba para la diferencia entre n proporciones

Las pruebas de hipótesis a partir de proporciones se realizan casi en la misma forma utilizada cuando nos referimos a las medias, cuando se cumplen las suposiciones necesarias para cada caso. Pueden utilizarse pruebas unilaterales o bilaterales dependiendo de la situación particular.

La proporción de una población.

Las hipótesis se enuncian de manera similar al caso de la media.

Ho: p = p0

H1: p ¹ p0

Regla de decisión: se determina de acuerdo a la hipótesis alternativa (si es bilateral o unilateral)

En el caso de muestras pequeñas se utiliza la *distribución binomial*. La situación más frecuente es suponer que existen diferencias entre las proporciones de dos poblaciones, para ello suelen enunciarse las hipótesis de forma similar al caso de las medias:

Ho: p1 = p2 Þ p1 – p2 = 0

H1: p1 ¹ p2

En este caso puede la hipótesis alternativa enunciarse unilateralmente.

5.4. Pruebas de bondad de ajuste a distribuciones teóricas

Una hipótesis estadística se definió como una afirmación o conjetura acerca de la distribución f(x,q) de una o más variables aleatorias. Igualmente se planteó que la distribución podía tener uno o más parámetros desconocidos, que denotamos por q y que la hipótesis se relaciona con este parámetro o conjunto de parámetros. En otros casos, se desconoce por completo la forma de la distribución y la hipótesis entonces se relaciona con una distribución específica f(x, q) que podamos asignarle al conjunto de datos de la muestra. El primer problema, relacionado con los parámetros de una distribución conocida o supuesta es el problema que hemos analizado en los párrafos anteriores. Ahora examinaremos el problema de verificar si el conjunto de datos se puede ajustar o afirmar que proviene de una determinada distribución. Las pruebas estadísticas que tratan este problema reciben el nombre general de *pruebas de bondad de ajuste*.

Se analizarán dos pruebas básicas que pueden aplicarse: La *prueba Ji cuadrada* y la prueba de *Smirnov-Kolmogorov*. Ambas pruebas caen en la categoría de lo que en estadística se denominan pruebas de “bondad de ajuste” y miden, como el nombre lo indica, el grado de ajuste que existe entre la distribución obtenida a partir de la muestra y la distribución teórica que se supone debe seguir esa muestra. Ambas pruebas están basadas en la hipótesis nula de que no hay diferencias significativas entre la distribución muestral y la teórica. Ambas pruebas están basadas en las siguientes hipótesis:

H0: $f(x,q) = f_0(x,q)$

H1: $f(x,q)\ ^{1}\ f_0(x,q)$

Donde $f_0(x,q)$ es la distribución que se supone sigue la muestra aleatoria. La hipótesis alternativa siempre se enuncia como si los datos no siguieran la distribución supuesta. Si se desea examinar otra distribución específica, deberá realizarse de nuevo la otra prueba suponiendo que la hipótesis nula es esta nueva distribución. Al especificar la hipótesis nula, el conjunto de parámetros definidos por *q* puede ser conocido o desconocido. En caso de que los parámetros sean desconocidos, es necesario estimarlos mediante alguno de los métodos de estimación analizados con anterioridad.

Para formular la hipótesis nula deberán tenerse en cuenta los siguientes aspectos o criterios:

a) La naturaleza de los datos a analizar. Por ejemplo, si tratamos de investigar la distribución que siguen los tiempos de falla de unos componentes, podríamos pensar en una distribución exponencial, o una distribución gama o una distribución *Weibull*, pero en principio no consideraríamos una distribución normal. Si estamos analizando los caudales de un río en un determinado sitio, podríamos pensar en una distribución logarítmica normal, pero no en una distribución normal.

b) Histograma. La forma que tome el histograma de frecuencia es quizás la mejor indicación del tipo de distribución a considerar[8].

5.5. Pruebas sobre la independencia entre dos variables

Cuando cada individuo de la población en estudio se puede clasificar según dos criterios A y B. El primero, "A", admite posibilidades diferentes y el segundo, "B", la representación de las frecuencias observadas en forma de una matriz "a x b" recibe el nombre de *tabla de contingencia*.

La hipótesis nula a contrastar admite que ambos caracteres, A y B, se presenten de forma independiente en los individuos de la población de la cual se extrae la muestra; siendo la alternativa la dependencia estocástica entre ambos caracteres. La realización de esta prueba requiere el cálculo del estadístico.

8 tomado de http://www.mitecnologico.com/Main/PruebaHipotesisParaProporcion.

El estadístico L se distribuye como una con (a - 1) (b - 1) grados de libertad. El contraste se realiza con un nivel de significación de 5%.

Para estudiar la dependencia entre la práctica de algún deporte y la depresión, se seleccionó una muestra aleatoria simple de 100 jóvenes, con los siguientes resultados:

	Sin depresión	Con depresión	Total
Deportista	38	9	47
No deportista	31	22	53
	69	31	100

L = (38 – 32,43)2/32,43 + (31 – 36,57)2/36,57 + (9 – 14,57)2/14,57 + (22 – 16,43)2/16,43

= 0,9567 + 0,8484 + 2,1293 + 1,8883 = 5,8227

El valor que alcanza el estadístico L es 5,8227. Buscando en la tabla teórica de ji cuadrado para 1 grado de libertad se aprecia Lt = 3,84146 < 5,8227 lo que permite rechazar la hipótesis de independencia de caracteres con un nivel de significación de 5%, admitiendo por tanto que la práctica deportiva disminuye el riesgo de depresión[9].

[9] **Tomado de** tomado de **http://www.mitecnologico.com/Main/PruebaHipotesisParaProporcion**.

5.6. Pruebas de homogeneidad

Se plantea el problema de la existencia de homogeneidad entre r poblaciones, para lo cual se realizan muestras independientes en cada una de ellas. Los datos muestrales vienen clasificados en s clases y sus frecuencias absolutas se presentan en forma de una matriz “r x s”.

El estadístico L se distribuye como una con (r - 1) (s - 1) grados de libertad. El contraste se realiza con un nivel de significación de 5%.

Un estudio sobre caries dental en niños de seis ciudades con diferentes cantidades de flúor en el suministro de agua, ha proporcionado los resultados siguientes:

Comunidad	No. niños sin caries	No. niños con caries	
A	38	87	125
B	8	117	125
C	30	95	125
D	44	81	125
E	64	61	125
F	32	93	125
	216	534	750

L = (38 – 36)2/36 + (8 – 36)2/36 + (30 – 36)2/36 + (44 – 36)2/36 + (64 – 36)2/36 + (32 – 36)2/36 + (87 – 89)2/89 (117 – 89)2/89 + (95 – 89)2/89 + (81 – 89)2/89 + (61 – 89)2/89 + (93 – 89)2/89

L = 0,1111 + 21,7778 + 1,0000 + 1,7778 + 21,7778 + 0,4444 + 0,0449 + 8,8089 + 0,4045 + 0,7191 + 8,8089 + 0,1797

L = 65,85

Se quiere saber si la incidencia de caries infantil es igual en las seis poblaciones.

