



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
FACULTAD DE CONTADURÍA Y ADMINISTRACIÓN
DIVISIÓN SISTEMA UNIVERSIDAD ABIERTA Y
EDUCACIÓN A DISTANCIA

LICENCIATURA EN **CONTADURÍA**

APUNTES DIGITALES PLAN 2011

SUAYED UNA OPCIÓN
PARA TI



ESTADÍSTICA I

Plan: 2011

| | |
|---------------------------------|-----------------------------------|
| Plan: 2011 | Créditos: 8 |
| Licenciatura: CONTADURÍA | Semestre: 1º. |
| Área: Matemáticas | Horas. Asesoría: |
| Requisitos: Ninguno | Horas. por semana: 4 |
| Tipo de asignatura: | Obligatoria (X) Optativa () |

AUTORES:

ANTONIO CAMARGO MARTÍNEZ

JORGE GARCÍA CASTRO



INTRODUCCIÓN GENERAL AL MATERIAL DE ESTUDIO

Las modalidades abierta y a distancia (SUAYED) son alternativas que pretenden responder a la demanda creciente de educación superior, sobre todo, de quienes no pueden estudiar en un sistema presencial. Actualmente, “con la incorporación de las nuevas tecnologías de información y comunicación a los sistemas abierto y a distancia, se empieza a fortalecer y consolidar el paradigma educativo de éstas, centrado en el estudiante y su aprendizaje autónomo, para que tenga lugar el diálogo educativo que establece de manera semipresencial (modalidad abierta) o vía Internet (modalidad a distancia) con su asesor y condiscípulos, apoyándose en materiales preparados ex profeso”¹.

Un rasgo fundamental de la educación abierta y a distancia es que no exige presencia diaria. El estudiante SUAYED aprende y organiza sus actividades escolares de acuerdo con su ritmo y necesidades; y suele hacerlo en momentos adicionales a su jornada laboral, por lo que requiere flexibilidad de espacios y tiempos. En consecuencia, debe contar con las habilidades siguientes.

¹ Sandra Rocha, *Documento de Trabajo. Modalidad Abierta y a Distancia en el SUA-FCA*, 2006.



- Saber estudiar, organizando sus metas educativas de manera realista según su disponibilidad de tiempo, y estableciendo una secuencia de objetivos parciales a corto, mediano y largo plazos.
- Mantener la motivación y superar las dificultades inherentes a la licenciatura.
- Asumir su nuevo papel de estudiante y compaginarlo con otros roles familiares o laborales.
- Afrontar los cambios que puedan producirse como consecuencia de las modificaciones de sus actitudes y valores, en la medida que se adentre en las situaciones y oportunidades propias de su nueva situación de estudiante.
- Desarrollar estrategias de aprendizaje independientes para que pueda controlar sus avances.
- Ser autodidacta. Aunque apoyado en asesorías, su aprendizaje es individual y requiere dedicación y estudio. Acompañado en todo momento por su asesor, debe organizar y construir su aprendizaje.
- Administrar el tiempo y distribuirlo adecuadamente entre las tareas cotidianas y el estudio.
- Tener disciplina, perseverancia y orden.
- Ser capaz de tomar decisiones y establecer metas y objetivos.
- Mostrar interés real por la disciplina que se estudia, estar motivado para alcanzar las metas y mantener una actitud dinámica y crítica, pero abierta y flexible.
- Aplicar diversas técnicas de estudio. Atender la retroalimentación del asesor; cultivar al máximo el hábito de lectura; elaborar resúmenes, mapas conceptuales, cuestionarios, cuadros sinópticos, etcétera; presentar trabajos escritos de calidad en contenido, análisis y reflexión; hacer guías de estudio; preparar exámenes; y aprovechar los diversos recursos de la modalidad.



- Además de lo anterior, un estudiante de la modalidad a distancia debe dominar las herramientas tecnológicas. Conocer sus bases y metodología; tener habilidad en la búsqueda de información en bibliotecas virtuales; y manejar el sistema operativo Windows, paquetería, correo electrónico, foros de discusión, chats, blogs, wikis, etcétera.

También se cuenta con materiales didácticos como éste elaborados para el SUAYED, que son la base del estudio independiente. En específico, este documento electrónico ha sido preparado por docentes de la Facultad para cada una de las asignaturas, con bibliografía adicional que te permitirá consultar las fuentes de información originales. El recurso comprende referencias básicas sobre los temas y subtemas de cada unidad de la materia, y te introduce en su aprendizaje, de lo concreto a lo abstracto y de lo sencillo a lo complejo, por medio de ejemplos, ejercicios y casos, u otras actividades que te posibilitarán aplicarlos y vincularlos con la realidad laboral. Es decir, te induce al “saber teórico” y al “saber hacer” de la asignatura, y te encauza a encontrar respuestas a preguntas reflexivas que te formules acerca de los contenidos, su relación con otras disciplinas, utilidad y aplicación en el trabajo. Finalmente, el material te da información suficiente para autoevaluarte sobre el conocimiento básico de la asignatura, motivarte a profundizarlo, ampliarlo con otras fuentes bibliográficas y prepararte adecuadamente para tus exámenes. Su estructura presenta los siguientes apartados.



1. *Información general de la asignatura.* Incluye elementos introductorios como portada, identificación del material, colaboradores, datos oficiales de la asignatura, orientaciones para el estudio, contenido y programa oficial de la asignatura, esquema general de contenido, introducción general a la asignatura y objetivo general.
2. *Desarrollo de cada unidad didáctica.* Cada unidad está conformada por los siguientes elementos.
 - Introducción a la unidad.
 - Objetivo particular de la unidad.
 - Contenidos.
 - Actividades de aprendizaje y/o evaluación. Tienen como propósito contribuir en el proceso enseñanza-aprendizaje facilitando el afianzamiento de los contenidos esenciales. Una función importante de estas actividades es la retroalimentación: el asesor no se limita a valorar el trabajo realizado, sino que además añade comentarios, explicaciones y orientación.
 - Ejercicios y cuestionarios complementarios o de reforzamiento. Su finalidad es consolidar el aprendizaje del estudiante.



- Ejercicios de autoevaluación. Al término de cada unidad hay ejercicios de autoevaluación cuya utilidad, al igual que las actividades de aprendizaje, es afianzar los contenidos principales. También le permiten al estudiante calificarse él mismo cotejando su resultado con las respuestas que vienen al final, y así podrá valorar si ya aprendió lo suficiente para presentar el examen correspondiente. Para que la autoevaluación cumpla su objeto, es importante no adelantarse a revisar las respuestas antes de realizar la autoevaluación; y no reducir su resolución a una mera actividad mental, sino que debe registrarse por escrito, labor que facilita aún más el aprendizaje. Por último, la diferencia entre las actividades de autoevaluación y las de aprendizaje es que éstas, como son corregidas por el asesor, fomentan la creatividad, reflexión y valoración crítica, ya que suponen mayor elaboración y conllevan respuestas abiertas.

3. *Resumen por unidad.*

4. *Glosario de términos.*

5. *Fuentes de consulta básica y complementaria.* Mesografía, Bibliografía, hemerografía y sitios web, considerados tanto en el programa oficial de la asignatura como los sugeridos por los profesores.

Esperamos que este material cumpla con su cometido, te apoye y oriente en el avance de tu aprendizaje.

Recomendaciones (orientación para el estudio independiente)

- Lee cuidadosamente la introducción a la asignatura, en ella se explica la importancia del curso.



- Revisa detenidamente los objetivos de aprendizaje (general y específico por unidad), en donde se te indican los conocimientos y habilidades que deberás adquirir al finalizar el curso.
- Estudia cada tema siguiendo los contenidos y lecturas sugeridos por tu asesor, y desarrolla las actividades de aprendizaje. Así podrás aplicar la teoría y ejercitarás tu capacidad crítica, reflexiva y analítica.
- Al iniciar la lectura de los temas, identifica las ideas, conceptos, argumentos, hechos y conclusiones, esto facilitará la comprensión de los contenidos y la realización de las actividades de aprendizaje.
- Lee de manera atenta los textos y mantén una actitud activa y de diálogo respecto a su contenido. Elabora una síntesis que te ayude a fijar los conceptos esenciales de lo que vas aprendiendo.
- Debido a que la educación abierta y a distancia está sustentada en un principio de auto enseñanza (autodisciplina), es recomendable diseñar desde el inicio un plan de trabajo para puntualizar tiempos, ritmos, horarios, alcance y avance de cada asignatura, y recursos.
- Escribe tus dudas, comentarios u observaciones para aclararlas en la asesoría presencial o a distancia (foro, chat, correo electrónico, etcétera).
- Consulta al asesor sobre cualquier interrogante por mínima que sea.
- Revisa detenidamente el plan de trabajo elaborado por tu asesor y sigue las indicaciones del mismo.

Otras sugerencias de apoyo

- Trata de compartir tus experiencias y comentarios sobre la asignatura con tus compañeros, a fin de formar grupos de estudio presenciales o a distancia (comunidades virtuales de aprendizaje, a través de foros de discusión y correo electrónico, etcétera), y puedan apoyarse entre sí.
- Programa un horario propicio para estudiar, en el que te encuentres menos cansado, ello facilitará tu aprendizaje.



- Dispón de periodos extensos para al estudio, con tiempos breves de descanso por lo menos entre cada hora si lo consideras necesario.
- Busca espacios adecuados donde puedas concentrarte y aprovechar al máximo el tiempo de estudio

TEMARIO OFICIAL

| | |
|-----------------------------------|----|
| 1. Introducción | 4 |
| 2. Estadística descriptiva | 18 |
| 3. Análisis combinatorio | 4 |
| 4. Teoría de la probabilidad | 16 |
| 5. Distribuciones de probabilidad | 18 |
| 6. Números índice | 4 |



INTRODUCCIÓN A LA ASIGNATURA

En esta asignatura el estudiante investigará lo relativo a la estadística descriptiva, la probabilidad y los números índice.

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

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



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

En la **Unidad 4** 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 5** 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 una distribución binomial, de una distribución de Poisson, de una distribución hipergeométrica, de una distribución multinomial, de una distribución normal y de una distribución exponencial. Por último, se enunciará la ley de los grandes números y su interpretación.

La **Unidad 6** está relacionada con diversos tipos de números índice, incluyendo los índices de precios al consumidor, al productor y el de precios y cotizaciones de la bolsa de valores.



Se trata en consecuencia de un curso introductorio a la estadística y la probabilidad, elementos imprescindibles en la toma de decisiones tanto por parte de las organizaciones gubernamentales y privadas como a nivel individual. Su rol ha crecido en importancia, a la par del desarrollo de los equipos de procesamiento de datos, a grado tal que actualmente es difícil encontrar un campo dentro de la investigación científica, las ciencias económico-administrativas y las ciencias sociales en que no tengan aplicación. En particular, los estudiantes de contaduría y administración encontrarán campo fértil para aplicar métodos estadísticos en las áreas de producción, investigación de mercados, auditoría, finanzas, mercado bursátil y desarrollo de sistemas, entre muchas otras.

La estadística y la probabilidad son 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 de manejo de hoja de cálculo.



OBJETIVO GENERAL DE LA ASIGNATURA

Que el alumno comprenda y aplique el proceso estadístico de datos, transformando datos en información útil para sustentar la toma de decisiones.



ESTRUCTURA CONCEPTUAL





UNIDAD 1

INTRODUCCIÓN

APUNTES DIGITALES PLAN 2011





OBJETIVO ESPECÍFICO DE LA UNIDAD

El alumno conocerá el marco teórico integrado por los conceptos básicos relacionados a la estadística descriptiva.



INTRODUCCIÓN

El mundo de los negocios, y en general cualquier actividad humana, se manifiesta fundamentalmente a través de datos de diferentes tipos, los cuales requieren, de acuerdo con su naturaleza, un tratamiento particular. Del correcto manejo de la información depende en gran medida el éxito de una organización, de un negocio, de una investigación científica o social, de un acuerdo comercial así como de una decisión individual. De aquí la importancia de contar con instrumentos que permitan establecer con claridad qué elementos u observaciones se van a considerar , qué atributos se desea conocer de ellos, cómo se les va a medir, qué tratamiento se puede dar a los datos, qué usos se piensa dar a la información generada y cómo puede ésta interpretarse correctamente.



LO QUE SÉ

La palabra estadística viene del latín *statisticum*, que significa la ciencia del Estado. A mediados del siglo XVIII se empieza a utilizar en textos alemanes la palabra *statistik* para referirse a las cifras o números que reflejan la situación del Estado. De ahí surgen los vocablos *statistique*, en francés; *statistics*, en inglés y *estadística* en español. Hoy día, uno de los principales usos de la estadística tiene que ver precisamente con las cifras o números que muestran el estado, la trayectoria y tendencias de la acción gubernamental en diferentes niveles. Por ejemplo, en el Reino Unido los denominados Archivos Nacionales son el archivo oficial para el gobierno central con información relativa a 900 años de historia. Dentro de los archivos se encuentran el Registro Nacional, mismo que durante la segunda guerra mundial permitió tomar decisiones vinculadas al racionamiento de alimentos y al traslado de niños a ciudades del interior de Inglaterra para evitar los bombardeos. También se encuentra el Registro Electoral. La información de estos Archivos Nacionales es la base de muchos estudios de mercado, de opinión, de salud pública y de ingreso-gasto, por mencionar sólo cuatro.

Como es de esperar, en México también hay una oficina que se encarga de recopilar, preparar, procesar y presentar información relativa a muchos aspectos de nuestro país. Tal oficina es el Instituto Nacional de Estadística, Geografía e Informática (INEGI). De acuerdo a tu experiencia describe en un documento dos tipos de información que ofrece esta oficina y sus aplicaciones, discútelas con tu asesor; en caso de que lo consideres necesario puedes visitar la página <http://www.inegi.gob.mx>



TEMARIO DETALLADO

1.1 Generalidades.

1.2 Poblaciones y muestras.



1.1 Generalidades

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

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



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



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

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



1.2 Poblaciones y muestras

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

A continuación se dan algunos ejemplos:

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

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



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

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

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

Limitaciones de tiempo

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

Limitaciones de recursos

El auditor de nuestro segundo ejemplo podría desear estudiar todos los cheques emitidos, pero la empresa auditada no puede



pagar el costo de una revisión tan exhaustiva. Por ello, el auditor debe basar su opinión en una investigación más limitada.

Imposibilidad física

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

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

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

A las características de las poblaciones las denominamos **parámetros** y a las características correspondientes en las muestras las denominamos **estadísticos**. Así, la **media** de la población (a la que conoceremos con la letra griega μ) es un parámetro y la media de la muestra (a la que conoceremos con el símbolo \bar{x}) es un estadístico.

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



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

En cualquier caso, la estadística es una herramienta que nos ayuda a obtener, registrar y procesar datos para generar y analizar información.

La estadística entonces, se divide en 2 tipos: la **estadística descriptiva** y la **estadística inferencial** o inferencia estadística.

Estadística descriptiva

Incluye aquellas técnicas que nos permiten resumir y describir datos. La preparación de tablas, la elaboración de gráficos y las técnicas para el cálculo de los diferentes parámetros de las poblaciones forman parte de las técnicas de la estadística descriptiva. Es en este contexto que adquiere singular importancia que los administradores, contadores e informáticos dominen las técnicas de estadística descriptiva para **resumir y caracterizar** sus **datos** con el objeto de **tomar decisiones** correctas.

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



Estadística inferencial

Comprende un conjunto de técnicas que nos permiten **estimar** (o inferir y de allí su nombre) las **parámetros** de una **población** a partir de una muestra de la misma y con ello tomar decisiones sobre esa población. Estas decisiones incluyen un factor de riesgo, ya que las características de la población **no se infieren con certeza**, lo que hace necesario medir la **probabilidad del error**.

Encontramos un ejemplo de aplicación de la estadística inferencial en las jornadas electorales, ya que en hacia al final de ellas se pronostican los resultados con base en lo que se ha dado en llamar “conteos” rápidos. Estos conteos se realizan registrando los datos de un pequeño conjunto de casillas electorales cuidadosamente seleccionadas. Estos conteos rápidos son un ejemplo de un estudio muestral, es decir, un estudio realizado mediante muestras con el objeto de inferir características de toda la población.

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

En este contexto, el crecimiento y desarrollo histórico de la estadística moderna puede trazarse desde dos fenómenos separados:

- La necesidad del gobierno de recabar datos sobre sus ciudadanos y



- El desarrollo en las matemáticas, de la teoría de probabilidades.