La propia tabla hace pensar que la incidencia de la enfermedad no es igual en todas las poblaciones; basta observar los datos correspondientes a las comunidades B y E. El contraste arroja un valor del estadístico L de 65,85, lo que lleva a rechazar la hipótesis de homogeneidad y aceptar que el contenido diferente de flúor en el suministro del agua puede ser la causa de la disparidad en el número de niños con caries. El Lt esperado según la tabla de la distribución Ji cuadrada es 11,0705, que es menor a 65,85 (Pérez, 2006).

RESUMEN

En esta unidad se revisó el concepto de prueba de hipótesis aplicado sobre varianzas, medias, etc.; lo que conlleva a concientizar la relevancia de las pruebas de hipótesis en la toma de decisiones de las empresas. Por lo que resulta importante que el alumno de informática enriquezca sus conocimientos en la materia, orientándolos al desarrollo de competencias en su desempeño profesional.

Como se analizó desde el comienzo de la unidad, los seres humanos actuamos con base en alguna creencia sobre la realidad, basadas en muchos casos en conjeturas o en proposiciones adelantadas, en otras palabras en hipótesis, las cuales se comprueban o se rechazan dando certidumbre o incertidumbre a la realidad.

En el desarrollo de la unidad vimos cómo una prueba de hipótesis es un método sistemático de evaluar creencias tentativas sobre la realidad, pues confronta a éstas con evidencia real y, así, determina si son razonables o deben desecharse.

BIBLIOGRAFÍA

SUGERIDA

Autor	Capítulo	Páginas
Levin y otros (1996)	11	447-501
Lind y otros (2004)	11	369-392
Christensen (1990)	9	459-498
Hanke y otros (1997)	9	275-297

Levin, Richard y David Rubin (1996). *Estadística para administradores*. México: Alfaomega, 1017 pp.

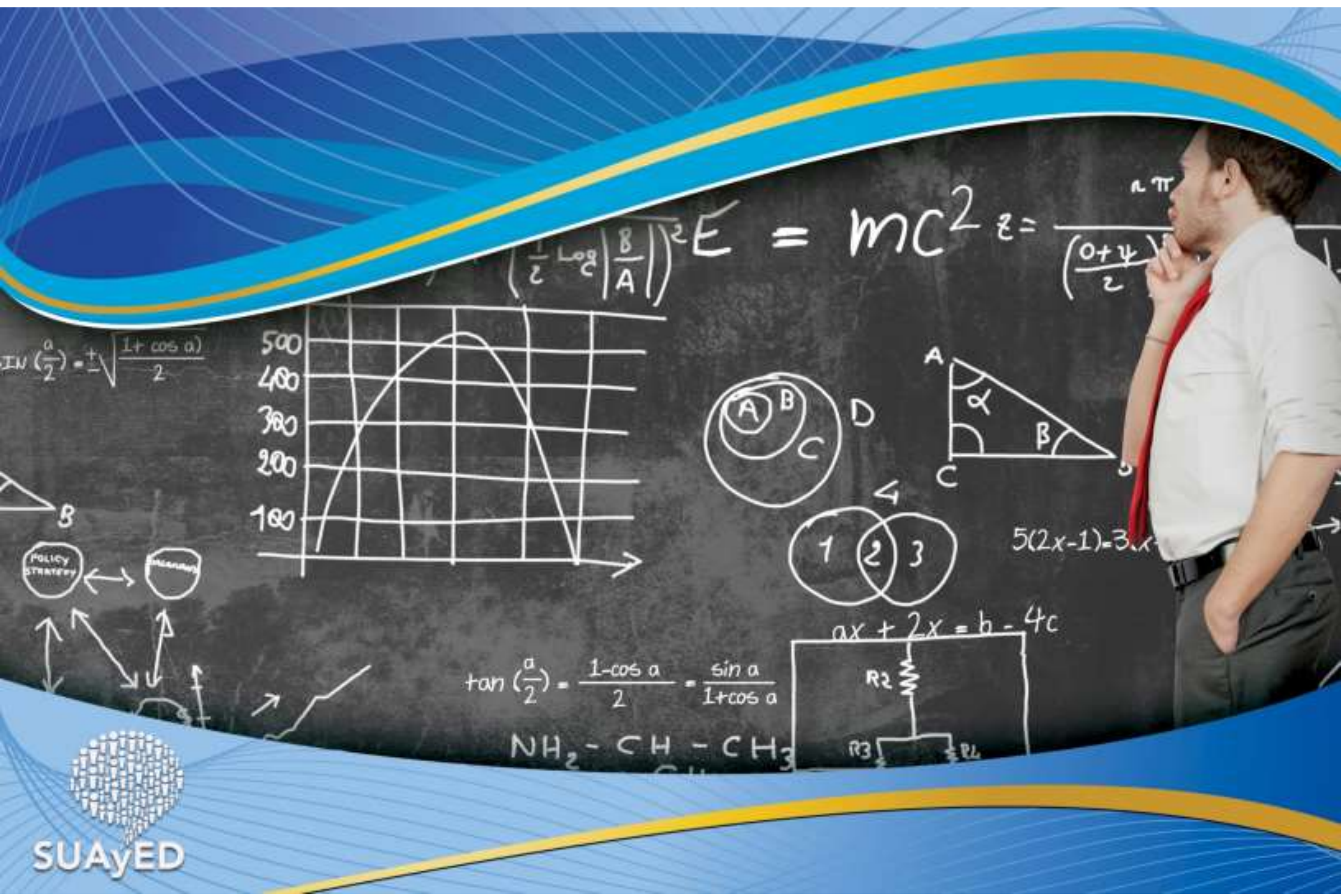
Lind A., Douglas, William Marchal y Robert Mason (2004). *Estadística para administración y economía* (11ª edición). México: Alfaomega.

Christensen, H. (1990). *Estadística paso a paso* (2ª ed.). México: Trillas, 682 pp.

Hanke, Jonh E. y Arthur Reitsch (1997). *Estadística para negocios*, México: Irwin McGraw-Hill, 955 pp.

Unidad 6.

Análisis de regresión lineal simple



OBJETIVO PARTICULAR

El alumno conocerá el método de regresión lineal simple así como su aplicación e interpretación.

TEMARIO DETALLADO

(10 horas)

6. Análisis de regresión lineal simple

6.1 Ecuación y recta de regresión
6.2 El método de mínimos cuadrados
6.3 Determinación de la ecuación de regresión
6.4 El modelo de regresión y sus supuestos
6.5 Inferencias estadísticas sobre la pendiente de la recta de regresión
6.6 Análisis de correlación

INTRODUCCIÓN

El uso de la regresión lineal simple es muy utilizado para observar el tipo de relación que existe entre dos variables y poder ejecutar la toma de decisiones correspondiente, dependiendo de la relación entre dichas variables; así, por ejemplo, pudiere darse el caso en el que después de aplicar la regresión lineal no existiere relación entre las variables involucradas y, en consecuencia, la decisión podría ser buscar cuál es la variable independiente que tiene influencia sobre la dependiente y volver a realizar el estudio completo; pero si fuera el caso en el cual sí existiera una relación positiva entre las variables involucradas, la obtención del coeficiente de correlación nos daría más información sobre el porcentaje de relación existente y podría determinarse si es necesaria la inclusión de otra variable independiente en el problema mismo, para lo cual el análisis de regresión ya sería del tipo múltiple.