Así por ejemplo, durante las civilizaciones egipcia, griega y romana, los datos se obtenían principalmente con propósitos de impuestos y reclutamiento militar. En la edad media, las instituciones eclesiásticas a menudo mantenían registros de nacimientos, muertes y matrimonios.

En nuestro país, como ya se ha mencionado, el organismo encargado de realizar levantamientos censales es el INEGI.

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

1. Cómo planear la obtención de los datos para que de ellos se puedan extraer conclusiones confiables.
2. Cómo analizar estos datos.
3. Qué tipo de conclusiones pueden obtenerse con los datos disponibles.
- 4.Cuál es la confianza que nos merecen los datos.

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



RESUMEN DE LA UNIDAD

La estadística nos permite establecer líneas de trabajo con los métodos adecuados para observar, medir, recopilar y analizar datos, referidos particularmente a situaciones dónde se generan volúmenes grandes de éstos, así como preparar, presentar e interpretar información. Su metodología se ha desarrollado básicamente en el último siglo y de manera muy rápida, gracias en parte al advenimiento de las computadoras y los sistemas de información.



GLOSARIO DE LA UNIDAD

Estadística

Rama de las matemáticas que se encarga del desarrollo de los procedimientos y métodos científicos - por lo tanto verificables-, para la recopilación, organización, presentación y análisis de datos así como de su transformación en información e interpretación de la misma.

Estadística descriptiva

Rama de la estadística que se encarga del desarrollo y aplicación de procedimientos para presentar y describir colecciones de datos a través de indicadores numéricos y gráficos

Estadística inferencial

Rama de la estadística que tiene como objetivo aplicar los resultados de estudios de una muestra a la población y emitir juicios o conclusiones sobre esa población en general.



Estimación

Es un valor estadístico obtenido a partir de la muestra.

Muestra

Una muestra es una colección de objetos obtenida de un conjunto mayor denominado población con el propósito de que represente a ésta.

Parámetro

Es una propiedad o cualidad de la población que puede expresarse numéricamente.

Población

Colección de elementos con características medibles comunes que los ubican o definen como miembros de una totalidad



ACTIVIDADES DE APRENDIZAJE

ACTIVIDAD 1

Consulta el Capítulo I: Historia de la estadística del Manual de Estadística de David Ruiz Muñoz publicado por la Universidad Pablo de Olavide <http://www.eumed.net/coursecon/libreria/drm/1a.htm> (ANEXO 1).

Elabora un resumen de un máximo de una cuartilla.



ACTIVIDAD 2

Lee los siguientes documentos:

Los Censos y los Impuestos, publicado en Historia clásica. La nueva forma de entender la historia, por Ramiro Sánchez-Crespo. ANEXO 2.

(<http://www.historiaclasica.com/2007/11/los-censos-y-los-impuestos.html>)

Cambio en la Historia de los Censos de América, publicado por el Departamento Administrativo Nacional de Estadística –DANE. ANEXO 3

(http://www.dane.gov.co/files/comunicados/cp_censosamer.pdf)

Antecedentes, publicado por el INEGI.

(<http://www.inegi.org.mx/inegi/default.aspx>)

Elabora, en media cuartilla, una nota relativa a la importancia de los censos, sus aplicaciones y el futuro de los mismos.

ACTIVIDAD 3

Estudia la sección **15.1 Inferencia**, del capítulo 15 La inferencia científica, del texto La investigación científica, de Mario Bunge, publicado por la editorial Ariel.

Elabora un cuadro sinóptico que muestre los distintos tipos de inferencia y sus características principales.



CUESTIONARIO DE REFORZAMIENTO

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



LO QUE APRENDÌ

Construye un **wiki** para anotar tus datos personales en el formato que se presenta.

| Nombre | Edad | Sexo | Estado Civil | Entidad Nacimiento | Número Hermanos | Años Interrupción Estudios |
|--------|------|------|--------------|--------------------|-----------------|----------------------------|
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |

Verifica que los datos de todos tus compañeros estén en la tabla para responder lo siguiente:

- a) Menciona cinco aplicaciones o preguntas que consideres importantes desde un punto de vista descriptivo.
- b) Menciona cinco aplicaciones o preguntas que consideres importantes desde un punto de vista inferencial.



EXAMEN DE AUTOEVALUACIÓN

I. A continuación te presentamos un cuadro donde se anotaron los datos de 10 personas que trabajan en una oficina.

| Nombre | Ingreso | Estado Civil | Número de hijos |
|-----------|---------|--------------|-----------------|
| Juan | 3500 | Casado | 2 |
| José | 2300 | Soltero | 1 |
| Alicia | 4700 | Casada | 4 |
| Mario | 6100 | Soltero | 0 |
| Mónica | 3200 | Soltera | 0 |
| Rebeca | 1700 | Soltera | 0 |
| Ramiro | 2000 | Casado | 3 |
| Josefina | 1800 | Soltera | 1 |
| Amira | 3100 | Casada | 2 |
| Prudencio | 4200 | Casado | 0 |

El jefe de ellos quiere saber si es verdad que los casados tienen en promedio un ingreso mayor que los solteros. También desea saber si en promedio los hombres casados tienen más hijos que las mujeres casadas. Ayúdalo a contestar estas preguntas.

Te recomendamos ordenar y clasificar los datos. Para ello te presentamos una tabla con las celdas en blanco, misma que puedes imprimir para trabajarla en papel.



II. Realiza la lectura del artículo “**Diferencia entre Dato, Información y Conocimiento**” publicado por Gestión del conocimiento en www.gestiondelconocimiento.com/conceptos_diferenciaentredato.htm.

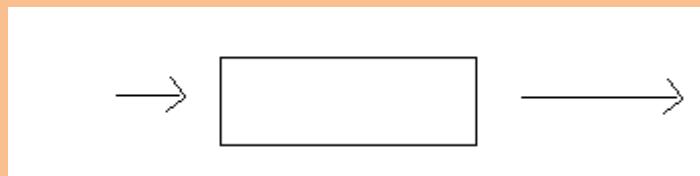
Construye el esquema sobre el rol de la estadística en el proceso de construcción de conocimiento.

Arrastra los conceptos al diagrama presentado, recuerda que sólo tendrás un intento para realizar esta actividad.

Dato
Sistema

Estadística
Interpretación

Información
Intuitivo





III. Revisa las siguientes definiciones, anota la letra que corresponda al concepto.

- 1. En términos generales, es el paso de las premisas a las conclusiones.
- 2. Se le puede ver como proceso o como stock que requiere de una serie de operaciones mentales como lo son comparar, conectar y comunicar.
- 3. Se obtiene asociando hechos en un contexto determinado.
- 4. Se les considera como hechos y cifras “puras” que pueden ser objeto de diversos procesos para adquirir significado.

- A** Conocimiento
- B** Inferencia
- C** Información
- D** Muestra
- E** Parámetro
- F** Datos
- G** Descriptiva



- 5. Conjunto de observaciones con sus respectivos atributos obtenido de un colectivo mayor.
- 6. Rama de la estadística que busca ofrecer una imagen de una situación o fenómeno dado.
- 7. Rasgo importante que caracteriza a una población.



MESOGRAFÍA

BIBLIOGRAFÍA RECOMENDADA

| Autor | Capítulo | Páginas |
|---|---|---------|
| 1, Berenson , Levine, y Krehbiel. | 1. Introducción y recopilación de datos, 1.1 ¿Por qué un administrador necesita estadística? | 2 |
| | 1.2 Crecimiento y desarrollo de la estadística moderna. | 2-3 |
| | 1.3 Pensamiento estadístico y administración moderna. | 4 |
| | 1.4 Estadística descriptiva versus inferencia estadística. | 5-6 |
| | 1.5 ¿Por qué se necesitan datos? | 6-7 |



| | | |
|---------------------------|--|-----------|
| 2. Bunge, Mario. | 15. La inferencia científica 15.1, Inferencia. | 712 - 718 |
| 3. Levin y Rubin. | 1. Introducción. 1.1 ¿Por qué hay que tomar este curso y quién utiliza la estadística? | 2-3. |
| | 2. Agrupación y presentación de datos para expresar significados: tablas y gráficas. Sección: 2.1, ¿Cómo podemos ordenar los datos? | 8-11 |
| 4. Lind, Marchal, Wathen. | 1. ¿Qué es la estadística? ¿Qué se entiende por estadística? | 4-5 |
| | 1. ¿Qué es la estadística? Tipos de estadística. | 6-8 |



BIBLIOGRAFÍA BÁSICA

1. Berenson, Mark L., David M. Levine, y Timothy C. Krehbiel, (2001), *Estadística para administración*, 2ª edición, México, Prentice Hall, 734 pp.
2. Bunge, Mario, (2000), *La investigación científica*, México, Siglo XXI. 805 pp.
3. Levin, Richard I. y David S Rubin, (2004), *Estadística para administración y economía*, 7a. Edición, México, Pearson Educación Prentice Hall, 826 pp.
4. 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.



BIBLIOGRAFÍA COMPLEMENTARÍA

1. Bowerman Bruce, *Pronósticos, series de tiempo y regresión; un enfoque aplicado*, México: Cengage Learning, 4ª edición, 2007, 720 pp.
2. Mendenhall William, *Introducción a la probabilidad y estadística*, México: Cengage Learning, 13ª edición, 2010, 776 pp.
3. Webster Allen L., *Estadística I aplicada a los negocios y la economía*, México: McGraw-Hill, 2ª. edición, 2002, 154 pp.



SITIOS ELECTRÓNICOS

- *Historia clásica. La nueva forma de entender la historia, en el blog:*
<http://www.historiaclasica.com/2007/11/los-censos-y-los-impuestos.html>
- Ruiz Muñoz, David, *Historia de la estadística*, capítulo 1 del hipertexto Manual de Estadística, editado por la Enciclopedia y biblioteca virtual de las ciencias sociales, económicas y jurídicas, EUMEDNET, 2004, en el sitio:
<http://www.eumed.net/cursecon/libreria/drm/cap1.pdf/>
- INEGI, *Programa de censos nacionales*, México, en el sitio:
http://www.inegi.org.mx/est/contenidos/espanol/proyectos/metadatos/censos/pcn_10.asp?s=est&c=1434
- Diferencia entre dato, información y conocimiento, publicado en Gestión del conocimiento por la Fundación Iberoamericana del Conocimiento (Barcelona, España), en el sitio:
http://www.gestiondelconocimiento.com/conceptos_diferenciaentre dato.htm



UNIDAD 2

ESTADISTICA DESCRIPTIVA

APUNTES DIGITALES PLAN 2011





SUAYED
UNA OPCIÓN
PARA TI

OBJETIVO ESPECÍFICO DE LA UNIDAD

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



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 dato



LO QUE SÉ

*Hemos caído bajo el embrujo de los números
We have fallen under the spell of numbers. H. G. Wells.*



Escribe un párrafo en el que manifiestes cómo te sientes en un mundo de números.

Para enviar tu respuesta, pulsa el botón **Editar mi envío**; se mostrará un editor de texto en el cual puedes redactar tu información; una vez que hayas concluido, salva tu actividad pulsando el botón **Guardar cambios**.



SUAYED
UNA OPCIÓN
PARA TI

TEMARIO DETALLADO

- 2.1. Tabulación de datos.
- 2.2. Distribuciones de frecuencia.
- 2.3. Presentación gráfica de datos.
- 2.4. Medidas de tendencia central.
- 2.5. Medidas de dispersión.
- 2.6. Teorema de Tchebysheff y regla empírica.



2.1 Tabulación de datos

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



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

Son aquellas en las que el **cero es convencional o arbitrario.**

Un ejemplo de este tipo de escalas es la de los grados Celsius o centígrados que se usan para medir la temperatura. En ella el cero es el punto de congelación del agua y, sin embargo, existen temperaturas más frías que se miden mediante números negativos. En 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.

Escalas de razón

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

Tal es el caso de los grados Kelvin, para medir temperaturas, o algunas otras medidas que utilizamos en nuestra vida cotidiana. Encontramos un ejemplo de esta escala cuando medimos la estatura 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**.

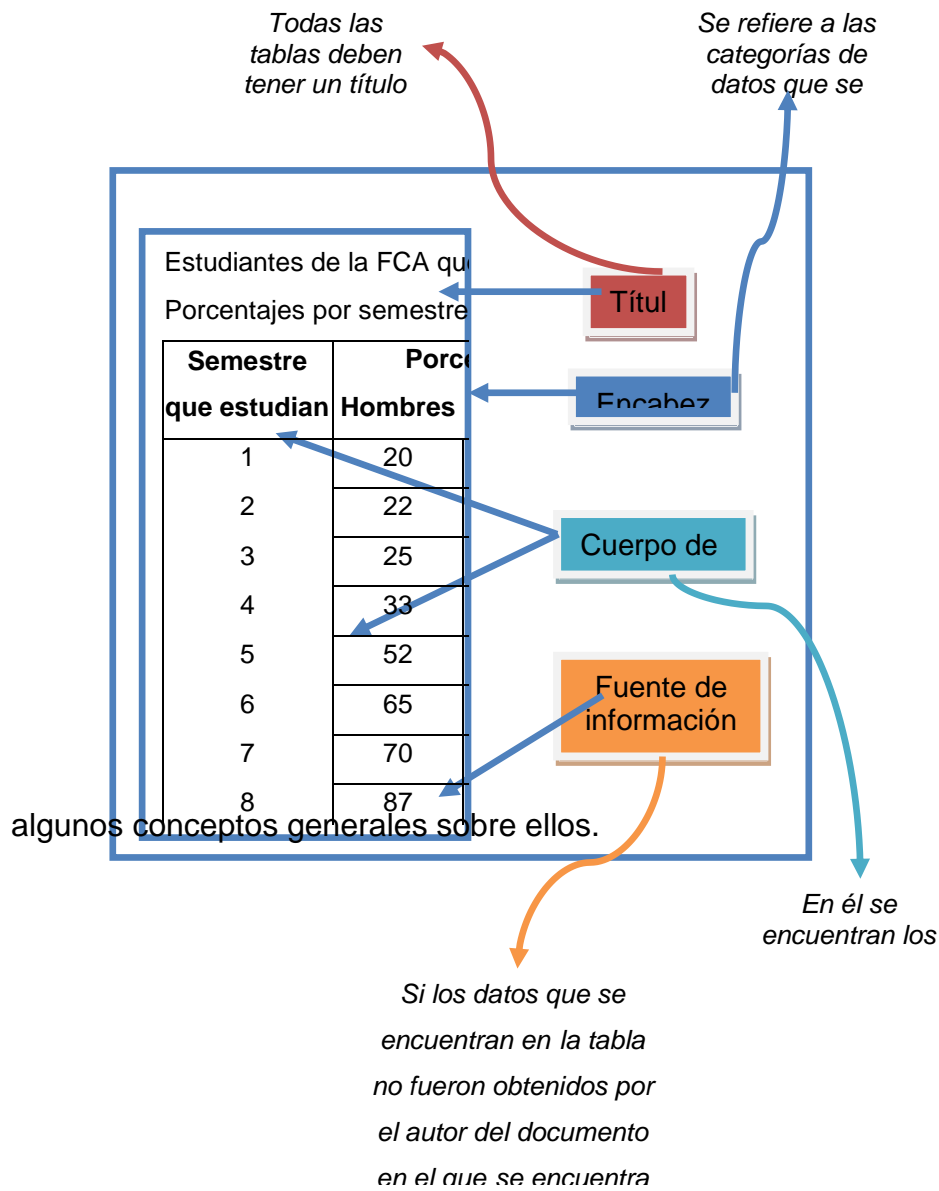
Existen herramientas de uso general como el **Excel o Lotus** que incluyen algunas funciones estadísticas y son útiles para muchas aplicaciones. Sin embargo, si se desea estudiar con mayor profundidad el uso de técnicas más avanzadas es importante contar con herramientas específicamente diseñadas para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





SUAYED
UNA OPCIÓN
PARA TI

Tabla sencilla de datos

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



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 la persona que la elabora, pero que tiene siempre, en una de las columnas, información sobre el número de veces (frecuencia) que se presenta cierto fenómeno.

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

| Deportes Batista, S.A. de C.V. | | |
|--|------------|----------------|
| Número de bicicletas vendidas por tienda | | |
| Primer trimestre de 20XX | | |
| Tienda | Unidades | 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 |
| <i>Los números en rojo representan los porcentajes</i> | | | | | | | | | | |

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



2.2 Distribuciones de frecuencia

Una distribución de frecuencias o tabla de frecuencias no es más que la presentación tabular de las frecuencias 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 f_i . 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 f_i . 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 (f_i) 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 cabecera.** 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 cabecera. 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.

10.

Cuadro 1.1 **Distribución de alumnos por días de ausencia**

Mes base enero

| Ausencia (valores de variable) | Número de alumnos (Frecuencia) |
|-----------------------------------|-----------------------------------|
| | |
| | |
| | |
| | |
| | |
| | |
| | |



Casillas. Es la intersección que forma 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 |
| 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.



2.3 Presentación gráfica de datos

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

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



Histogramas y polígonos de frecuencias

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

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



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

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

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

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



Considera el siguiente conjunto de datos:

| | | | | | | |
|---------------|------------|------------|------------|------------|------------|---------------------|
| 8.9 | 8.3 | 9.2 | 8.4 | 9.1 | | |
| 8.6 | 8.9 | 9.1 | 8.8 | 8.8 | | |
| 8.8 | 9.1 | 8.9 | 8.7 | 8.8 | | |
| 8.9 | 9.0 | 8.6 | 8.7 | 8.4 | | |
| 8.6 | 9.0 | 8.8 | 8.9 | 9.1 | | |
| 9.4 | 9.0 | 9.2 | 9.1 | 8.8 | | |
| 9.1 | 9.3 | 9.0 | 9.2 | 8.8 | | |
| 9.7 | 8.9 | 9.7 | 8.3 | 9.3 | | |
| 8.9 | 8.8 | 9.3 | 8.5 | 8.9 | | |
| 8.3 | 9.2 | 8.2 | 8.9 | 8.7 | | |
| 8.9 | 8.8 | 8.5 | 8.4 | 8.0 | | |
| 8.5 | 8.7 | 8.7 | 8.8 | 8.8 | | |
| 8.3 | 8.6 | 8.7 | 9.0 | 8.7 | | |
| 8.4 | 8.8 | 8.4 | 8.6 | 9.0 | | |
| 9.3 | 8.8 | 8.5 | 8.7 | 9.6 | | |
| 8.5 | 9.1 | 9.0 | 8.8 | 9.1 | | |
| 8.6 | 8.6 | 8.4 | 9.1 | 8.5 | | |
| 9.1 | 9.2 | 8.8 | 8.5 | 8.3 | | |
| 9.3 | 8.6 | 8.7 | 8.7 | 9.1 | | |
| 8.8 | 8.7 | 9.0 | 9.0 | 8.5 | | |
| 8.5 | 8.8 | 8.9 | 8.2 | 9.0 | | |
| 9.0 | 8.7 | 8.7 | 8.9 | 9.4 | | |
| 8.3 | 8.6 | 9.2 | 8.7 | 8.7 | | |
| 8.7 | 9.7 | 8.9 | 9.2 | 8.8 | | |
| 8.3 | 8.6 | 8.5 | 8.6 | 9.7 | | |
| Máximo | 9.7 | 9.7 | 9.7 | 9.2 | 9.2 | máximo = 9.7 |
| Mínimo | 8.3 | 8.3 | 8.2 | 8.2 | 8.0 | 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 mas grande como el mas 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 $R = \text{MÁX} - \text{MÍN} = 9.7 - 8.0 = 1.7$.

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 menos de 6 ni más de 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 es necesario dar una regla general que se mantenga para lograr una normalización del histograma. En el ejemplo, la lectura menor fue de 8.0 por lo que se podría tomar este 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 | 8.00-8.19 | 8.1 | I | 1 |
| 2 | 8.20-8.39 | 8.3 | IIII IIII | 9 |
| 3 | 8.40-8.59 | 8.5 | IIII IIII IIII I | 16 |
| 4 | 8.60-8.79 | 8.7 | IIII IIII IIII IIII IIII II | 27 |
| 5 | 8.80-8.99 | 8.9 | IIII IIII IIII IIII IIII IIII I | 31 |
| 6 | 9.00-9.19 | 9.1 | IIII IIII IIII IIII III | 23 |
| 7 | 9.20-9.39 | 9.3 | IIII IIII 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:

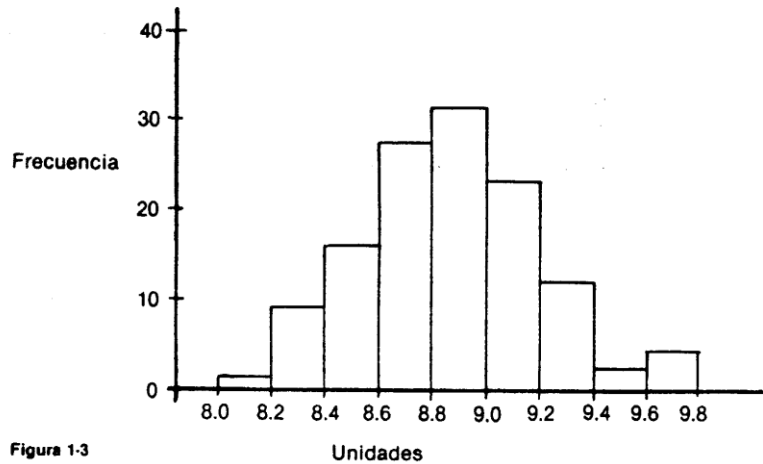


Figura 1-3

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:

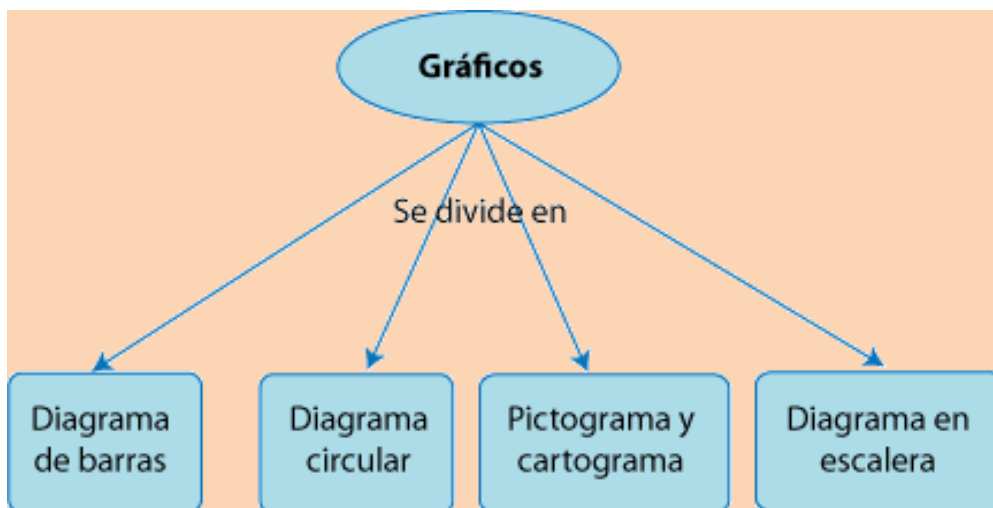
- **Distribuciones sin agrupar**
- **Distribuciones agrupadas en intervalos**



Gráficas 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 y consiste en representar datos mediante una barra o columna simple, la cual puede ser colocada horizontal o verticalmente.



Este gráfico permite comparar las proporciones que guardan cada una de las partes con respecto al todo, por lo que pueden construirse usando valores absolutos, proporciones o bien porcentajes. Suelen utilizarse cuando se comparan gráficamente las distribuciones de iguales conceptos en dos o más periodos. Asimismo, constituye la representación gráfica más utilizada, por su capacidad para adaptarse a numerosos conjuntos de datos.

La forma de elaborar estos diagramas es la siguiente:

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

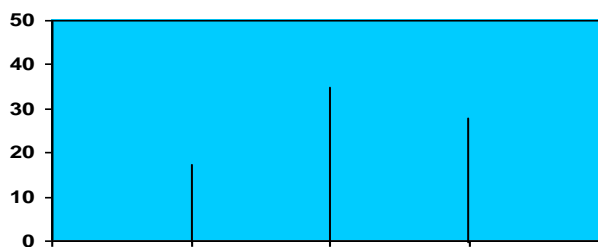
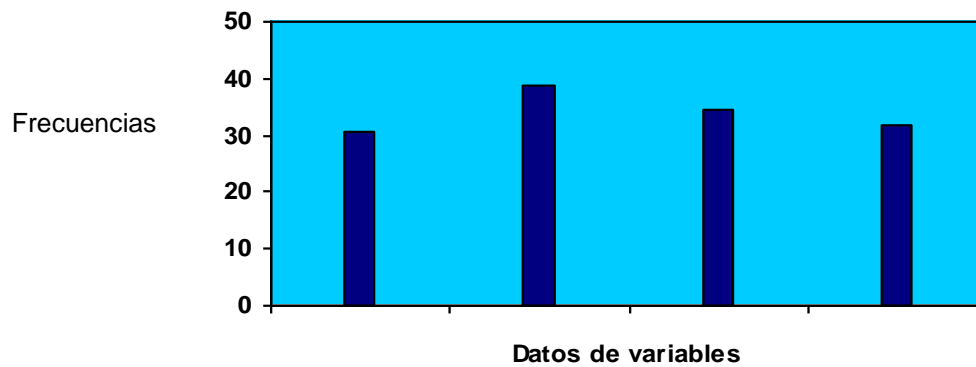


Diagrama de barras

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



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



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

Diagrama de rectángulos

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



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

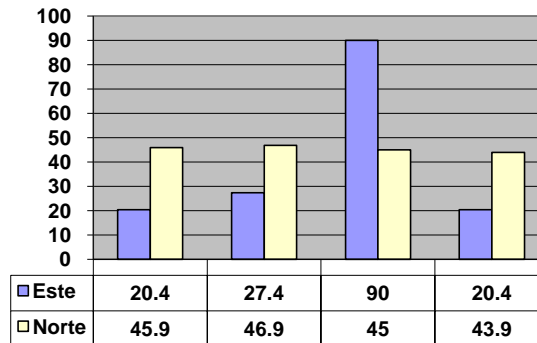


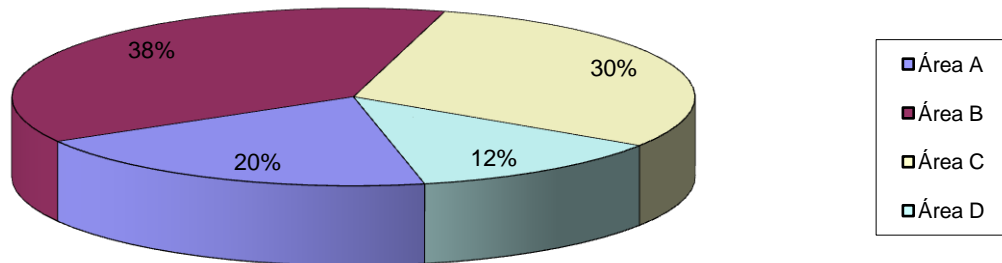
Diagrama de rectángulos

b) Diagrama circular.

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

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

- Se 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 360° , éstos se reparten proporcionalmente a las frecuencias absolutas de cada componente.



Grafica 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.



SUAYED
UNA OPCIÓN
PARA TI

Por ejemplo:

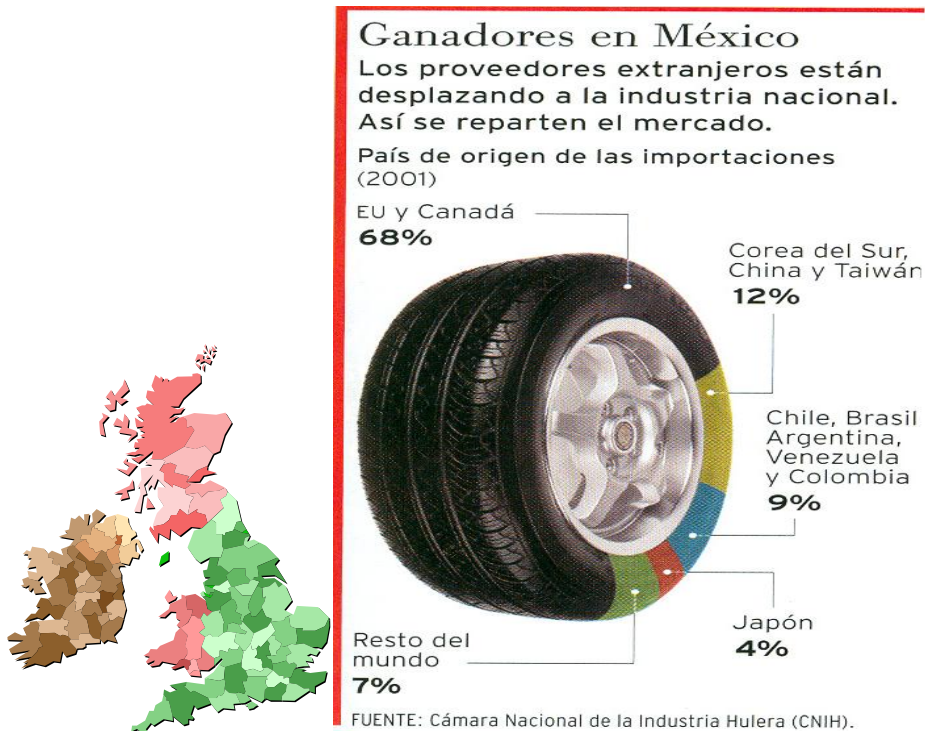
 = 100 viviendas

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*, No. 852 (octubre 30 del 2002), p. 69.



d) Diagrama en escalera.

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

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

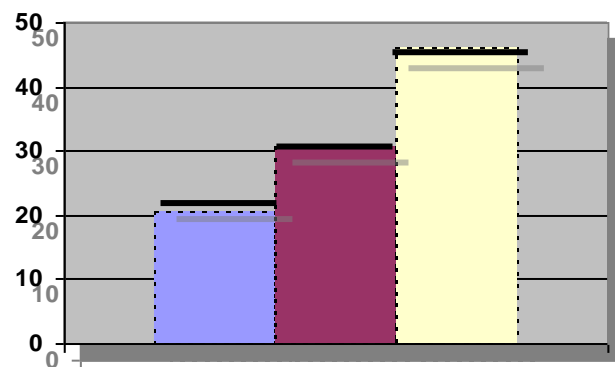


Diagrama en escalera

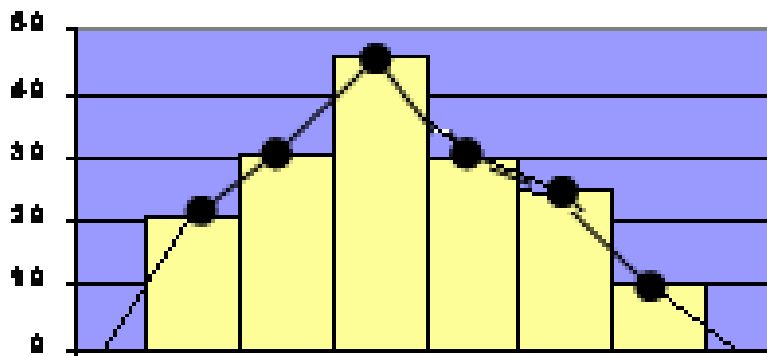


Gráficas para distribuciones de frecuencias agrupadas en clases

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

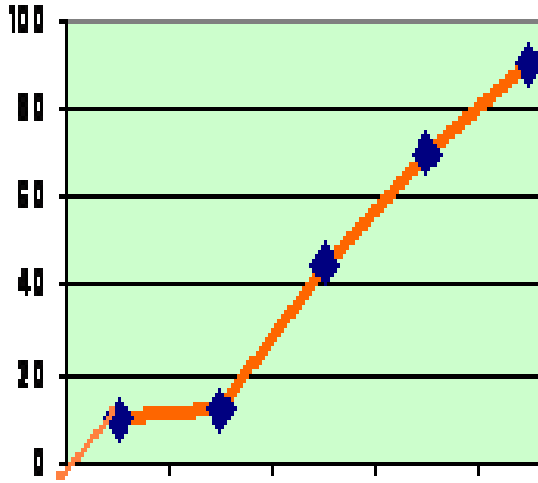
Si en lugar de frecuencias absolutas utilizamos las acumuladas, obtendremos, en vez del histograma, una representación gráfica en forma de línea creciente que se conoce con el nombre de ojiva. Estos gráficos son especialmente adecuados cuando se tiene interés en saber cuántas observaciones se acumulan 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.