6.1. Ecuación y recta de regresión

Observando el diagrama de dispersión, podemos obtener una primera idea de si existe o no relación entre las variables estadísticas. Con el coeficiente de correlación podemos medir la correlación lineal, en caso de existir. Vamos ahora a calcular las líneas que mejor se aproximen a la nube de puntos. A estas líneas se les llama líneas de regresión.

La función que mejor se aproxima a la nube de puntos puede ser lineal, de segundo grado, exponencial, logarítmica. En este tema vamos a calcular únicamente funciones lineales, que vamos a llamar rectas de regresión.

La forma de obtener estas rectas es por el procedimiento conocido como el método de los mínimos cuadrados. Buscamos una recta de ecuación $y=mx+n$ que sea la mejor aproximación. Cada punto x_i de la primera variable tendrá, por una parte, el valor correspondiente a la segunda variable y_i, y, por otra, su imagen por la recta de regresión $y=mx_i+n$. Entre estos dos valores existirá una diferencia $d_i=mx_i+n-y_i$. Vamos a calcular la recta con la condición de que la suma de los cuadrados de todas estas diferencias $\Sigma(mx_i+n-y_i)^2$ sea mínima. Derivando respecto de m y de n y realizando los cálculos matemáticos necesarios, llegamos a la recta de regresión de Y sobre X, que tiene por ecuación en la forma punto-pendiente (**Barrios 2005)**.

$$y - \overline{y} = \frac{\sigma_{xy}}{\sigma_x^2} \cdot (x - \overline{x})$$

6.2. El método de mínimos cuadrados

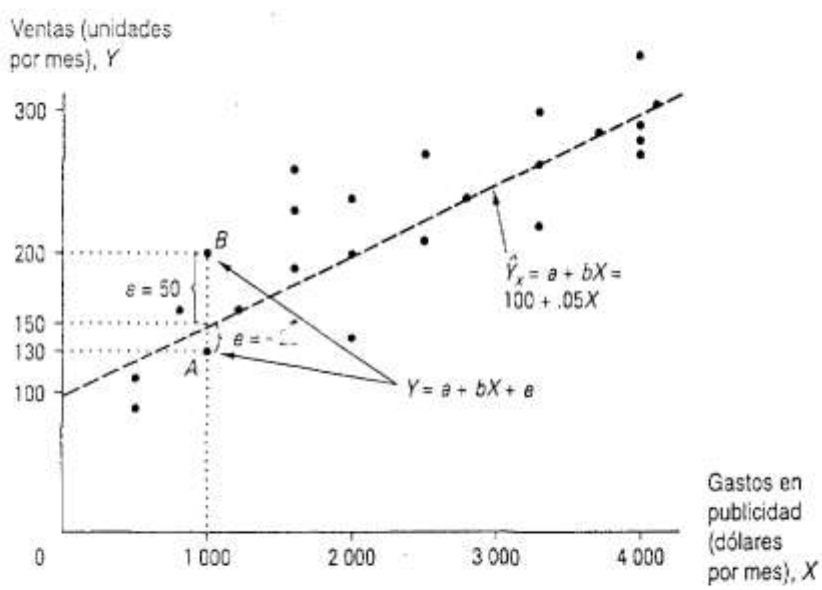
Cualquier método estadístico que busque establecer una ecuación que permita estimar el valor desconocido de una variable, a partir del valor conocido de una o más variables, se denomina análisis de regresión.

El método de mínimos cuadrados, es un procedimiento para encontrar la ecuación de regresión que se origina al estudiar la relación estocástica que existe entre dos variables. Fue Karl Friedrich Gauss (1777-1855) quien propuso el método de los mínimos cuadrados y fue el primero en demostrar que la ecuación estimada de regresión minimiza la suma de cuadrados de errores.

En el análisis de regresión[10], una variable cuyo valor se suponga conocido y que se utilice para explicar o predecir el valor de otra variable de interés se llama variable independiente y se simboliza por "X". Por el contrario, una variable cuyo valor se suponga desconocido y que se explique o prediga con ayuda de otra, se llama variable dependiente y se simboliza por "Y".

[10] Heinz Kohler. *Estadística para negocios y economía,* pp. 528-529.

Una relación estocástica[11] entre dos variables cualesquiera, *X* y *Y*, es imprecisa en el sentido de que muchos valores posibles de "Y" se pueden asociar con cualquier valor de "*X*". Sin embargo, un resumen gráfico de la relación estocástica entre la variable independiente "*X*" y la variable dependiente "*Y*" estará dado por una línea de regresión, que reduce al mínimo los errores cometidos cuando la ecuación de esa línea se utilice para estimar *Y* a partir de *X*.



Gráfica que muestra la relación existente entre los gastos de publicidad y las ventas.

De esta gráfica podemos ver claramente que las ventas dadas en unidades por mes (variable dependiente), en este caso, sí guardan relación con los gastos en publicidad y que dicha relación puede ser denotada por la "recta de regresión".

[11] Heinz Kohler. *Estadística para negocios y economía,* p. 530.

De este análisis de relación estocástica que se da entre dos variables, surgen las ecuaciones que nos provee el método de mínimos cuadrados, que a saber son:

Ecuación de la recta de regresión: $\hat{y}_i = b_0 + b_1 X_i$

En la que:

x_i = es un valor dado de la variable independiente para el cual se quiere estimar el valor correspondiente de la variable dependiente,

b_0 = ordenada al origen de la línea estimada de regresión,

b_1 = pendiente de la línea estimada de regresión,

$\hat{Y}_i$ = valor estimado de la variable dependiente, para el i-ésimo valor de la variable independiente.

Resulta claro que para poder determinar la recta de regresión, es necesario que antes sean calculados los valores correspondientes a la pendiente de la recta y a la ordenada al origen.

La pendiente de la recta de regresión se calcula mediante la siguiente fórmula:

$$b_1 = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n}}{\sum_{i=1}^{n} X_i^2 - \frac{(\sum_{i=1}^{n} X_i)^2}{n}}$$

Y la ordenada al origen se calcula mediante la fórmula:

$$b_0 = \overline{Y} - b_1 \overline{X}$$

Antes de continuar, es necesario advertir que el análisis de regresión no se puede interpretar como un procedimiento para establecer una relación de causa a efecto entre variables. Sólo puede indicar cómo o hasta qué grado las variables están asociadas entre sí. Cualquier conclusión acerca de causa y efecto se debe basar en el juicio del o los individuos con más conocimientos sobre la aplicación. Por ejemplo, un estadista puede llegar a determinar que la relación entre las ventas y el presupuesto asignado a mercadotecnia es positiva y que se tiene un coeficiente de correlación de 0.96, lo cual prácticamente nos indica que es recomendable incrementar el presupuesto al departamento de mercadotecnia para obtener mejores ingresos dentro de la compañía, sin embargo, el director de operaciones puede llegar a determinar que debido a condiciones internas del país en el que se encuentre la empresa, o bien la aparición de una nueva ley que regule los medios utilizados por el mencionado departamento de mercadotecnia, pueden llegar a frenar o, incluso, generar conflictos dentro de la empresa si incrementamos el presupuesto al departamento correspondiente.