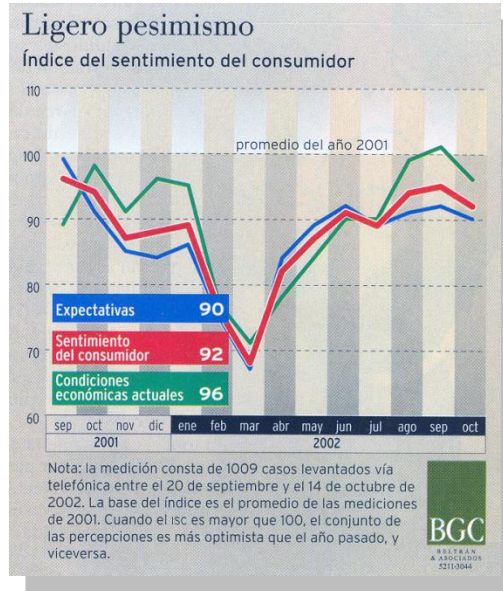
La ojiva también es un polígono que se puede construir con la frecuencia acumulada relativa.



SUAYED UNA OPCIÓN PARA TI



Ojivas



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

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

Columnas.

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

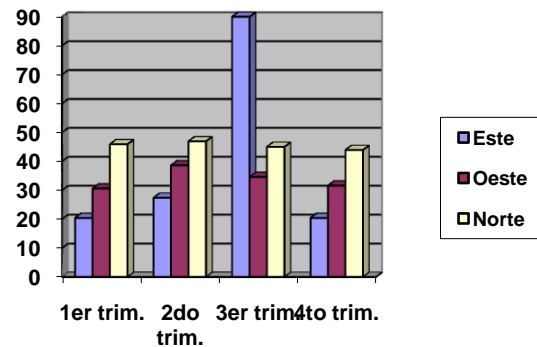
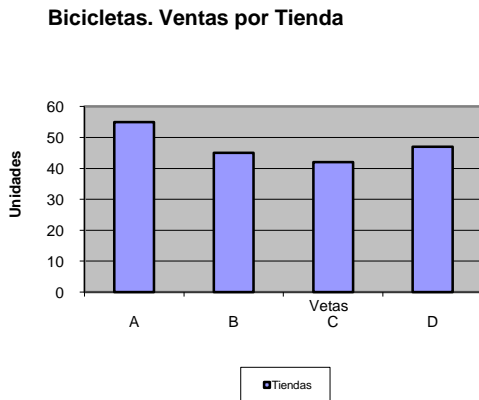




Diagrama de columnas

Barras.

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

Bicicletas. Ventas por tienda

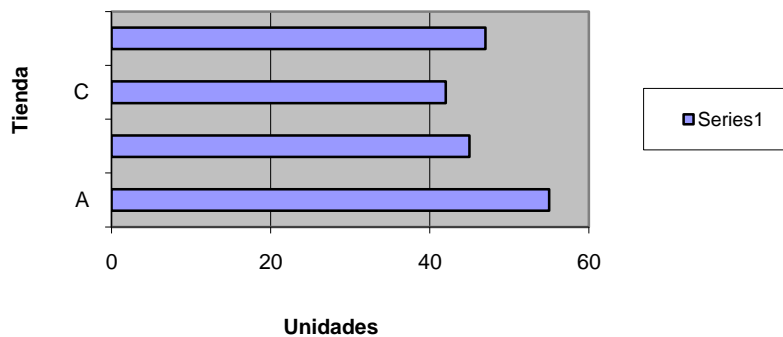


Diagrama de barras

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.



Bicicletas, ventas por tienda

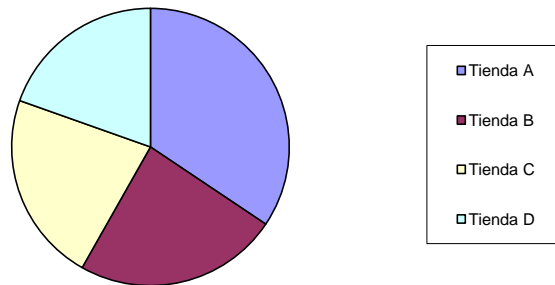


Diagrama circular

2.4 Medidas de tendencia central

Hemos visto que tanto las tablas como las gráficas pueden ser 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 de ésta. A estas medidas se les llama “parámetros” de acuerdo con lo visto en la unidad 1. Se dividen en **medidas de posición y medidas de dispersión.**



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 medidas de posición.

Las **medidas de tendencia central** son aquellas 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) el 25% de la información y a la derecha (datos más grandes), el 75% de la información. A este punto se le denomina primer cuartil o Q1.

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

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



Media aritmética

La media aritmética es el promedio que todos conocemos desde nuestros años de infancia. Se obtiene sumando todos los datos y 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.

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

La media aritmética está dada por la siguiente expresión:

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

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



Mediana

Es el valor que divide la distribución en dos partes iguales 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 nos hay dos datos iguales; por lo mismo, este conjunto de datos no tiene moda.

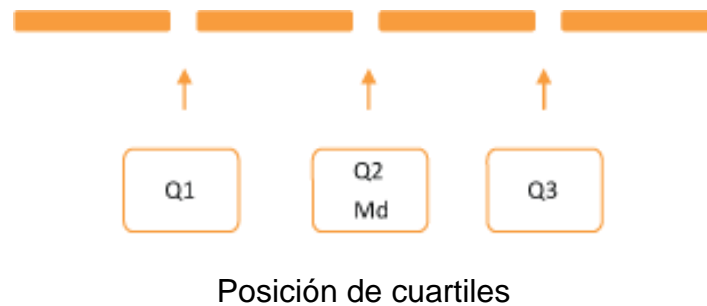
Puede darse el caso, en conjuntos más grandes de datos, que el “honor” de ser el valor más frecuente sea compartido por dos datos. En ese caso se afirma que la distribución es **bimodal**, pues tiene dos modas. Algunos autores llegan a hablar de distribuciones **primordiales** e incluso más.

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**:



Cada una de las barras amarillas 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.



2.5 Medidas de dispersión

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

Tienda A. \$10,000; \$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 siguientes 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, que tan alejado está cada dato de la media ya citada. En la siguiente tabla se realiza ese trabajo.



Tabla de desviaciones de datos

| Tienda C | | Tienda D | |
|----------|--------------------------|----------|--------------------------|
| datos | cada dato menos la media | datos | cada dato menos la media |
| 5000 | $5000-10000=-5000$ | 5000 | $5000-10000=-5000$ |
| 10000 | $10000-10000=0$ | 6000 | $6000-10000=-4000$ |
| 10000 | $10000-10000=0$ | 10000 | $10000-10000=0$ |
| 10000 | $10000-10000=0$ | 14000 | $14000-10000=4000$ |
| 15000 | $15000-10000=5000$ | 15000 | $15000-10000=5000$ |
| | Suma = 0 | | Suma = 0 |

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 |
| 5000 | 5000 | 25,000,000 | 5000 | -5000 | 25,000,000 |
| 10000 | 0 | 0 | 6000 | -4000 | 16,000,000 |
| 10000 | 0 | 0 | 10000 | 0 | 0 |
| 10000 | 0 | 0 | 14000 | 4000 | 16,000,000 |
| 15000 | 5000 | 25,000,000 | 15000 | 5000 | 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 $50000000/5$, 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 (5000) Kg.

Para manejar el ejemplo de manera sencilla, supongamos que en un día determinado solamente vendió tres costales y que además salieron tres camiones cargados; para verificar el trabajo de los operarios, se pesaron tanto unos como otros en presencia de un supervisor. Sus pesos, la



media de los mismos y sus desviaciones estándar aparecen en la siguiente tabla (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 | 4990Kg |
| 50 Kg. | 5000 Kg. |
| 60 Kg. | 5010 Kg. |

Tabla de datos

- Media de los costales 50 Kg.
- Media de los camiones 5000 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 del 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 ni menos de cinco ni más de 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 se debemos asignarlo a la segunda o a la tercera clase. Para remediar esta situación existen varios caminos, pero el más práctico de ellos (y el que usaremos para los efectos de este trabajo) es el de hacer intervalos abiertos por un lado y cerrados en el otro.



Esto se logra de 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 X_i . En nuestro ejemplo se tendría:

| Clase | Incluye datos iguales o mayores a: | Incluye datos menores a: | Marca de clase (X_i) |
|---------|------------------------------------|--------------------------|--------------------------|
| 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

Estas 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 clasificar éstos. 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 | | 11 |
| Segunda | 600 | 700 | | 11 |
| Tercera | 700 | 800 | | 7 |
| Cuarta | 800 | 900 | | 6 |
| Quinta | 900 | 1000 | | 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:

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

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:

$$Md = L_M + \frac{\frac{n}{2} - F_M}{f_M} \cdot i$$

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:

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

En donde:

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

d_1 es la diferencia entre la frecuencia de clase f_{Mo} del intervalo que contiene a la moda y la frecuencia de la clase inmediata anterior f_1 .

d_2 es la diferencia entre la frecuencia de clase f_{Mo} del intervalo que contiene a la moda y la frecuencia de la clase inmediata posterior f_2 .

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:

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

En donde:

x_i es la marca de clase.

f_i es la frecuencia de clase.

\bar{x} es la media.

n es el número de datos.

c) Desviación estándar:

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

d) Coeficiente de variación:

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

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 si es importante manejar esta distinción ya que se trabaja con muestras para inferir los parámetros poblacionales.

En el ejemplo siguiente se muestra la utilización de las fórmulas descritas:



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

| Límites reales | x_i | f_i | $f_i acum$ | $x_i f_i$ | $x_i - \bar{x}^2 f_i$ |
|----------------|-------|-------|------------|-----------|-----------------------|
| 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:**

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

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

c) Mediana:

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

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 .

d) Moda o modo:

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



$$Mo = 80 + \frac{2}{2+7} \cdot 5 = 80.11, \quad \text{en donde: } \begin{array}{l} d_1 = 20 - 18 = 2 \\ d_2 = 20 - 13 = 7 \end{array} \quad \begin{array}{l} f_{Mo} = 20 \\ f_1 = 18 \\ f_2 = 13 \\ i = 5 \end{array}$$

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.

2.6 Teorema de Tchebysheff y regla empírica

El teorema de Tchebysheff y la regla empírica nos permiten inferir el porcentaje de elementos que deben quedar dentro de una cantidad específica de desviaciones estándar respecto a la media. 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 dos desviaciones estándar arriba. Al aplicar el teorema de Tchebysheff, cuando menos el 0.75 o 75% de los elementos deben tener valores a menos de dos desviaciones estándar del promedio. Así, cuando menos 75 de los 100 alumnos deben haber obtenido calificaciones entre 60 y 80.

Para las calificaciones entre 58 y 82, 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 el 82.6% de los alumnos deben tener calificaciones entre 58 y 82.

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

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

Sin embargo, en las aplicaciones prácticas se ha encontrado que muchos conjuntos de datos tienen una distribución en forma de colina o de campana, 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 el 68% de los elementos están a menos de una desviación estándar de la media.
- aproximadamente el 95% de los elementos están a menos de dos desviaciones estándar de la media.
- casi todos los elementos están a menos de tres desviaciones estándar de la media.

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

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



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

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



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 de la manera en que se midan los atributos dependerá el tipo de indicador que se puede construir.



GLOSARIO

Desviación estándar

Es una medida de variabilidad que corresponde a la raíz cuadrada de la varianza.

Distribución de frecuencias

Es una herramienta tabular que permite mostrar cómo se distribuyen las observaciones en los distintos intervalos o clases en que se ha dividido una colección de datos.

Histograma

Es una representación gráfica de la distribución de frecuencias que consiste en una serie de columnas o barras verticales, cada una de las cuales tiene como base un intervalo de clase y como altura, la frecuencia respectiva.

Intervalo de clase

Cualquiera de los rangos en los que se acomodan los distintos datos en una distribución de frecuencias.

Marca de clase

Es el punto medio de un intervalo de clase.



Media

Es un nombre genérico con el que por lo general se hace referencia a la media aritmética o promedio que se obtiene de dividir entre el número de observaciones un total obtenido mediante la suma de los valores.

Mediana

Es el valor en el que se acumula el 50% de los datos u observaciones. En el caso de una distribución simétrica es igual a la media.

Moda

En un conjunto de datos es el valor más común o el que se repite más veces. En caso de haber más de una moda se habla de distribuciones multimodales.

Ojiva

Es una representación gráfica de la distribución acumulada de frecuencias, por lo general relativas, que se construye uniendo con segmentos de línea la sucesión de puntos (Li, Fi) donde Li es el límite superior del intervalo i , y Fi la frecuencia acumulada respectiva.

Percentil

Es el valor X_p en el que se acumula una frecuencia o número de casos equivalente a $Np/100$, donde N es el total de datos (esto es, el $p\%$ de los mismos). Dentro de los percentiles más comúnmente usados se encuentran la mediana (percentil 50); primer, segundo y tercer cuartil (percentiles 25, 50 y 75), y los deciles (percentiles 10, 20, ..., 90).



Polígono de frecuencias

Es una representación gráfica de la distribución de frecuencias que se construye uniendo con segmentos de línea la sucesión de puntos (m_i, f_i) donde m_i es la marca de clase del intervalo y f_i la frecuencia respectiva.

Varianza

Es una medida de variabilidad que indica qué tanto se alejan (o se acercan) las distintas observaciones al punto indicado por la media aritmética de los propios datos.



ACTIVIDADES DE APRENDIZAJE

ACTIVIDAD 1

Un profesor aplicó el primer día de clase una encuesta a sus alumnos de primer ingreso. Días después le comentó al grupo que 53% de ellos no tenía coche y que el 62% trabajaba. Les dijo además que le sorprendía el hecho de que 35% no trabajaba pero tenía coche.

Estructura una tabla donde puedas incorporar estos datos. Complétala y señala cuál es el porcentaje de estudiantes que no trabaja y no tiene coche.

ACTIVIDAD 2

Realiza la lectura del documento **Operacionalización de variables**, de Betancur López, Sonia Inés, publicado por la revista Hacia la promoción de la salud, No 5, Departamento de Salud Pública, Universidad de Caldas, Colombia.

Elabora un cuadro sinóptico sobre las escalas de medición.



ACTIVIDAD 3

Considera la siguiente situación.

En un hotel de playa recibieron durante un fin de semana a 140 personas. El Gerente quiso saber su perfil de edad por lo que preguntó a los encargados de recepción algunos datos. Le indicaron que 39 personas tenían 65 o más años de edad, que otros 12 tenían de 18 a menos de 25 años de edad, que 80 tenían menos de 35 años, que 50 tenían menos de 18 años, que 30 tenían menos de 12 años de edad y que 20 tenían menos de 4 años de edad.

- Elabora con los datos proporcionados la respectiva distribución de frecuencias incorporando las frecuencias acumuladas.
- Explica en un párrafo con tus propias palabras qué diferencia hay entre la frecuencia y la frecuencia acumulada y cómo se calcula una a partir de la otra.

ACTIVIDAD 4

Considera la actividad integradora de la Unidad 1 en que se te pidió elaborar un fichero para asentar los datos de los estudiantes que integran el grupo.

- a) Establece para cada variable considerada en el fichero, la escala de medida que le corresponde
- b) Define y elabora por lo menos seis tablas de doble entrada, anotando sus títulos y encabezados
- c) Determina la distribución de frecuencias de cada variable.



ACTIVIDAD 5

Las autoridades de una institución educativa han realizado desde hace años un seguimiento de egresados. Uno de los aspectos que siempre incluyen en su estudio es el relativo al monto de las ventas mensuales de la empresa donde labora el egresado. Los datos que se presentan a continuación se refieren a la distribución de egresados según las ventas mensuales por año considerado.

- a) Elabora dos formas alternativas de presentar gráficamente los datos. Puedes apoyarte en las facilidades que a este respecto ofrece el software de hoja de cálculo.
- b) Explica qué ventajas tendría una y otra alternativa. Imagina que estás ante un auditorio que no ha tenido acceso a la presentación tabular de los datos.

Escribe un texto donde expliques e intérpretes las gráficas respectivas.

| Año | Ventas Mensuales de la Empresa | | | | | Total |
|-------------|--------------------------------|-----------------------|-----------------------|-----------------------|------------------|-----------|
| | Menos de \$100 000 | \$100 000 a \$200 000 | \$200 000 a \$300 000 | \$300 000 a \$400 000 | Mas de \$400 000 | |
| 1980 | 28 | 7 | 3 | 3 | 1 | 42 |
| 1985 | 25 | 9 | 4 | 3 | 1 | 42 |
| 1990 | 25 | 8 | 4 | 4 | 1 | 42 |
| 1995 | 24 | 10 | 3 | 5 | 0 | 42 |
| 2000 | 24 | 14 | 2 | 1 | 1 | 42 |
| 2005 | 31 | 5 | 0 | 5 | 1 | 42 |



ACTIVIDAD 6

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

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

Considerando la forma de la distribución de los datos, y con el propósito de determinar el porcentaje de datos que está a menos de 2 veces la desviación estándar respecto del promedio, indica qué sería mejor: aplicar el teorema de Tchebysheff o la regla empírica.

ACTIVIDAD 7

Considera la actividad Lo que aprendí de la Unidad 1 en que se te pidió elaborar un fichero para asentar los datos de los estudiantes que integran el grupo. En la sección 2 de esta unidad se te pidió determinar la distribución de frecuencias de cada variable. Con estos antecedentes, determina:

- a) las medidas de tendencia central de cada variable.
- b) la varianza y desviación estándar de cada variable



ACTIVIDAD 8

En un banco de sangre se han presentado solicitudes cuya distribución por tipo de sangre solicitado, para dos periodos de observación, son como se muestra en la siguiente tabla. Al dueño del banco no le parece clara esta presentación y te pide que elabores un gráfico que describa el comportamiento de las solicitudes.

| Grupo de sangre | Periodo | |
|-----------------|----------------|----------------|
| | 1 | 2 |
| A | 17.65 % | 14.81 % |
| B | 25 % | 27.78 % |
| AB | 35.29 % | 33.33 % |
| O | 22.06 % | 24.07 % |

Elabora la gráfica, anotando en ella la información que consideres pertinente, incluyendo un título. Supón que en el periodo uno se registraron 68 solicitudes y en el periodo dos, 54.

Indica además :

- Si es cierto que en el periodo uno hay más boletas llenas con solicitudes de sangre tipo O que en el periodo dos.
- ¿Por qué en ambos casos la suma es 100%?



CUESTIONARIO DE REFORZAMIENTO

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



LO QUE APRENDÌ

Considera las actividades que has venido desarrollando desde la Unidad 1 con los datos de los estudiantes que integran el grupo del cual tú formas parte. Redacta un informe con tres conclusiones salientes obtenidas a partir de las medidas de tendencia central y de dispersión y el teorema de Tchebysheff o la regla empírica para construir intervalos o rangos en cada variable cuantitativa, de modo que en cada caso se abarque al 90% de la población que has calculado para esa población.



EXAMEN DE AUTOEVALUACIÓN

I. A continuación se presentan 9 atributos o variables. Señala a qué escala de medición corresponde cada uno de ellos, Nominal (N), Ordinal (O), Intervalar (I) o de Razón (R).

| N | O | I | R | |
|-----------------------|-----------------------|-----------------------|-----------------------|--------------------------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | 1. Religión |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | 2. Preferencia sexual |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | 3. Peso corporal |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | 4. Temperatura |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | 5. Ingreso |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | 6. Idioma |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | 7. Promedio de calificaciones |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | 8. Longitud |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | 9. Gusto por platillos |



II. Había una vez un hombre muy rico que se jactaba de tener empresas en diversas partes de nuestro continente. Le gustaba comentar que sus empleados estaban en todos lados.

La tabla siguiente condensa información al respecto, tomando datos de un informe de labores del año 2005, sólo que tiene 7 errores. Algunos son pequeños, casi intrascendentes, pero a final de cuentas, son errores.

Escribe en los espacios vacíos de la tabla los conceptos que hacen falta (recuerda colocar acentos).

Distribución de Empleados por Sector

| <input type="text"/> | México | Sudamérica | Total | <input type="text"/> |
|----------------------|----------------------|----------------------|-------|----------------------|
| Agroindustria | 458 | 59 | 517 | <input type="text"/> |
| Agropecuario | <input type="text"/> | 712 | 1057 | <input type="text"/> |
| Transporte | 125 | 41 | 166 | <input type="text"/> |
| Minería | 326 | <input type="text"/> | 441 | <input type="text"/> |
| Servicios | 1578 | 321 | 1899 | <input type="text"/> |
| <input type="text"/> | 2832 | 1248 | 4080 | <input type="text"/> |

Fuente: Informe Labores.



III. Se tomo una muestra de 8 calificaciones de un grupo de alumnos que presentó un examen final de ocho preguntas. Las calificaciones son:
4.5, 5.3, 3.3, 2.9, 3.9, 7.6, 5 y 3.8

Se desea conocer el valor de la media y la mediana. Para ello te pedimos que completes el siguiente cuadro

| | |
|--|--|
| Suma de calificaciones | |
| Media de calificaciones | |
| Observación en donde cae la mediana | |
| Mediana | |

IV. Considera los datos y los resultados del ejercicio de autoevaluación de la sección anterior. Para tu comodidad aquí te volvemos a presentar el texto base del problema

Se tomo una muestra de 8 calificaciones de un grupo de alumnos que presentó un examen final de ocho preguntas. Las calificaciones son:

4.5, 5.3, 3.3, 2.9, 3.9, 7.6, 5 y 3.8

¿Cuál es el porcentaje de observaciones que caen entre 4.08 y 6.00?

- a) 9.35%
- b) 31.08%
- c) No se puede determinar porque los límites del intervalo corresponde a una distancia de menos de una desviación estándar alrededor de la media
- d) - 9.35%



MESOGRAFÍA

BIBLIOGRAFÍA RECOMENDADA

| Autor | Capítulo | Páginas |
|----------------------------------|--|----------------|
| 1. Berenson, Levine, y Krehbiel. | 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 variados | 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 forma. | 103 - 127 |
| | 3.4 Obtención de medidas descriptivas de resumen a partir de una población. | 133 -139 |
| 2. Levin y Rubin. | 2. Agrupación y presentación de datos para expresar significados: tablas y gráficas. | 8-11 |



| | | |
|---------------------------|---|-----------|
| | Secciones: 2.1, ¿Cómo podemos ordenar los datos? | |
| | 2.3 Ordenamiento de datos en arreglos de datos y distribuciones de frecuencias. | 12- 20 |
| | 2.4 Construcción de una distribución de frecuencias. | 20- 29 |
| | 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 |
| 3. Lind, Marchal, Wathen. | 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 |



BIBLIOGRAFÍA BÁSICA

- 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.

BIBLIOGRAFÍA COMPLEMENTARIA

1. Bowerman Bruce, *Pronósticos, series de tiempo y regresión; un enfoque aplicado*, México: Cengage Learning, 4ª edición, 2007, 720 pp.
2. Mendenhall William, *Introducción a la probabilidad y estadística*, México: Cengage Learning, 13ª edición, 2010, 776 pp.
3. Webster Allen L., *Estadística I aplicada a los negocios y la economía*, México: McGraw-Hill, 2ª. edición, 2002, 154 pp.



SITIOS DE INTERNET

- *Desigualdad de Chebyshev*, en la página de wikipedia, en el sitio:

[http://es.wikipedia.org/wiki/Desigualdad de Chebyshev](http://es.wikipedia.org/wiki/Desigualdad_de_Chebyshev), en donde se presenta la desigualdad de Chebyshev a nivel de teorema así como algunas de sus consecuencias.

- *Gráficas estadísticas*, del Taller estadístico del programa Estadística para todos, en el sitio:

<http://www.estadisticaparatodos.es/taller/graficas/graficas.html> ; en donde se presentan los pasos para construir diagramas de tallos y hojas así como de caja y bigotes y se proporcionan elementos para la interpretación de los mismos.

- *Gráficas estadísticas*, publicado por Plan Ceibal, Uruguay, en el sitio:

[http://www.ceibal.edu.uy/contenidos/areas_conocimiento/mat/estadistica/grficas estadsticas.html](http://www.ceibal.edu.uy/contenidos/areas_conocimiento/mat/estadistica/grficas_estadsticas.html); en donde se presenta un pequeño catálogo de diversos recursos gráficos.

- *Parámetros muestrales y poblacionales*, de Mullor, Rubén, en el sitio:

<http://www.eio.ua.es/licmat/cursocero/apuntes/introduccion%20a%20la%20estadistica.doc>; sección 2 del texto *Introducción a la estadística*, apuntes de la asignatura Razonamiento matemático y matemáticas discretas. En esta liga, el profesor Mullor, de la



licenciatura en matemáticas de la Universidad de Alicante, España, hace una presentación puntual de la media, varianza, desviación estándar y su interpretación y se proponen algunos problemas sencillos relativos a estos conceptos.

- *Tema 2, Medidas de centralización*, de Mateu Mahiques, Jorge, catedrático del Departamento de Matemáticas de la Universidad Jaume I, España, en el sitio:

<http://www3.uji.es/~mateu/Tema2-D37.doc>, en donde se presenta una lista de los principales indicadores a través de los cuales se pueden cuantificar las características fundamentales de grupos de datos y se desarrollan y ejemplifican las expresiones algebraicas y propiedades de las medidas de tendencia central.

- *Varianza y desviación estándar o desviación típica*, en el sitio:

<http://dieumsnh.qfb.umich.mx/estadistica/varianza.htm>, publicado por la Coordinación de Innovación Educativa de la Escuela de Químico Farmacobiología de la Universidad Michoacana de San Nicolás de Hidalgo, en donde se desarrolla el tema de la varianza y desviación estándar y se presentan algunos resultados importantes respecto de estos indicadores.



UNIDAD 3

ANÁLISIS COMBINATORIO

APUNTES DIGITALES PLAN 2011



OBJETIVO ESPECÍFICO DE LA UNIDAD

El alumno diferenciara los procesos de multiplicación, permutación y combinación



INTRODUCCIÓN

Desde hace miles de años el hombre realiza juegos de azar, por lo que la necesidad de plantear soluciones a los problemas que se presentan en estos juegos no es nueva. Con el paso del tiempo se han enunciado principios aplicables al problema de obtener la máxima probabilidad de éxito de una estrategia aplicada. Como veremos, en muchos casos, se ha observado que el problema radica en contar de cuántas formas puede ocurrir cierta situación.

Establecer procedimientos eficientes y eficaces de conteo es la esencia del análisis combinatorio.

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



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

Muchas de las decisiones comerciales requieren que se cuente el número de subconjuntos que se pueden obtener de un conjunto. Como ejemplos tenemos:

- 1) de las ventas de una línea que consta de 10 productos, ¿cuántos subconjuntos de tres productos se pueden ofrecer a los clientes?;
- 2) ¿cuántos números telefónicos distintos de 8 dígitos pueden asignarse a una oficina, si todos deben empezar con el código 55 32?

Existen muchos ejemplos más en los cuales será importante utilizar diferentes reglas de conteo para su solución.

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



LO QUE SÉ

¿Sabías que el universo en que vivimos tiene espacio para 10^{118} partículas subatómicas, lo que significa que hay 2 a la 10^{118} combinaciones distintas de tales partículas?

(Fuente: Universos paralelos, Scientific American México, año 2, núm. 3, pág. 39)



Pensando en estos datos, ¿qué opinas de la frase, muy coloquial, que dice que este es un mundo de posibilidades infinitas?



TEMARIO DETALLADO

- 3.1 Principios fundamentales.
- 3.2 Ordenaciones, permutaciones y combinaciones.



3.1 Principios fundamentales

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

Principio de multiplicación

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

Si hubiese un tercer evento independiente que puede ocurrir de "o" maneras diferentes, entonces el número de arreglos distintos estaría dado por el producto $m \times n \times o$.

El principio es por entero aplicable para cualquier número finito de eventos

A continuación, analizaremos algunos ejemplos.

Ejemplo 1:



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

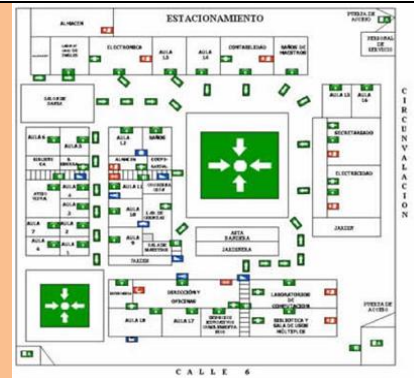
Solución:

$m = 3$ formas de llegar a la escuela

$n = 2$ entradas

Por lo que $m \times n = 3 \times 2 = 6$.

Existen 6 maneras de hacer su arribo



Ejemplo 2:

Juan acostumbra comer todos los días en el Restaurante La Deliciosa. El menú consta de cuatro tiempos. En el primero se puede escoger de entre dos opciones, en el segundo el platillo es fijo, en el tercero hay cuatro posibilidades y en el cuarto otras tres. ¿De cuántas maneras distintas puede Juan ordenar su comida?

Solución:

Utilizaremos la letra m con subíndices del 1 al 4 para indicar el número de opciones que hay en





cada tiempo de la comida.

m_1 = número de opciones en el primer tiempo = 2

m_2 = número de opciones en el segundo tiempo = 1

m_3 = número de opciones en el tercer tiempo = 4

m_4 = número de opciones en el cuarto tiempo = 3

Entonces, hay $m_1 \times m_2 \times m_3 \times m_4 = 2 \times 1 \times 4 \times 3 = 24$ formas distintas de armar el menú

Ejemplo 3:

La contraseña de acceso a un sistema de asesoría en línea está formada por 3 letras y cuatro números distintos entre sí. ¿Cuántas contraseñas distintas se pueden formar?

Solución:

Podemos considerar que hay 7 eventos, cada uno asociado a una posición en la contraseña. Para cada uno de los tres primeros eventos hay 26 opciones. Para los cuatro últimos eventos hay 9, 8, 7 y 6 opciones respectivamente porque no pueden repetirse los números. Entonces hay $26 \times 26 \times 26 \times 9 \times 8 \times 7 \times 6 = 53\ 149\ 824$ contraseñas distintas





Principio de adición

Supongamos que tenemos dos eventos o sucesos A y B , que no pueden ocurrir simultáneamente, lo que en la terminología de conjuntos significa que la intersección es vacía, $(A \cap B = \emptyset)$, y además, que el evento A se puede realizar de “ m ” maneras mientras que B se puede realizar de “ n ” maneras diferentes. Entonces el evento $(A \cup B)$ puede ocurrir $(m + n)$ maneras.

Ejemplo 4. Juan acostumbra todos los días tomar alguna bebida a las 10:30 de la mañana. Sus opciones son café, té o chocolate. La máquina que proporciona el servicio cuenta con café americano, café moka y café descafeinado; además hay siete diferentes sabores de té y chocolate frío o caliente. ¿Cuántas opciones tiene Juan?

Solución:

Si aceptamos que sólo puede tomar una bebida, es claro que cuenta con $3 + 7 + 2 = 12$ opciones

Aquí hemos aplicado el principio de adición

Ejemplo 5. Para trasladarse a la Facultad, la novia de Juan puede abordar el metro, tomar uno de 3 servicios de autobús o uno de 6 servicios de colectivo. ¿Cuántas opciones tiene la novia de Juan?

Solución:

En vista de que sólo necesita un medio de transporte, podemos aplicar el principio de adición, de modo que cuenta con $1 + 3 + 6 = 10$ formas distintas de trasladarse



3.2 Ordenaciones, permutaciones y combinaciones

Es importante antes de abordar estos conceptos exponer el concepto de factorial.

Los factoriales nos ayudan a cuantificar rápidamente el número de maneras distintas en que se pueden acomodar **n** objetos en **n** lugares.

El factorial de n se denota como $n!$ y se lee n factorial. Se define como:

$$n! = n(n-1)(n-2)(n-3)\dots 1$$

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



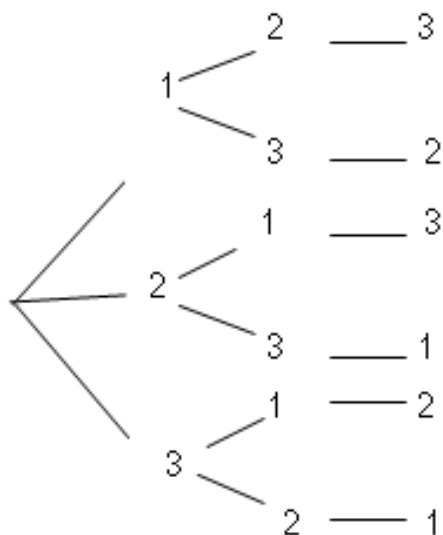
La información que proporciona el factorial es equivalente a la que obtendríamos de un diagrama de árbol, de hecho, podríamos decir que son diagramas de árbol simplificados. El tipo de posibilidades en que se utilice este método dependerá del orden en que se presenten las variables.