6.3. Determinación de la ecuación de regresión

En estadística la **regresión lineal** o **ajuste lineal** es un método matemático que modeliza la relación entre una variable dependiente *Y*, las variables independientes X_i y un término aleatorio ε. Este modelo puede ser expresado como:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon$$

Donde:

$β_0$ es la intersección o término "constante", las $\beta_i \ (i > 0)$ son los parámetros respectivos a cada variable independiente, y *p* es el número de parámetros independientes a tener en cuenta en la regresión. La regresión lineal puede ser contrastada con la regresión no lineal.

6.4. El modelo de regresión y sus supuestos

Con frecuencia, nos encontramos en economía con modelos en los que el comportamiento de una variable, *Y,* se puede explicar a través de una variable *X;* lo que representamos mediante:

$$Y = f\,(X)\ (1)$$

Si consideramos que la relación *f,* que liga *Y* con *X*, es lineal, entonces (1) se puede escribir así:

$$t\ 1\ 2\ t\ Y = β + β\ X\ (2)$$

Como quiera que las relaciones del tipo anterior raramente son exactas, sino que más bien son aproximaciones en las que se han omitido muchas variables de importancia secundaria, debemos incluir un término de perturbación aleatoria, *t u*, que refleja todos los factores —distintos de X— que influyen sobre la variable endógena, pero que ninguno de ellos es relevante individualmente (**Uriel, 2004, p. 1**). Con ello, la relación quedaría de la siguiente forma:

Modelo de regresión simple

Yt = β + β Xt + ut

6.5 Inferencias estadísticas sobre la pendiente de la recta de regresión

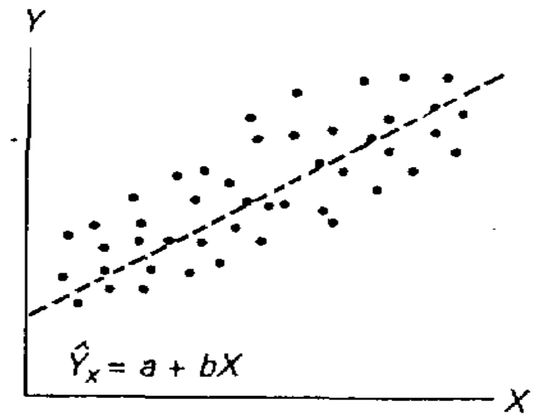
Las inferencias acerca de la pendiente de la recta de regresión son importantes dado que la relación entre las dos variables en cuestión depende de ella precisamente, es decir, si la pendiente de la recta de regresión es positiva, entonces la naturaleza de la relación entre ambas variables será positiva, y la pendiente de la recta es negativa, entonces la relación entre las variables será negativa también, con lo cual podemos iniciar la toma de decisiones dependiendo del contexto del problema mismo. Como se mencionó anteriormente la ecuación de la recta de regresión:

$$\hat{y}_i = b_0 + b_1 X_i$$

b_0 representa la ordenada al origen de la línea estimada de regresión, y

b_1 es la pendiente de la línea estimada de regresión.

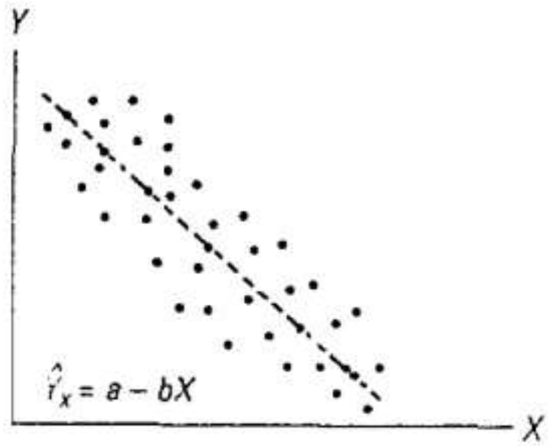
Donde b_0 es en sí, el punto donde la recta corta al eje de las “X” y b_1 nos da el grado de inclinación de la recta, de tal forma que cuando la pendiente de la recta es positiva, se dice que la relación que existe entre las dos variables, dependiente e independiente, es de naturaleza positiva, es decir, que posee una gráfica como la indicada a continuación:



Relación positiva entre dos variables en regresión lineal

En este tipo de relación, los incrementos en los valores de la variable independiente traen como consecuencia un incremento en los valores correspondientes de la variable dependiente y la gráfica tiene como podemos apreciar una forma ascendente.

Pero cuando la pendiente de la recta de regresión es negativa, es decir, que dicha ecuación tuviera la forma $\hat{y}_i = b_0 - b_1 X_i$ entonces la relación existente entre las variables es de tipo negativa, lo cual quiere decir que a incrementos en los valores de la variable independiente, la variable dependiente responde con decrementos; la gráfica resultante tendría la forma siguiente:



Relación negativa entre dos variables en regresión lineal

En esta gráfica podemos observar que la tendencia de la recta de regresión es descendente, lo cual implica, como ya habíamos mencionado, que la relación entre ambas variables es negativa.

6.6. Análisis de correlación

Cuando es necesario resumir aún más los datos (de una gráfica por ejemplo), se utiliza un solo número, que de alguna forma mide la fuerza de asociación entre dos variables como son el ingreso real y el nivel de educación escolar, en nuestro caso. El análisis de correlación nos ayuda a obtener dicho número que se conoce como: coeficiente de correlación. Los valores de coeficiente de correlación siempre están entre –1 y +1; un valor de +1 indica que las dos variables tienen una relación lineal positiva perfecta. Esto es, todos los puntos de datos están en una línea recta con pendiente positiva. Un valor de –1 indica que las variables tienen una relación lineal negativa perfecta, y que todos los puntos de datos están en una recta con pendiente negativa. Los valores del coeficiente de correlación cercanos a cero indican que las variables no tienen relación lineal.[12]

A continuación presentamos la ecuación para calcular el coeficiente de correlación de la muestra. Si ya se ha hecho un análisis de regresión y se ha calculado el coeficiente de determinación, entonces, el coeficiente de correlación se puede calcular como sigue:

$$r = (signodeb_1)\sqrt{r^2}$$

donde b1 es la pendiente de la ecuación de regresión.

De esta fórmula, resulta claro que el signo del coeficiente de correlación es positivo si la ecuación de regresión tiene pendiente positiva (b1 > 0), y negativo si la ecuación de regresión tiene pendiente negativa (b1 < 0).

Por tanto, tendríamos que:

$$r = (signodeb_1)\sqrt{r^2}$$

$$r = +\sqrt{0.9027}$$

$$r = +0.9501$$

[12] Anderson, Sweeney & Williams (1999). *Estadística para administración y economía*, p. 555.

RESUMEN

En esta unidad se revisó el método de regresión lineal simple así como su aplicación e interpretación, la importancia de este método radica en que se utiliza para observar el tipo de relación que existe entre dos variables y poder llevar a cabo la toma de decisiones correspondiente, dependiendo de la relación entre dichas variables. Si fuera el caso en el cual existiera una relación positiva entre las variables involucradas, la obtención del coeficiente de correlación nos daría más información sobre el porcentaje de relación existente y con esto se determinaría si es necesario incluir otra variable independiente en el problema mismo.

BIBLIOGRAFÍA

SUGERIDA

Autor	Capítulo	Páginas
Levin y otros (1996)	12 y 13	509-612
Lind y otros (2004)	13	458-489
Christensen (1990)	10	557-609
Hanke (1997)	14	522-561

Levin, Richard y Rubin David S. (1996). *Estadística para administradores*. México: Alfaomega, 1017 pp.

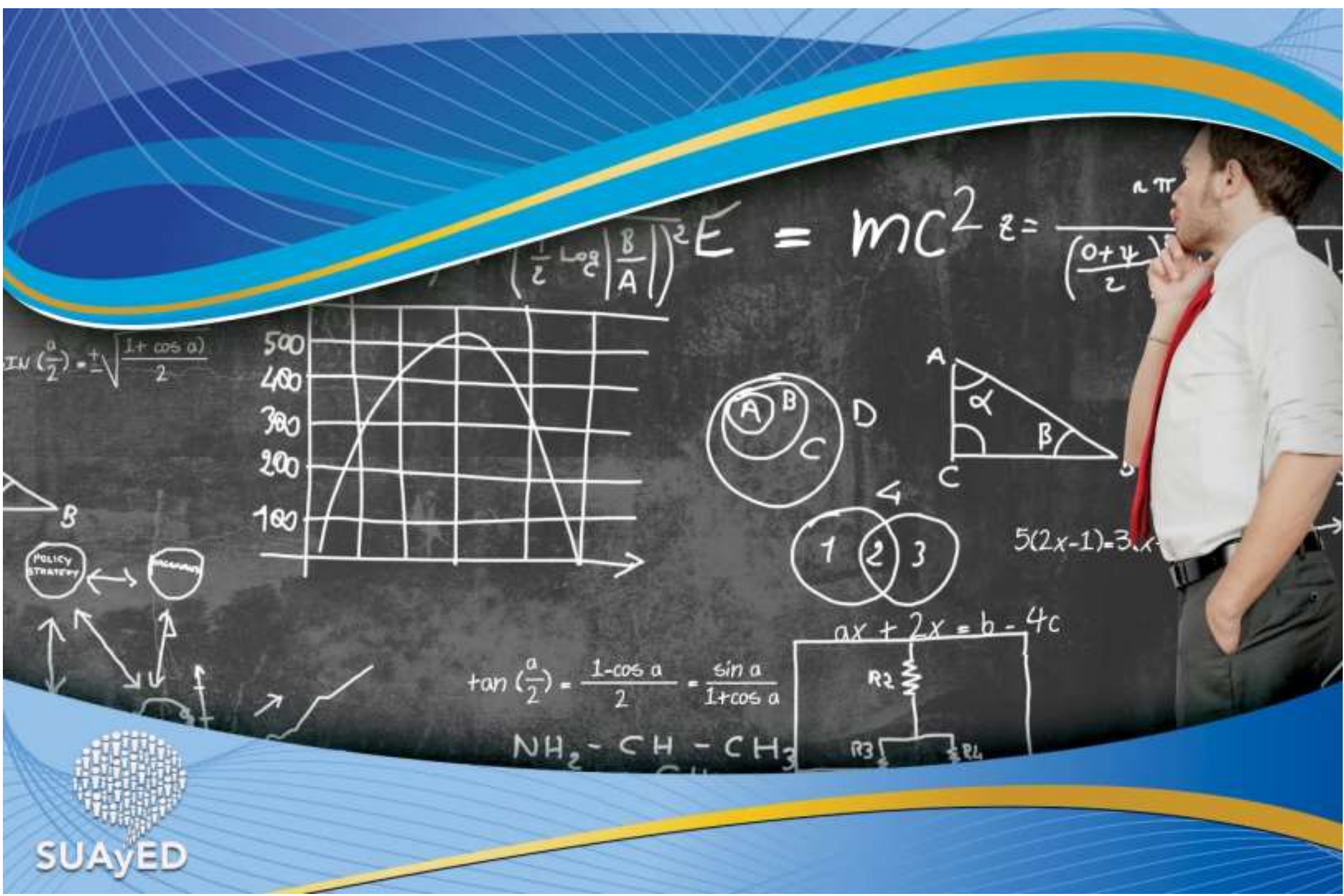
Lind, Douglas, William Marchal y Robert Mason (2004). *Estadística para administración y economía* (11ª ed.). México: Alfaomega.

Christensen, H. (1990). *Estadística paso a paso* (2ª ed.). México: Trillas, 682 pp.

Hanke, Jonh E. y Arthur Reitsch (1997). *Estadística para negocios*. México: Irwin McGraw-Hill, 955 pp.

Unidad 7.

Análisis de series de tiempo



OBJETIVO PARTICULAR

El alumno conocerá el método de regresión lineal simple así como su aplicación e interpretación.

TEMARIO DETALLADO

(8 horas)

7. Análisis de series de tiempo

7.1 Los cuatro componentes de una serie de tiempo
7.2 Análisis gráfico de la tendencia
7.3 Tendencia secular
7.4 Variaciones estacionales
7.5 Variaciones cíclicas
7.6 Fluctuaciones irregulares
7.7 Modelos autorregresivos de promedios móviles

INTRODUCCIÓN

Una serie de tiempo es el conjunto de datos que se registran a través del tiempo sobre el comportamiento de una variable de interés, generalmente los registros se realizan en periodos iguales.

Las series de tiempo resultan especialmente útiles cuando se requiere realizar un pronóstico sobre el comportamiento futuro que puede tener una variable determinada, imaginemos por ejemplo la necesidad de tomar una decisión sobre el comportamiento a futuro de la demanda, el precio y las ventas de un producto, los ingresos en el próximo año, los precios de bienes y servicios, los valores de los energéticos, etc. En todas estas situaciones resulta útil el análisis de las series de tiempo que los representan, bajo la hipótesis de que los factores que han influenciado su comportamiento en el pasado, estarán presentes de manera similar en el futuro. De esta manera, el objetivo principal del conocimiento de las series de tiempo es la identificación de los factores que intervienen y la separación de cada uno de ellos, con el fin de pronosticar cuál será el comportamiento en el futuro.

7.1. Los cuatro componentes de una serie de tiempo

La componente cíclica es la fluctuación que puede observarse ocurre alrededor de la tendencia. Cualquier patrón regular de variaciones arriba o debajo de la recta que representa a la tendencia puede atribuirse a la componente cíclica.

Estacionalidad (E)

La componente estacional muestra un comportamiento regular en los mismos periodos de tiempo, reflejando costumbres o modas que se repiten regularmente dentro del periodo de observación. En la gráfica la estacionalidad quedaría representada por ejemplo por las variaciones semanales en los rendimientos, no visibles por el periodo de información que se está manejando.

Componente irregular (I)

Es la componente que queda después de separar a las otras componentes, es el resultado de factores no explicables que siguen un comportamiento aleatorio, siendo por ello una parte no previsible de la serie.