Un diagrama de árbol es una herramienta gráfica que tiene un punto de origen a partir del cual se abren ramas que representan las diferentes posibilidades, mismas que a su vez dan lugar a otras ramas, tantas como nuevas posibilidades haya, y así sucesivamente.

Ejemplo 1: Utilizar un diagrama de árbol para representar la situación que corresponde al caso 3!.



Solución:



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

$$3! = 3 * 2 * 1$$

$$3! = 6$$

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

Ejemplo 2: Obtener 5!

Solución:

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

$$5! = 5 * 4 * 3 * 2 * 1$$

$$5! = 120$$

Esto nos indica que si elaborásemos el árbol correspondiente éste tendría 120 ramas finales



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

Solución:

$$6! \cdot 3! \cdot 0! = 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 \cdot 3 \cdot 2 \cdot 1 \cdot 1$$

$$6! \cdot 3! \cdot 0! = 720 \cdot 6 \cdot 1$$

$$6! \cdot 3! \cdot 0! = 4320$$

Ejemplo 4: Dividir 10! entre 5!

Solución:

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

Ordenaciones

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

Ordenaciones en n objetos $n! = (n) \cdot (n-1) \cdot \dots \cdot (2) \cdot (1)$

Como ves, las ordenaciones de n objetos corresponden al factorial de n

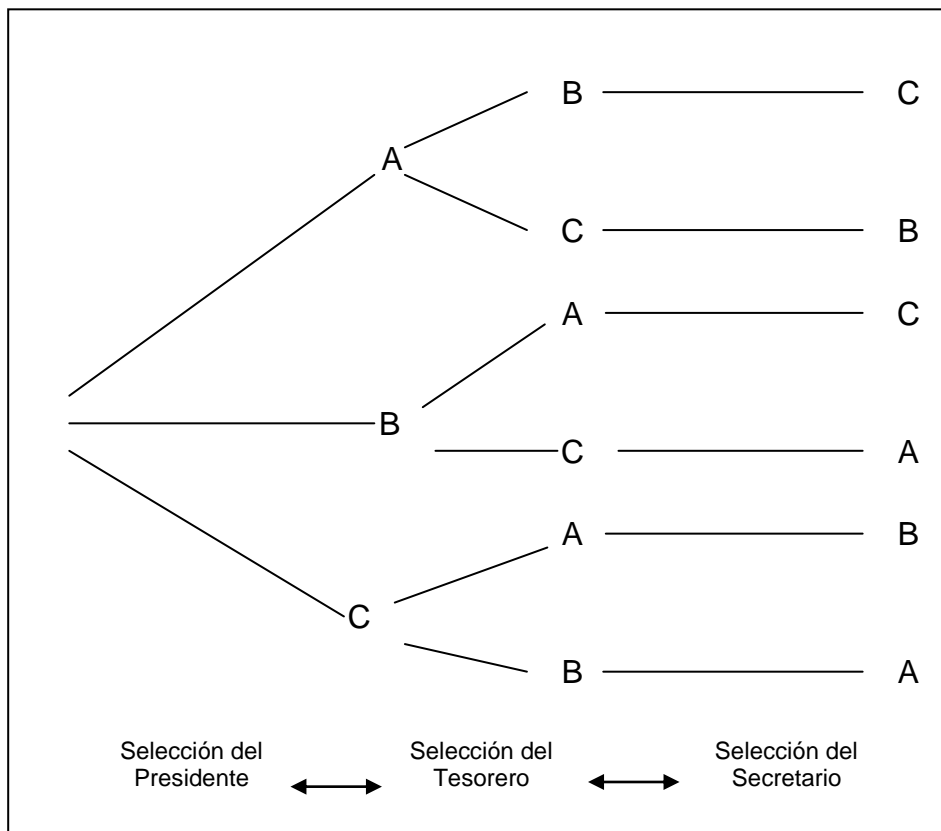


Ejemplo1: Tres miembros de una organización se han ofrecido a fungir, en forma voluntaria como presidente, tesorero y secretario. Obtener el número de formas en que los tres podrían asumir los puestos.

Solución:

$$n! = 3! = 3 \times 2 \times 1 = 6 \text{ formas}$$

Desde luego, de acuerdo a lo que hemos visto esto también se puede representar mediante un diagrama, poniéndoles letras a las tres personas A, B Y C.





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

Solución:

$$n! = 4! = 4 \times 3 \times 2 \times 1 = 24$$

Se pueden obtener 24 arreglos de los libros sin repetición.

Permutaciones

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

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

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

Solución:

Sabemos que la fórmula por aplicar es la anterior, por lo que al sustituir los datos correspondientes tenemos que:

Sustituyendo los datos correspondientes tenemos que:



$${}_{10}P_3 = \frac{10!}{(10-3)!} = \frac{10!}{7!} = \frac{10 \cdot 9 \cdot 8 \cdot 7!}{7!} = 10 \cdot 9 \cdot 8 = 720$$

Así, existen 720 formas diferentes de colocar a estos tres gerentes en tres de las diez plantas que posee la empresa.

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

Solución:

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

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

o bien podemos encontrar el resultado utilizando la fórmula:

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

de donde sustituyendo datos tenemos que:

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

y al aplicar la definición de factorial:

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



Finalmente el resultado se reduce a:

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

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

Combinaciones

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

$${}_n C_r = \frac{n!}{r!(n-r)!}$$

Ejemplo 3: Al auditar las 87 cuentas por pagar de una compañía, se inspecciona una muestra de 10 cuentas. ¿Cuántas muestras posibles hay? Suponiendo que 13 de las cuentas contienen un error, ¿cuántas muestras contienen exactamente dos cuentas incorrectas?

Solución:

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

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



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

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

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

$${}_{13}C_2 \times {}_{74}C_8 = 1,200,000,000,000$$

muestras con dos cuentas erróneas y ocho correctas.

Ejemplo 4: Si un club tiene 20 miembros, ¿Cuántos comités diferentes de cuatro miembros son posibles?

Solución:

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

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

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



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

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

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

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

Como veremos en la siguiente unidad, el enfoque clásico del cálculo de probabilidades de ocurrencia de determinados eventos, establece que el valor de probabilidad se determina en función de la cantidad de resultados igualmente probables y que sean favorables así como del número total de resultados posibles. Cuando los problemas son sencillos, el número de resultados puede contarse directamente. Sin embargo, para problemas o situaciones más complejas se requieren las herramientas del análisis combinatorio aquí estudiadas.



RESUMEN DE LA UNIDAD

El análisis combinatorio, como rama de las matemáticas, integra los principios y herramientas relativas a los métodos de conteo que apoyan el cálculo de probabilidades.



GLOSARIO

Análisis combinatorio

Rama de las matemáticas que se encarga del desarrollo de las técnicas y/o procedimientos para acomodar una colección de objetos o formar subconjuntos de la misma de acuerdo a reglas específicas, determinando cuántos de tales subconjuntos se pueden formar.

Combinaciones

Se refiere al número de formas en que pueden obtenerse grupos de k objetos tomados de una colección de n de ellos ($n \geq k$), considerando que todos los arreglos del mismo grupo de k objetos son indistinguibles entre sí, porque no importa el orden en que aparecen. Tal número se denota como nCk .

Factorial

El factorial de un número entero no negativo n se define como el número de maneras en que n objetos pueden ser permutados. Se denota como $n!$ y se obtiene mediante el producto de todos los números naturales desde el 1 hasta n , esto es, $n! = (1)(2)(3)\dots(n)$

Permutación

Es cada uno de los rearreglos o acomodados de los elementos de un conjunto. El término también hace referencia al número de formas en que pueden obtenerse grupos de k objetos tomados de una colección de n de ellos ($n \geq k$), considerando que el orden en que se toman los objetos distingue un arreglo de otro. Tal número se denota como nPk



ACTIVIDADES DE APRENDIZAJE

ACTIVIDAD 1

Supón que tres clientes de un restaurante olvidan en el interior de éste sus paraguas. La gerencia, que conoce de antaño a las tres personas, decide hacerles llegar sus paraguas, aunque no sabe cuál es el de cada quien, de modo que tendrá que escogerlos al azar. Se desea saber de cuántas formas puede ocurrir que

- a) nadie reciba el paraguas correcto
- b) dos de los clientes reciban el paraguas correcto
- c) los tres clientes reciban el paraguas correcto.

Elabora un texto en el que expliques de qué manera se puede responder a estas interrogantes, detallando cómo se aplicarían, de ser el caso, los principios de adición y multiplicación.



ACTIVIDAD 2

Considera una baraja de 24 cartas, con los siguientes valores en orden ascendente: 9, 10, J (jack), Q (reina), K (rey) y A (as). Además, cada carta muestra una de cuatro posibles figuras, a saber: espada, trébol, corazón y diamante.

De este modo, de cada valor hay cuatro figuras y de cada figura hay seis valores (por ejemplo, hay una reina de espadas, una de tréboles, una de corazones y una más de diamantes).

La tabla siguiente te muestra, para mayor claridad, la distribución de las cartas.

| Figura | Valor | | | | | Total |
|----------|-------|----|---|---|---|-------|
| | 9 | 10 | J | Q | K | |
| Espada | 1 | 1 | 1 | 1 | 1 | 6 |
| Trébol | 1 | 1 | 1 | 1 | 1 | 6 |
| Corazón | 1 | 1 | 1 | 1 | 1 | 6 |
| Diamante | 1 | 1 | 1 | 1 | 1 | 6 |
| Total | 4 | 4 | 4 | 4 | 4 | 24 |

A cada jugador se le entregan cinco cartas. Se desea saber el número de formas distintas que se tienen para formar:



- Un par, definido por dos cartas del mismo valor y las otras diferentes entre sí y al par (por ejemplo, las cartas 9, 9, J, Q y K definen un juego con par de nueves).
- Dos pares, definidos por dos grupos de cartas del mismo valor, pero diferente entre sí y la quinta carta de otro valor diferente a los de los dos pares (por ejemplo, las cartas 9, 9, J, J y K definen un juego con dos pares, uno de nueves y otro de jacks).
- Una tercia, definida por tres cartas del mismo valor y dos cartas de valores distintos entre sí y al valor que define la tercia (por ejemplo, las cartas 9, 9, 9, Q y A definen un juego con una tercia de nueves).
- Full, definido por una tercia y un par (por ejemplo, las cartas 9, 9, 9, Q y Q definen un juego con una tercia de nueves y un par de reinas).
- Póker, definido por cuatro cartas del mismo valor (por ejemplo, las cartas 9, 9, 9, 9 y K definen un juego con póker de nueves).
- Flor imperial, definido por cinco cartas de la misma figura y con valores sucesivos (por ejemplo, las cartas 9, 10, J, Q, K y A, todas ellas de espadas, definen una flor imperial de espadas).

Determina los valores que se solicitan. Explica en cada caso el procedimiento que seguiste.



CUESTIONARIO DE REFORZAMIENTO

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



LO QUE APRENDÍ

Encontrarás 8 ejercicios en la hoja de trabajo del documento Principios de Adición y Multiplicación, de la unidad didáctica Combinatoria, elaborada por Barrios Calmaestra, Luis para el proyecto Descartes, del Ministerio de Educación y Ciencia, España, cuya dirección es:

http://descartes.cnice.mecd.es/materiales_didacticos/Combinatoria/hojasdetrabajo/hoja2_principios.pdf

Forma un equipo de trabajo con un compañero de grupo. Escojan uno de los ejercicios de la página mencionada.

Desarrollenlo y preséntenlo en un blog. Revisen los comentarios y observaciones que se les haga llegar. Realicen las correcciones necesarias.



EXAMEN DE AUTOEVALUACIÓN

I. Anota en el espacio correspondiente tu respuesta.

Considera el experimento en que se arroja una moneda tres veces:

| | |
|--|--|
| a) ¿Qué principio utilizarías para determinar el número de arreglos o cadenas distintas de resultados? | |
|--|--|

| | |
|---------------------------------------|--|
| b) ¿Cuántos resultados distintos hay? | |
|---------------------------------------|--|

Tres personas llegan al elevador de un edificio de 8 pisos (descontando la planta baja). Determine de cuántas maneras las tres personas pueden:

| | |
|-------------------------------|--|
| a) Designar sus pisos destino | |
|-------------------------------|--|

| | |
|---|--|
| b) Designar sus pisos destino si los tres deben ir a pisos diferentes | |
|---|--|

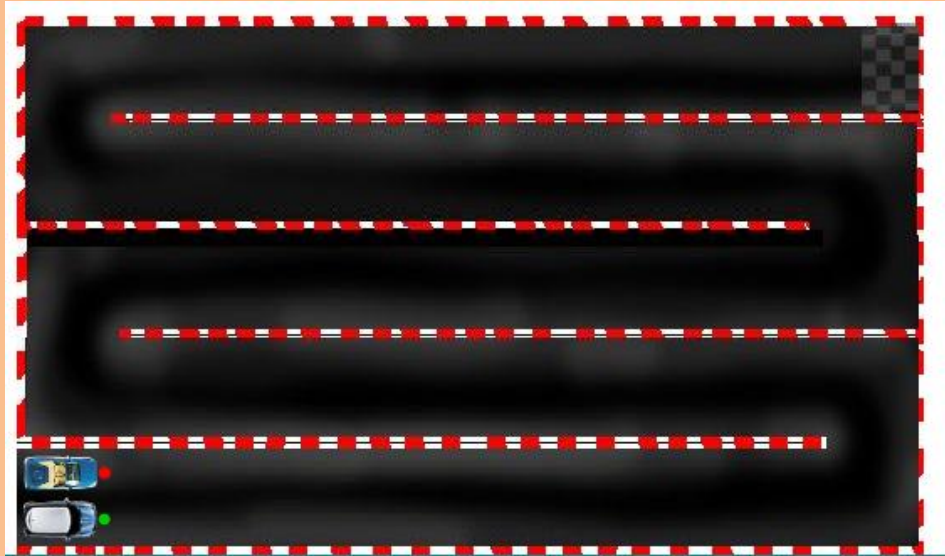


II. Indica cuáles de las siguientes aseveraciones son verdaderas y cuáles son falsas.

| | Verdadero | Falso |
|---|-----------------------|-----------------------|
| ${}_{36}C_4 = {}_{36}C_{32}$ | <input type="radio"/> | <input type="radio"/> |
| Se desea acomodar 6 libros en un estante donde sólo hay espacio para 4. Entonces se pueden escoger los dos libros que sobran de 6! maneras | <input type="radio"/> | <input type="radio"/> |
| ${}_{87}C_{80} = 5843355957$ | <input type="radio"/> | <input type="radio"/> |
| $({}_nC_r)(r!) = {}_nP_r$ | <input type="radio"/> | <input type="radio"/> |
| Si una moneda se arroja 7 veces, entonces hay 2^7 secuencias distintas de resultados | <input type="radio"/> | <input type="radio"/> |
| Se desea acomodar 6 libros en un estante donde sólo hay espacio para 4. Entonces se pueden escoger los dos libros que sobran de ${}_6C_4$ maneras | <input type="radio"/> | <input type="radio"/> |

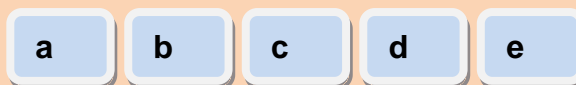


III. Una vez que has revisado la información de esta unidad te invitamos a que participes en este rally, si tu puntuación no es favorable, revisa nuevamente tus contenidos.



1. El producto de $4!$ por $3!$ Es igual a:

- a. 49
- b. 81
- c. 121
- d. 144
- e. 169



2. Diga si la siguiente igualdad es correcta:

$$n! = n(n + 1)(n + 2)\dots\dots 1$$

- a. Es correcta
- b. Es correcta sólo para $n = 0$
- c. Es correcta sólo para $n < 0$
- d. Incorrecta
- e. Se requieren más datos



a

b

c

d

e

3. Si un problema se puede resolver de tres maneras diferentes y otro problema se puede resolver de cinco formas distintas, ¿de cuántas maneras se pueden resolver ambos problemas?

- a. 15
- b. 120
- c. 81
- d. 720
- e. 1024

a

b

c

d

e

4. Si tres números se toman de dos en dos, ¿cuántas cantidades se pueden formar?

- a. 12
- b. 6
- c. 9
- d. 3
- e. 10

a

b

c

d

e



5. Para el cálculo de permutaciones de n objetos tomados de n en n diga si la siguiente fórmula es correcta: ${}_nP_n = n(n-1)(n-2)\dots 1$

- a. Es correcta
- b. Es correcta sólo para $n = 0$
- c. Es correcta sólo para valores negativos de n
- d. Es incorrecta
- e. Se requieren más datos

a b c d e

6. El valor de ${}_7C_4$ es:

- a. 128
- b. 56
- c. 28
- d. 35
- e. 121

a b c d e

7. ¿De cuántas maneras se puede formar un comité de cinco estudiantes si se tienen nueve candidatos?

- a. 45
- b. 135
- c. 128
- d. 99
- e. 126

a b c d e



8. Con un total de cinco profesores de matemáticas y siete de estadística se integra un comité donde deben participar dos de matemáticas y tres de estadística, ¿de cuántas formas puede formarse dicho comité?

- a. 365
- b. 350
- c. 35
- d. 175
- e. 208

a b c d e

9. General Motors de México ofrece cinco modelos de vehículos con tres tipos distintos de equipamiento, ¿cuántos modelos diferentes pueden ofrecerse a sus clientes?

- a. 720
- b. 15
- c. 30
- d. 45
- e. 120

a b c d e



10. Con siete administradores y cinco contadores se quiere formar un consejo que conste de cuatro administradores y tres contadores, ¿de cuántas maneras diferentes se puede integrar?

- a. 121,480
- b. 3,000,000
- c. 50,400
- d. 1,764,000
- e. 1,024,000

a b c d e



MESOGRAFÍA

BIBLIOGRAFÍA RECOMENDADA

| Autor | Capítulo | Páginas |
|---------------------------------|--|----------|
| 1. Anderson, Sweeney, Williams. | 4. Introducción a la probabilidad, Sección 4.1 Experimentos, reglas de conteo y asignación de probabilidades. | 135 -143 |
| 2. Lind, Marchal, Wathen, | 5. Estudio de los conceptos de la probabilidad Sección: Principios de conteo. | 165-170 |
| 3. Webster. | Capítulo 4. Principios de probabilidad, Sección 4.9 Técnicas de conteo | 93- 96 |



BIBLIOGRAFÍA BÁSICA

- Anderson, David R., Sweeney, Dennis J.; Williams, Thomas A., (2005). *Estadística para administración y economía*, 8ª edición, International Thompson editores, México, 888 páginas más apéndices.
- 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.
- Webster, Allen L. (2000), *Estadística aplicada a los negocios y la economía*, 3ª edición, Colombia, Irwin McGraw-Hill, 640 pp.

BIBLIOGRAFÍA COMPLEMENTARIA

1. Bowerman Bruce, *Pronósticos, series de tiempo y regresión; un enfoque aplicado*, México: Cengage Learning, 4ª edición, 2007, 720 pp.
2. Mendenhall William, *Introducción a la probabilidad y estadística*, México: Cengage Learning, 13ª edición, 2010, 776 pp.
3. Webster Allen L., *Estadística I aplicada a los negocios y la economía*, México: McGraw-Hill, 2ª. edición, 2002, 154 pp.



SITIOS DE INTERNET

- *Enumeración*, publicado como material didáctico para la materia Combinatoria de la licenciatura en ciencias matemáticas de la facultad de ciencias exactas y naturales de la Universidad de Buenos Aires, Argentina, en:

<http://www.dm.uba.ar/materias/optativas/combinatoria/2002/2/introd.pdf>.

- Dos Santos María José Marques, *Principios Fundamentales en el proceso de contar*, del Colegio de posgraduados de la Facultad de Estudios Superiores Zaragoza de la UNAM, México, en el sitio:

<http://colposfesz.galeon.com/est501/probabi/teo/cap302/cap302.htm>.

- Barrios Calmaestra, Luis, *¿Qué es la combinatoria?*, de la Unidad didáctica: Combinatoria, elaborada para el proyecto Descartes, del Ministerio de Educación y Ciencia, España, en el sitio:

http://descartes.cnice.mec.es/materiales_didacticos/Combinatoria/combinatoria.htm.



SUAYED UNA OPCIÓN PARA TI

UNIDAD 4

TEORÍA DE LA PROBABILIDAD

APUNTES DIGITALES PLAN 2011

SUAYED UNA OPCIÓN PARA TI



OBJETIVO ESPECÍFICO

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



INTRODUCCIÓN

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



LO QUE SÉ

En 1693, Samuel N. Pepys, quien había sido alto funcionario del Almirantazgo inglés, le solicitó a Isaac Newton su ayuda en torno a un problema de decisión cuyo sentido general era más o menos el siguiente:



“Me presentan tres sobres, cada uno con una tarjeta marcada con un número distinto. Los números son el 1, el 2 y el 3.

Me ofrecen dos alternativas:

- I. Extraer dos sobres con reemplazo. Gano si por lo menos una vez sale el número 3*
- II. Extraer cuatro sobres con reemplazo. Gano si por lo menos dos veces sale el número 3.*

¿Cuál alternativa es mejor?”

¿Tú que hubieras respondido?



TEMARIO DETALLADO

- 4.1 Interpretaciones de la probabilidad.
 - 4.1.1 Teórica o clásica.
 - 4.1.2 La probabilidad como frecuencia relativa.
 - 4.1.3 Interpretación subjetiva de la probabilidad.
- 4.2 Espacio muestral y eventos.
- 4.3 Los axiomas de la probabilidad.
- 4.4 La regla de la suma de probabilidades.
- 4.5 Tablas de contingencias y probabilidad condicional.
- 4.6 Independencia estadística.
- 4.7 La regla de multiplicación de probabilidades.
- 4.8 Teorema de Bayes.



4.1 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.

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.



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*.

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.

4.2 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 que 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 solo 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 $A \cap B$.



Eventos simultáneos.

A manera de resumen en la siguiente tabla te mostramos cuatro operaciones que serán muy útiles para manejar eventos aleatorios y su equivalencia con operaciones lógicas.



| Operación Lógica | Operación en conjuntos |
|------------------|---------------------------------------|
| o | Unión (U) |
| y | Intersección (\cap) |
| 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 $A \cup C$
- c) Probabilidad de A^c
- d) Probabilidad de $M \cap I$
- 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) $P(A \cup C)$
- c) $P(A^c)$
- d) $P(M \cap I)$
- e) $P(H - A^c)$

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



4.3 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 .

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

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 $N \rightarrow \infty$ 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.

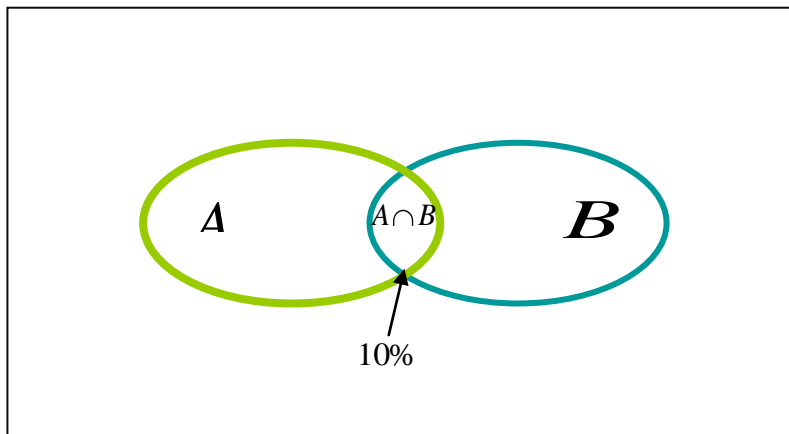
Ejemplo 1. En una investigación de mercado se encontró que entre los integrantes de un club, el 30% de los hombres usan loción para después de afeitarse, en tanto que el 40% de ellos utiliza desodorante y el 10% utiliza ambos productos. Si elegimos al azar a un varón de ese club, ¿qué probabilidades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

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 afeitarse, sabemos que tal pregunta en lenguaje probabilístico se transforma en:

$$P(A \cup B)$$



Intrínsecamente la pregunta se refiere a aquéllos 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, $P(A \cup B) = 30\% + 40\% - 10\%$.

$P(A \cup B) = 60\%$

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

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) $P(\phi) = 0$, 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),

$$\begin{aligned} P(A \cup A^c) &= P(A) + P(A^c) - P(A \cap A^c), \\ \rightarrow 1 &= P(A) + P(A^c) - P(A \cap A^c), \end{aligned}$$

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,

$$P(A \cup B) = 2/11 + 6/11 - 1/11$$

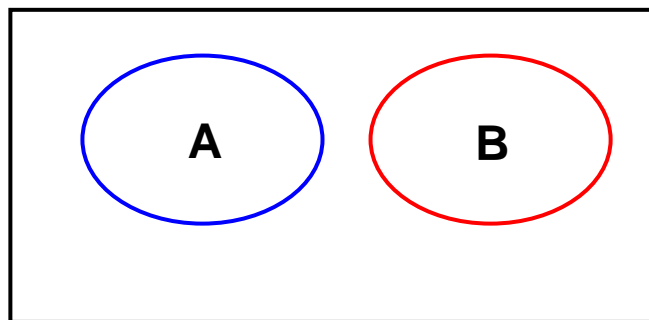
$$P(A \cup B) = 7/11$$



4.4 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) $P(A \cap B)$



b) $P(A \cup B)$

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, $P(A \cap B) = 0$.

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.

$$P(A \cup B) = P(A) + P(B) \quad \text{si } A \cap B = \phi$$

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, el 20% de los socios pertenece al equipo de natación y el 10% al equipo de waterpolo. Ningún socio pertenece a ambos equipos simultáneamente. Diga 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.

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



4.5 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 avocaremos 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 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 del 60%.

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

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

Es decir, la probabilidad de B dado que A ya ocurrió es igual a la probabilidad de que ocurran ambos eventos simultáneamente (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 el 50%.

Deseamos averiguar lo siguiente.

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

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:

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

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



En el inciso (b) deseamos conocer 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 el 60% de los días y despejado el 40% de ellos, podemos sustituir en la fórmula.

$$P(B/A^c) = \frac{0.10}{0.40} = 0.25 = 25\%$$

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

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:

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

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

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

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

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

$$P(M | E_2) = \frac{P(E_2 \cap M)}{P(E_2)} = \frac{0.197}{0.337} = 0.585 = 58.5\%$$

f)

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:

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

o sea:

$$P H = 0.117 + 0.140 + 0.170 = 0.427 = 42.7\%$$



4.6 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:

$$P(B/A) = P(B) \quad \text{y} \quad P(A/B) = P(A)$$

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,

$$P(S|J) = 40 / 70$$

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.

4.7 La regla de multiplicación de probabilidad

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

4.8 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.

$$P A_i / B = \frac{P B / A_i \cdot P A_i}{P B / A_1 \cdot P A_1 + P B / A_2 \cdot P A_2 + \dots + P B / A_k \cdot P A_k}$$

En donde:

| | |
|---|--|
| $P A_i =$ Probabilidad previa | Es la probabilidad de un evento posible antes de cualquier otra información. |
| $P B / A_i =$ Probabilidad condicional | Es la probabilidad de que el evento "B" ocurra en cada posible suceso de A_i . |
| $P B / A_i \cdot P A_i =$ Probabilidad conjunta | Equivalente a la probabilidad de $A_i \cap B$ determinada por la regla general de la multiplicación. |



| | |
|---|--|
| $P(A_i B) =$ Probabilidad a <i>posteriori</i> | 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. D = Clientes dueños de sus casas.



P_2 = Clientes pagan tarde.

D' = Clientes no

son dueños de sus casas

P_3 = Clientes que no pagan.

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:

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

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



b) Probabilidad de que un nuevo solicitante pague tarde:

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

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

c) Probabilidad de que un nuevo solicitante no pague.

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

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

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

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

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



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

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

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 DE LA UNIDAD

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.



GLOSARIO

Axiomas de probabilidad

Son los postulados básicos sobre los que se construyó la teoría moderna de la probabilidad. Los axiomas establecen que:

- La probabilidad de todo evento es por lo menos cero
- La probabilidad de que algo pase es 1
- Si se tienen dos eventos mutuamente excluyentes, la probabilidad de ocurra alguno de ellos es la suma de sus respectivas probabilidades

Ensayo

Cada realización del experimento aleatorio.

Espacio muestral

Es el conjunto de todos los valores que pueden resultar de la realización de un experimento aleatorio.

Evento

Es un subconjunto del espacio muestral por lo que está formado por resultados del experimento aleatorio.



Eventos independientes

Se dice que se tiene una colección de eventos independientes si la probabilidad conjunta de todos ellos es igual al producto de las probabilidades marginales.

Eventos mutuamente excluyentes

Se dice que una colección de eventos es mutuamente excluyente si el hecho de que uno de ellos ocurra impide o excluye que cualquier otro evento ocurra.

Experimento aleatorio

Es una situación en la que no puede anticiparse con certeza el resultado.

Probabilidad

Es un número que expresa las oportunidades o chance que tiene una situación de ocurrir o no.



Probabilidad condicional

Expresa el valor relativo que tiene la probabilidad de un subconjunto del espacio muestral ante la probabilidad de otro evento denominado condición y que se supone ya ha ocurrido.

Probabilidad marginal

Es la probabilidad ordinaria o simple de un evento individualmente considerado.

Regla de Bayes

Establece que la probabilidad condicional de un evento X dado que la condición Y se ha verificado (ha ocurrido) se puede expresar en términos de la probabilidad condicional del evento Y dado que la condición X ha ocurrido.



ACTIVIDADES DE APRENDIZAJE

ACTIVIDAD 1

Considera la siguiente situación.

Deseas trasladarte a un cierto destino para lo cual debes abordar un autobús. Te diriges a la parada más cercana y esperas. Entonces te das cuenta que en esa parada pasan autobuses de 3 distintas rutas; dos de ellas te llevan a tu destino, la otra te dejaría muy lejos de éste.

Supón que en un lapso de una hora pasan por esa parada 30 autobuses y que no hay una secuencia predeterminada de rutas. ¿Cuáles de las siguientes afirmaciones te parecen válidas?

- “Como máximo tendré que esperar al segundo autobús”
- “Hay una probabilidad de un medio (0.5) de que el primer autobús sea de la ruta que no me conviene”
- “Dos de cada tres autobuses pasan cerca de donde yo tengo que ir”
- “La probabilidad de que tenga que esperar hasta cuatro autobuses es mayor a 10%”



ACTIVIDAD 2

Considera la situación que se te presenta en la fotografía que se muestra a continuación.

Construye y define en torno a ella tres eventos probabilísticos.



ACTIVIDAD 3

Dados dos eventos A y B respecto de los cuales se sabe que $P(A)=0.3$, $P(B^c)=0.4$ y $P(A \cup B)=0.7$, determina los valores que se solicitan a continuación. Para obtener las respuestas puedes auxiliarte con diagramas de Venn-Euler.

- a) $P(B)$
- b) $P(A^c)$
- c) $P(A - B)$
- d) $P(A \cap B)$
- e) $P(B - A)$
- f) $P[(A \cup B)^c]$
- g) $P[(A \cap B)^c]$



ACTIVIDAD 4

Elabora un mapa mental con los conceptos básicos de probabilidad que has estudiado hasta ahora. Tu mapa debe contener por los menos los conceptos de evento, evento universo, probabilidad marginal y conjunta, eventos mutuamente excluyentes, así como las escuelas o paradigmas de probabilidad, entre otros.

ACTIVIDAD 5

Mediante el empleo de los diagramas de Venn-Euler da una interpretación intuitiva de las siguientes dos relaciones:

- a) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
- b) Si A y B son mutuamente excluyentes, $P(A \cup B) = P(A) + P(B)$

Analiza la situación para el caso de tres eventos A, B y C, apoyándote, si lo crees necesario, en los diagramas de Venn-Euler y desarrolla una expresión que permita determinar $P(A \cup B \cup C)$ en función de $P(A)$, $P(B)$, $P(C)$, $P(A \cup B)$, $P(A \cup C)$, $P(B \cup C)$ y $P(A \cap B \cap C)$. Explica qué ocurre si A, B y C son mutuamente excluyentes.



ACTIVIDAD 6

En un estudio de hábitos de lectura de periódico se clasificó a las personas en tres grupos de edad, a saber, el grupo J integrado por personas de 18 a 30 años de edad, el grupo A por aquéllos de 31 a 45 y el grupo M por aquéllos de más de 45 años de edad. Por otro lado, se encontró que 21 personas del grupo M leen el periódico Cambio Ligero, otros 26 leen este mismo periódico pero están en el grupo A de edad. De los que leen el periódico El Infinito, 4 están en el grupo J de edad, 12 en el grupo A y 24 en el grupo M. En total, 61 personas leen el periódico Cambio Ligero y otros 53 leen El Apalancamiento. Además hay 26 personas en total en el grupo J y 68 en el grupo A. Se desea conocer la probabilidad de que si se extrae a una persona al azar, ésta.