Ejemplo:

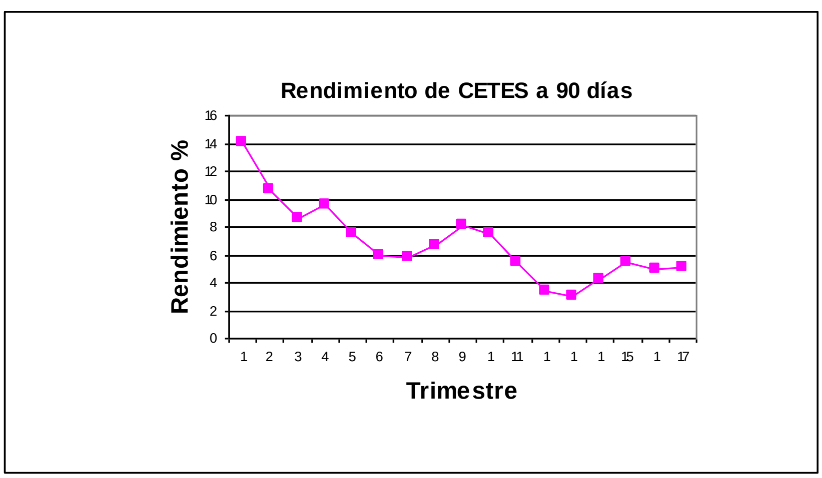
Supongamos que tenemos la siguiente información, correspondiente al comportamiento del rendimiento de los Certificados de la Tesorería, denominados CETES a 90 días, el tiempo está expresado en trimestres y el valor de la variable en valores de la tasa de interés que ganan en cada trimestre.

Trimestre	%
1	14.03
2	10.69
3	8.63
4	9.58
5	7.48
6	5.98
7	5.82
8	6.69
9	8.12
10	7.51
11	5.42
12	3.45
13	3.02
14	4.29
15	5.51
16	5.02
17	5.07

Rendimiento de CETES a 90 días

El registro de rendimientos trimestrales de los CETES representa una serie de tiempo, ya que se han obtenido en periodos sucesivos.

Si se analiza el registro podemos observar que hay una disminución en los valores de rendimiento, de mayor a menor, pero nos resulta difícil afirmar en qué proporción ha ocurrido y de cuánto han sido las variaciones. Si este registro lo analizamos como una serie tendremos la gráfica siguiente:



Rendimiento de los certificados de la tesorería a 90 días.

Utilizando el ejemplo anterior procederemos a descomponer la serie de tiempo en cada uno de sus componentes, lo cual haremos en los siguientes incisos.

La separación de la tendencia, utiliza la metodología de la línea de regresión, hemos mencionado que esta línea puede ser una recta o una curva, en este curso únicamente analizaremos el modelo lineal, por su simpleza y facilidad de cálculo, de esta manera podemos representar a la tendencia por medio de la expresión matemática siguiente:

$$Yt = b_o + b_1X$$

En donde:

Yt tasa de rendimiento calculada

X tiempo, en este caso expresado en trimestres

b_o valor de Y cuando el valor del tiempo es cero

b_1 pendiente de la recta de tendencia

Una vez definido el modelo, se procede a la determinación de los valores de los coeficientes b_o y b_1 de la recta de regresión. En nuestro problema en particular, la ecuación de regresión, que representa a la tendencia del comportamiento de la tasa de rendimiento de los CETES a 90 días aplicando las fórmulas correspondientes para el cálculo primero de “b_1”

$$b_1 = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n}}{\sum_{i=1}^{n} X_i^2 - \frac{(\sum_{i=1}^{n} X_i)^2}{n}}$$

y posteriormente para el cálculo de "b_0"

$$b_0 = \bar{Y} - b_1 \bar{X}$$

es:

Yt = 10.8553676 – 0.44595588 X

Además, aplicando las fórmulas correspondientes primero al cálculo del coeficiente de determinación:

$$r^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{Y})^2}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}$$

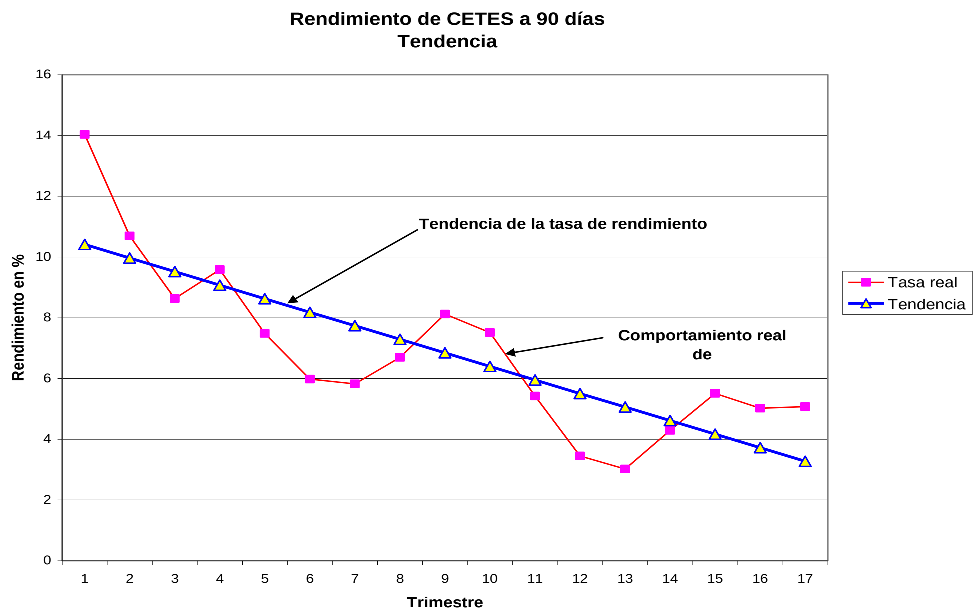
y finalmente al cálculo del coeficiente de correlación:

$$r = (signodeb_1)\sqrt{r^2}$$

Tenemos que el valor del coeficiente de correlación es de r = –0.8078, lo que nos indica que el ajuste logrado con la recta de regresión es adecuado, recordemos que el coeficiente de correlación es una medida de la precisión lograda en el ajuste, valores del coeficiente de correlación iguales a +1 ó –1 son la indicación de un ajuste perfecto, un valor igual a cero nos dirá que éste no existe.

7.2. Análisis gráfico de la tendencia

Una vez definida la ecuación de la recta de tendencia, es posible compararla gráficamente con los valores de la serie, como se muestra en la gráfica siguiente, en ella podemos observar que la tendencia de las tasas de rendimiento es descendente, el signo del coeficiente b1, que representa la pendiente de la recta, ya nos lo había indicado. También podemos observar que son evidentes valores por arriba y por debajo de esta línea, éstos representan a los valores cíclicos de la serie.



Gráfica de comparación de la recta de tendencia contra el comportamiento real de los CETES a 90 días.

En el análisis de tendencias podemos ver clara y rápidamente mediante el cálculo de la pendiente de la recta de regresión (b_1) si la tendencia de la variable de medición (en nuestro caso en particular “el rendimiento de los CETES a 90 días”) es a la baja (pendiente negativa), a la alza (pendiente positiva) o se mantiene sin variación (pendiente cero); lo cual es muy importante dentro del análisis de la serie de tiempo.

7.3. Tendencia secular

Se denomina tendencia secular o simplemente tendencia a la trayectoria temporal de crecimiento, decrecimiento o estabilidad que sigue una serie cronológica a largo plazo. Movimiento unidireccional y persistente que describe la evolución temporal de una determinada variable, una vez depurada de sus variaciones estacionales, cíclicas y accidentales. Para obtener la tendencia secular de una serie temporal se pueden emplear diferentes métodos, como por ejemplo el de las medias móviles o el de los mínimos cuadrados.

7.4. Variaciones estacionales

Método de la razón a la media móvil para determinar la componente estacional en una serie temporal:

1º Se determina la tendencia por el método de las medias centradas en los períodos (Y t) (estamos aplicando cuatro observaciones para el cálculo de la media aritmética).

2º Cómo este método se basa en la hipótesis multiplicativa, si dividimos la serie observada Y t, por su correspondiente media móvil centrada, eliminamos de forma conjunta las componentes del largo plazo (tendencia y ciclo), pero la serie seguirá manteniendo el efecto de la componente estacional.

3º Para eliminar el efecto de la componente estacional, calcularemos las medias aritméticas a nivel de cada estación (cuatrimestre). Estas medias representan de forma aislada la importancia de la componente estacional.

4º Calcularemos los índices de variación estacional, para lo que previamente calcularemos la media aritmética anual de las medias estacionales (M1, M2, M3, M4), que será la base de los índices de variación estacional. Existirán tantos índices como estaciones o medias estacionales tengan las observaciones.

5º Una vez obtenidos los índices de variación estacional puede desestacionalizarse la serie observada, dividiendo cada valor de la correspondiente estación por su correspondiente índice.

Método de la Tendencia por Ajuste Mínimo-Cuadrático. El objetivo sigue siendo aislar la componente estacional de la serie por eliminación sucesiva de todos los demás. La diferencia con el método anterior es que, en este caso, las componentes a l/p (tendencia-ciclo) las obtenemos mediante un ajuste mínimo-cuadrático de las medias aritméticas anuales yt calculándose bajo la hipótesis aditiva.

Sigue los siguientes pasos:

• Se calculan las medias anuales de los datos observados y:

i las observaciones son trimestrales estas medias se obtienen con 4 datos, si son mensuales con 12 datos, etc. para el caso de que el periodo de repetición sea el año.

• Se ajusta una recta por mínimos cuadrados y a b t t = + que nos representa, como sabemos, la tendencia, siendo el coeficiente angular de la recta el incremento medio anual de la tendencia, que influirá de forma distinta al pasar de una estación a otra.

• Se calculan, con los datos observados, las medias estacionales (M1, M2, M3,…) con objeto de eliminar la componente accidental. Estas medias son brutas pues siguen incluyendo los componentes a l/p (tendencia-ciclo) que deben someterse a una corrección.

• Empleando el incremento medio anual dado por el coeficiente, se obtienen las medias estacionales corregidas de las componentes a largo plazo (M’1, M’2, M’3,…) bajo el esquema aditivo.

• Los índices de variación estacional se obtienen con la misma sistemática del método anterior: con las medias estacionales corregidas se obtiene la media aritmética anual M’A que sirve de base para calcular los índices.

• Obtenidos estos índices, podemos desestacionalizar la serie como en el método anterior **(**Ruiz, 2004, §5.4).

7.5. Variaciones cíclicas

Las fluctuaciones de los valores de rendimientos alrededor de la línea de tendencia, constituyen la componente cíclica, éstas son el resultado de la ocurrencia de fenómenos que pueden tener origen social, económico, político, costumbres locales, etc., pero que pueden afectar el comportamiento de la variable, de ahí que su separación resulte importante.

Supongamos ahora que nos interesa conocer la variación que han tenido los rendimientos respecto de la tendencia, es decir, la componente cíclica, la cual queda representada en la gráfica (Gráfica de apreciación de la componente cíclica de los CETES a 90 días) por los valores mayores y menores respecto de la tendencia. Si deseamos conocer el valor numérico de este comportamiento debemos proceder como sigue:

Calcular para cada trimestre el valor del rendimiento de acuerdo con la ecuación de la tendencia (Yt) y compararlo con el correspondiente del registro, estableciendo una proporción entre estos dos valores de la manera siguiente:

$$c = \frac{Y}{Yt}100$$

En donde:

Yrepresenta el rendimiento registrado.

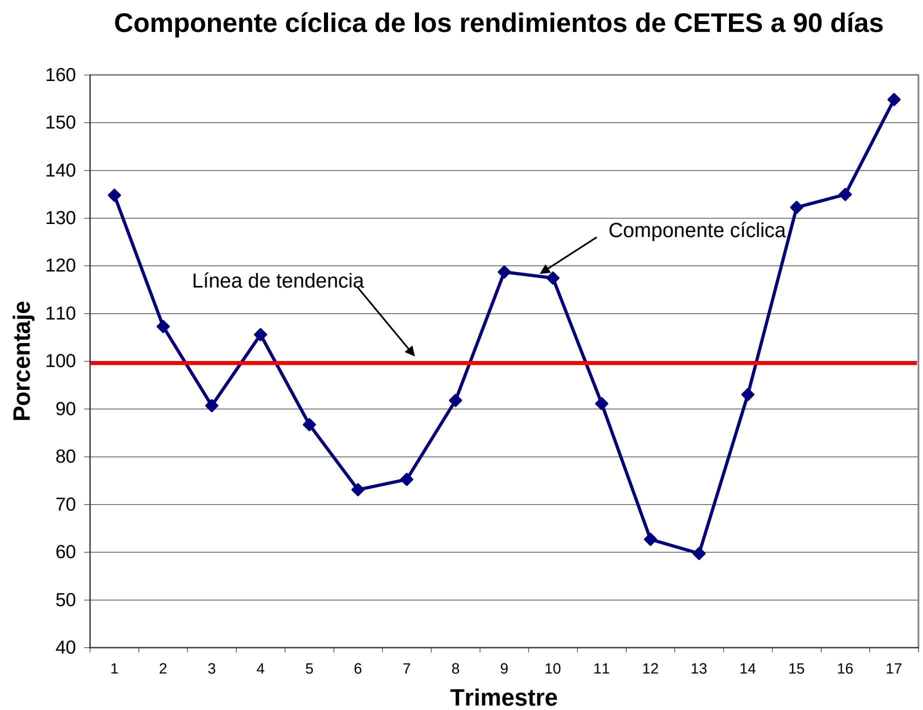
Ytrepresenta el rendimiento calculado con la ecuación de tendencia.

Los valores así calculados se muestran en la tabla siguiente, expresados en porcentaje respecto del valor de la tendencia, los valores que estén por encima de la recta de tendencia alcanzarán un porcentaje superior a cien, mientras que los que se encuentren por debajo de ella tendrán valores inferiores a cien.

Valores de la componente cíclica

	Rendimiento		Componente
Trimestre	Real	Tendencia	cíclica
	Y	Yc	%
1	14.03	10.41	134.78
2	10.69	9.96	107.29
3	8.63	9.52	90.68
4	9.58	9.07	105.60
5	7.48	8.63	86.72
6	5.98	8.18	73.11
7	5.82	7.73	75.26
8	6.69	7.29	91.80
9	8.12	6.84	118.68
10	7.51	6.40	117.42
11	5.42	5.95	91.09
12	3.45	5.50	62.68
13	3.02	5.06	59.71
14	4.29	4.61	93.02
15	5.51	4.17	132.26
16	5.02	3.72	134.94
17	5.07	3.27	154.85

Las componentes cíclicas, pueden ser graficadas para observar los posibles patrones que se presentan, la línea de la tendencia corresponde en la gráfica a la línea del 100%, observemos que la variación cíclica se presenta hacia arriba y hacia abajo de la recta de tendencia.



Es posible ver con mucha claridad cuál ha sido el comportamiento de los rendimientos respecto de la tendencia. Podemos observar que las fluctuaciones a la baja han sido más importantes que las correspondientes a la alza. Esto muy importante, pues si alguna persona compró CETES a 90 días durante el primer trimestre, podemos observar que el rendimiento de estos bajo a continuación y apenas pudieron igualarse los rendimientos alrededor del trimestre 16, presentando una alza alrededor del trimestre 17, lo cual puede representar una pérdida de tiempo y dinero para la persona que bien pudo invertir algunos otros instrumentos que tuvieran mejores rendimientos.

7.6. Fluctuaciones irregulares

Finalmente, una vez separada la componente estacional, procedemos a calcular la componente irregular, lo cual se realiza utilizando nuevamente la ecuación del modelo multiplicativo, relacionándola con el producto de las componentes conocidas hasta ahora, es decir, obteniendo la relación:

$$\frac{(T)(C)(E)(I)}{(T)(C)(E)} = I$$

Los valores obtenidos se expresan en porcentaje, el cálculo de esta componente se muestra en la tabla siguiente:

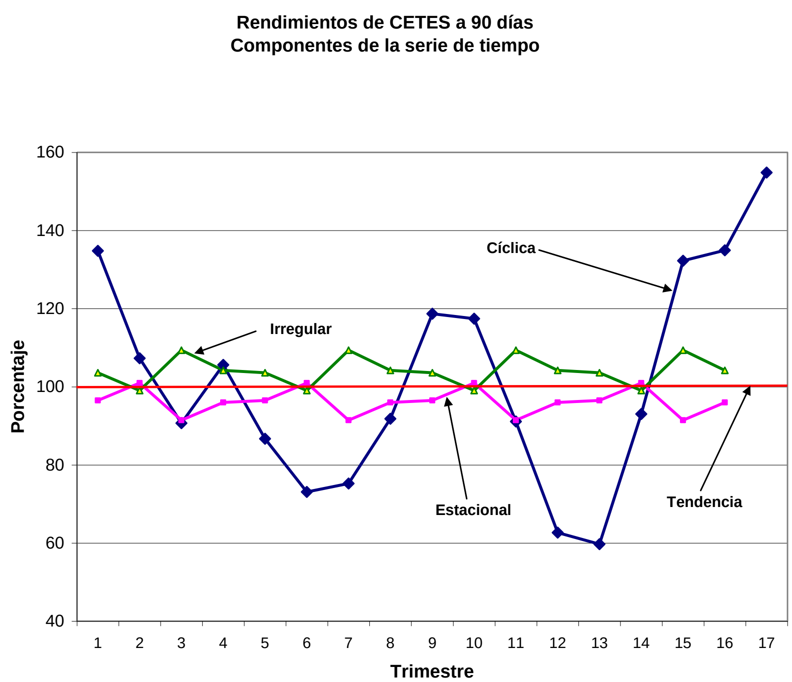
	Rendimiento	Componentes			
Trimestre	Real	tendencia	cíclica	temporal	Irregular
		Yc	C	E	I
1	14.03	10.41	134.78	96.52	103.61
2	10.69	9.96	107.29	100.96	99.05
3	8.63	9.52	90.68	91.46	109.34
4	9.58	9.07	105.60	95.98	104.19
5	7.48	8.63	86.72	96.52	103.61
6	5.98	8.18	73.11	100.96	99.05
7	5.82	7.73	75.26	91.46	109.34
8	6.69	7.29	91.80	95.98	104.19
9	8.12	6.84	118.68	96.52	103.61
10	7.51	6.40	117.42	100.96	99.05

11	5.42	5.95	91.09	91.46	109.34
12	3.45	5.50	62.68	95.98	104.19
13	3.02	5.06	59.71	96.52	103.61
14	4.29	4.61	93.02	100.96	99.05
15	5.51	4.17	132.26	91.46	109.34
16	5.02	3.72	134.94	95.98	104.19
17	5.07	3.27	154.85		

Cuadro 7.5 Cálculo de la componente irregular

En la tabla se presentan los valores de cada una de las componentes, los correspondientes a la cíclica, estacional e irregular se expresan como un porcentaje del valor de la tendencia, la gráfica que relaciona todos los valores se presenta enseguida.

Gráfica de los componentes de la serie de tiempo para nuestro ejemplo del rendimiento de los CETES a 90 días.



Una vez separadas cada una de los componentes es posible conocer la influencia que cada una de ellas tiene sobre el valor del rendimiento, y tomar una decisión sobre las consideraciones que deban realizarse para llevar a cabo una predicción, en este caso deberá analizarse con mucha atención la relación que cada una de ellas haya tenido con los fenómenos económicos y hacer la consideración de las probabilidades que tiene de ocurrir de la misma manera, para considerar o no su participación en la predicción sobre el comportamiento del rendimiento de los CETES.

7.7. Modelos autorregresivos de promedios móviles

Un proceso estocástico { zt } con índice temporal discreto se dice estacionario si las distribuciones conjuntas de probabilidad asociadas con un vector *(zt^,z^2,...,z,k)* son idénticas a las asociadas con el vector *(zl1+h,z ►Z+h,...,z^,^+h)* obtenido por una traslación temporal, y esto para todo conjunto *(tl,t,,...,t^)* de índices, para todo *k* y para todo *h.* Un proceso estacionario tiene todos sus momentos invariantes a cambios en el tiempo. Un proceso se dice "estacionario débil" si sus momentos de primer y segundo orden (esperanzas matemáticas, varianzas, covarianzas) son invariantes a cambios en el tiempo.

RESUMEN

Esta unidad es una introducción básica a los métodos elementales de análisis y pronóstico de series de tiempo; primero se muestra, que para explicar el comportamiento de una serie de tiempo es conveniente suponer que la serie está formada por sus cuatro componentes básicos: tendencia, cíclico, estacional e irregular. Posteriormente separamos cada uno de estos componentes para medir su efecto, con lo cual logramos pronosticar valores futuros de la serie de tiempo.

También se mencionan los métodos de suavizamiento como medio para pronosticar una serie de tiempo que no presenta algunos de sus componentes de manera apreciable. Además, se ejemplifica el uso del análisis de regresión lineal en series de tiempo que sólo tengan una tendencia a largo plazo.

Finalmente, es fácil observar que las series de tiempo son métodos cualitativos de pronóstico que se utilizan cuando se tienen pocos datos históricos o carecemos de ellos. Las series de tiempo también se utilizan cuando se espera que su comportamiento continúe en el futuro.

BIBLIOGRAFÍA

SUGERIDA

Autor	Capítulo	Páginas
Levin y otros (1996)	15	673-712
Christensen (1990)	12	625-643
Lind y otros (2004)	12	602-624
Hanke y otros (1997)	16	668-691

LEVIN Richard I. y Rubin David S. (1996). *Estadística para administradores*. México: Alfaomega, 1017 pp.

CHRISTENSEN H. (1990). *Estadística paso a paso* (2ª edición). México: Trillas, 682 pp.

LIND A. Douglas, Marchal G. William, Mason D. Robert. (2004). *Estadística para Administración y Economía*. México: Alfaomega, 11ª edición.

HANKE Jonh E. y Reitsch Arthur G. (1997). *Estadística para negocios*. México: Irwin McGraw-Hill, 955 pp.

Facultad de Contaduría y Administración
Sistema Universidad Abierta y Educación a Distancia