Sea del grupo A

Lea El Apalancamiento

Sea del grupo M

Sea del grupo M y lea El Infinito

Lea El Infinito

Lea El Infinito si es del grupo M

Sea del grupo M si lee El Infinito



ACTIVIDAD 7

Descarga la actividad, realízala en un procesador de textos, guárdala en tu computadora y una vez concluida, presiona el botón **Examinar**. Localiza el archivo, ya seleccionado, presiona **Subir este archivo** para guardarla en la plataforma.

Dar click para descargar la actividad: **unidad 4_tema5_actividad 2**

Una compañía de seguros está revisando sus estadísticas sobre siniestralidad en el ramo de automóviles. Con este motivo se toma una muestra de 145 pólizas que en el año pasado registraron por lo menos un siniestro. Uno de los puntos del estudio requiere revisar la relación entre dos variables, de un lado la variable edad del conductor (E) y del otro la suma cubierta por la aseguradora (S).

De las 145 pólizas, 59 pertenecen a conductores de 18 a 24 años de edad, de las cuales en 31 se pago un daño de menos de \$10,000, en otras 10 se pago un daño de \$30,000 a \$99,999.99 y en otras 4 de \$100,000 a \$300 000. Del grupo de edad de entre los 25 y los 35 años de edad, hubo 30 pólizas con daños menores a los \$10,000 y otras 18 con daños entre \$10,000 y \$30,000. Otras 6 pólizas con daños reportados entre los \$10,000 y los \$30,000 eran de conductores de más de 35 años. En este último grupo de edad, hubo además 4 pólizas con daños entre \$30,000 y \$100,000 y una póliza con un daño reportado de entre \$100,000 y \$300,000. En total, hubo 26 pólizas con montos entre los \$30,000 y los \$100,000 y 23 que pertenecían a conductores de más de 35 años.

Suponga que se selecciona al azar una póliza. Se desea saber ...



- a) ¿Cuál es la probabilidad de que la suma pagada haya sido una cantidad entre \$10,000 y \$30,000?
- b) ¿Cuál es la probabilidad de que el conductor tenga una edad registrada menor a 35 años?
- c) ¿Cuál es la probabilidad de que el conductor tenga entre 18 y 24 años y la suma pagada sea mayor a \$100,000?
- d) ¿Cuál es la probabilidad de que la suma pagada no exceda los \$30,000, si se sabe que la edad del conductor está entre 25 y 35 años?

Contesta las interrogantes anteriores, elaborando para el efecto una tabla de contingencia. Escribe en los espacios en blanco tu respuesta.

| EDAD | SUMA PAGADA POR SINIESTRO | | | | Total |
|-----------|---------------------------|------------------------|-------------------------|--------------------------|-------|
| | Hasta \$10,000 | De \$10,000 a \$30,000 | De \$30,000 a \$100,000 | De \$100,000 a \$300,000 | |
| 18-24 | 31 | | 10 | 4 | 59 |
| 25-35 | 30 | 18 | | | |
| Más de 35 | | 6 | 4 | 1 | 23 |
| Total | | | 26 | | 145 |



ACTIVIDAD 8

Considera una tabla de contingencia cualquiera de dos renglones y dos columnas (aparte del renglón y columna de totales), como la que se muestra a continuación.

| | Evento X | Evento Y | Total |
|----------|----------|----------|-------|
| Evento A | | | |
| Evento B | | | |
| Total | | | |

Explica por qué si los eventos A y X son independientes probabilísticamente, las parejas de eventos A y Y, B y X así como B e Y también son independientes probabilísticamente.

ACTIVIDAD 9

Una organización civil que agrupa a profesionistas de tres diferentes licenciaturas, L1, L2 y L3, va a elegir a su nuevo Presidente. Hay tres candidatos, C1, C2 y C3.

Se tienen los siguientes datos:

- 1) Hay 350 miembros en la agrupación
- 2) La probabilidad condicional de que al seleccionar al azar a un miembro de la agrupación éste sea un profesionista de la licenciatura 1, dado que apoya al candidato 1 es 0.20
- 3) Ningún profesionista de la licenciatura 3 apoya al candidato 2



- 4) Si se selecciona al azar a un miembro de la agrupación, la probabilidad de que sea un profesionalista de la licenciatura 3 es 0.40
- 5) El candidato 1 cuenta con 30% de las preferencias
- 6) Hay independencia probabilística entre L2 y C1
- 7) Si se selecciona al azar a un miembro de la agrupación, la probabilidad de que sea un profesionalista de la licenciatura 1 que apoye al candidato 3 es 0.12
- 8) Los egresados de la licenciatura 2 constituyen el 20% del total de la agrupación
- 9) Si se selecciona al azar a un miembro de la agrupación, la probabilidad de que sea un profesionalista de la licenciatura 3 o de que apoye al candidato 2 es 0.24

Con los datos que se te han proporcionado, completa la siguiente tabla, anotando el número de casos que corresponde en cada celda:

| Licenciatura | Candidato 1 | Candidato 2 | Candidato 3 | Total |
|--------------|-------------|-------------|-------------|-------|
| 1 | | | | |
| 2 | | | | |
| 3 | | | | |
| Total | | | | |

Cuando tengas tus respuestas, incorpora la tabla en un archivo tipo texto, junto con los desarrollos que hayas realizado.



ACTIVIDAD 10

La gerencia de una empresa de publicidad ha solicitado a sus dos especialistas A y B, que le presenten sus respectivos proyectos para la campaña publicitaria de un nuevo producto. Para que la decisión respecto de cuál proyecto apoyar sea imparcial se les ha solicitado que los remitan bajo seudónimo. De experiencias anteriores se sabe que un 45% de los proyectos de A son aprobados mientras que para B la cifra correspondientes es 60%. Si ya se seleccionó al proyecto ganador, ¿cuál es la probabilidad de que sea el proyecto de B?

ACTIVIDAD 11

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

ACTIVIDAD 12

Se arrojan tres dados al mismo tiempo. Se sabe que considerados por pares, las caras que mostraron los dados no fueron iguales.

Determina las siguientes probabilidades.

- a) Probabilidad de que haya salido una vez el número 2
- b) Probabilidad de que la suma sea 5
- c) Probabilidad de que haya salido el número 4 si la suma es 12



CUESTIONARIO DE REFORZAMIENTO

1. Indique la diferencia entre una probabilidad frecuencial y una probabilidad subjetiva.
2. ¿Cuáles son los pasos del procedimiento para calcular la probabilidad simple de un evento?
3. ¿Cuál es la diferencia entre eventos excluyentes y eventos independientes?
4. Explique las características de la regla de la adición.
5. Defina las propiedades de una probabilidad condicional.
6. Explique las características de la regla de la multiplicación.
7. ¿En qué consiste una tabla de probabilidad conjunta?
8. ¿En qué consiste una tabla de contingencia?, ¿cuál es su relación con una tabla de probabilidades?
9. ¿A qué se refiere cuando se habla de una probabilidad marginal?
10. ¿Cuáles son los objetivos de un teorema de Bayes?, ¿qué tipo de probabilidades intervienen?



LO QUE APRENDÍ

Al inicio de la Unidad, te planteamos el problema que le consultó Samuel N. Pepys a Isaac Newton. Con los conocimientos que has adquirido a lo largo de estas secciones desarrolla un planteamiento más formal que te permita contestar la pregunta planteada. Para tu comodidad aquí anotamos nuevamente el texto del problema.



“Me presentan tres sobres, cada uno con una tarjeta marcada con un número distinto. Los números son el 1, el 2 y el 3.

Me ofrecen dos alternativas:

I. Extraer dos sobres con reemplazo. Gano si por lo menos una vez sale el número 3

II. Extraer cuatro sobres con reemplazo. Gano si por lo menos dos veces sale el número 3.

¿Cuál alternativa es mejor?”



EXAMEN DE AUTOEVALUACIÓN

I. Determina si las siguientes aseveraciones son verdaderas (V) o falsas (F). Una vez que termines obtendrás tu calificación de manera automática.

| | Verdadero | Falso |
|---|-----------|-------|
| 1. La probabilidad frecuentista está vinculada a la recopilación de datos. | () | () |
| 2. Se puede decir que hay tres definiciones de probabilidad por lo menos. | () | () |
| 3. En la escuela frecuentista no importa si el experimento o situación se repite bajo condiciones distintas. | () | () |
| 4. La probabilidad subjetiva se puede definir como el grado de certidumbre que tiene un observador respecto de que pase algo. | () | () |
| 5. La frase "Hasta no ver no creer" es aplicable por entero a la escuela clásica. | () | () |
| 6. Es en la escuela clásica donde se encuentra | () | () |



| | | |
|--|-----|-----|
| un campo natural de aplicación del análisis combinatorio. | | |
| 7. Se requiere un número pequeño de ensayos para determinar una probabilidad frecuentista. | () | () |

II. Relaciona las siguientes dos columnas anotando en el paréntesis del lado derecho la letra que corresponda. Una vez que termines obtendrás tu calificación de manera automática.

| Si A, B y C tres eventos aleatorios cualquiera, entonces... | | |
|---|---------------------------------|-----|
| a) $P[A \cup (B \cap C)]$ es equivalente a... | $P(A)$ | () |
| b) $P[B \cap (A \cup C)]$ es equivalente a... | $P(A \cap B)$ | () |
| c) $P(B - A^c)$ es equivalente a... | $P[(A \cup B) \cap (A \cup C)]$ | () |
| c) $P[B \cup (B \cap A)]$ es equivalente a... | $P[(B \cap A) \cup (B \cap C)]$ | () |
| d) $P\{(A \cap B) \cup [(A \cap C) - (A \cap B \cap C)]\}$ es equivalente a... | $P(B)$ | () |



III. Considera la siguiente situación y calcula las probabilidades que se solicitan. Anota tus respuestas en la columna derecha del cuadro.

En una urna hay 7 bolas rojas, 3 azules y 5 verdes. Si se saca una bola al azar, ¿cuál es la probabilidad de que ésta ...

| | |
|----------------------|--|
| sea roja? | |
| no sea verde? | |
| sea roja o verde? | |
| sea azul y verde? | |
| no sea azul ni roja? | |



IV. Para un experimento se han definido dos eventos, A y B, respecto de los cuales se sabe que $P(A^c) = 0.4$ y $P(B) = 0.6$. Se sabe también que $P(A \cup B) = 0.8$. Determina los valores que se solicitan a continuación. Te sugerimos que elabores los diagramas de Venn-Euler.

Escribe tus respuestas en la zona sombreada con notación decimal.

| | |
|------------------------------|--|
| $P(A)$ | |
| $P(B^c)$ | |
| $P(A \cap B)$ | |
| $P(A - (A \cap B))$ | |
| $P(A^c \cap B^c)$ | |
| $P((A \cap B) \cap B)$ | |
| $P(A^c \cap B)$ | |
| $P((A \cup B) - (A \cap B))$ | |
| $P(A^c \cup B^c)$ | |



V. En una encuesta de opinión aplicada a 200 personas sobre sus preferencias en materia de color de automóviles, se obtuvieron los siguientes resultados:

De los 100 hombres entrevistados, 20 preferían el color rojo, 10 el color azul y 60 el negro. Además, 15 eran menores de 30 años y de ellos, tres preferían el rojo; otros 15 tenían entre 30 y 40 años y de ellos tres preferían el rojo, tres más preferían el negro y otros cinco preferían el blanco; otros 50 hombres tenían más de 50 años y de ellos, dos preferían el blanco, 38 preferían el negro y 8 más preferían el rojo. Otros dos hombres preferían el azul y estaban en el rango de edad de 40 a 50 años de edad y uno más de este mismo grupo de edad prefería el blanco.

En el caso de las mujeres, ocurrió que 40 eran menores de 30 años, de las cuales, tres preferían el negro y 32 preferían el blanco; además, 15 más tenían entre 40 y 50 años de edad, de las cuales dos preferían el azul y cuatro más el blanco; otras 25 tenían más de 50 años y de ellas, seis preferían el rojo y dos más preferían el negro. Por otro lado, un total de 15 mujeres preferían el rojo, 10 preferían el azul y 15 el negro. Nueve mujeres que preferían el blanco tenían entre 30 y 40 años de edad y en este mismo grupo de edad había otras cuatro mujeres que preferían el negro y otras tres que preferían el rojo.

En relación a las siguientes preguntas, anota tu respuesta en el cuadro del lado derecho.

Si se selecciona una persona al azar, ...

| | |
|--|--|
| 1. ¿cuál es la probabilidad de que sea mujer si se sabe que prefiere el rojo o el negro? | |
|--|--|



| | |
|--|--|
| | |
| 2. y resulta que ésta prefiere el color blanco, ¿cuál es la probabilidad de que sea hombre y tenga entre 40 y 50 años de edad? | |
| 3. ¿cuál es la probabilidad de que prefiera el azul y tenga entre 30 y 40 años de edad? | |
| 4. ¿cuál es la probabilidad de que prefiera el rojo o tenga menos de 30 años de edad? | |



VI. Se tienen dos eventos A y B, respecto de los cuales se sabe que:

$$P(A) = 0.3$$

$$P(B^c) = 0.4$$

$$P(A \cup B) = 0.7$$

Con estos datos, calcula las siguientes probabilidades anotando tu respuesta en el cuadro de la derecha:

| | |
|-----------------|--|
| $P(B) =$ | |
| $P(A \cap B) =$ | |
| $P(A B) =$ | |
| $P(B A^c) =$ | |

Sean A y B dos eventos tales que $P(B^c) = 0.42$ y $P(A \cup B) = 0.63$.

VII. Determina las siguientes probabilidades.

| | |
|-------------------------------------|--|
| $P(A)$ si A y B son independientes | |
| $P(B)$ | |
| $P(A - B)$ si B está contenido en A | |



VII. Considera la siguiente situación:

Una empresa de servicios turísticos coloca en Internet una convocatoria para participar en un concurso. Quienes lo deseen se registran a través de la red y se les cita a presentarse en las instalaciones de la empresa, todos a la misma hora, para llevar a cabo el concurso. Para la semana siguiente se han registrado J, A, R y C. El concurso consiste en contestar una pregunta. El ganador es el primero que conteste correctamente. El orden en que se formula a cada quien su respectiva pregunta se determina al azar. En este orden de ideas, se sabe que la probabilidad de que el ganador se defina a la primera pregunta es $1/10$, de que se defina a la segunda es $2/10$, de que se defina a la tercera es $3/10$ y de que se defina hasta la cuarta pregunta es $4/10$.

Se desea determinar las siguientes probabilidades:

- De que gane J
- De que gane A
- De que no gane nadie
- De que gane R en la segunda pregunta

Anota tus respuestas en los cuadros del lado derecho de la siguiente tabla.

| Evento | Probabilidad |
|-----------------------------------|--------------|
| que gane J | |
| que gane A | |
| que no gane nadie | |
| que gane R en la segunda pregunta | |



MESOGRAFÍA

BIBLIOGRAFÍA BÁSICA

| Autor | Capítulo | Páginas |
|---------------------------------|--|------------|
| 1. Anderson, Sweeney, Williams | 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 |
| 2. Berenson, Levine y Krehbiel. | 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. |
| 3. Levin y Rubin. | 4. Probabilidad I: Ideas introductorias Sección: | 129-131 |



| | | |
|---------------------------|--|-----------|
| | 4.2 Terminología básica en probabilidad. | |
| | 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 |
| 4. Lind, Marchal, Wathen. | 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 |



BIBLIOGRAFÍA BÁSICA

1. Anderson, David R.; Sweeney, Dennis J.; Williams, Thomas A., (2005). *Estadística para administración y economía*, 8ª edición, México, International Thomson Editores, 888 páginas más apéndices.
2. Berenson, Mark L., David M. Levine, y Timothy C. Krehbiel, (2001), *Estadística para administración*, 2ª edición, México, Prentice Hall, 734 páginas.
3. Levin, Richard I. y David S Rubin, (2004), *Estadística para administración y economía*, 7a. Edición, México, Pearson Educación Prentice Hall, 826 páginas más anexos.
4. 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.



BIBLIOGRAFÍA COMPLEMENTARÍA

4. Bowerman Bruce, *Pronósticos, series de tiempo y regresión; un enfoque aplicado*, México: Cengage Learning, 4ª edición, 2007, 720 pp.
5. Mendenhall William, *Introducción a la probabilidad y estadística*, México: Cengage Learning, 13ª edición, 2010, 776 pp.
6. Webster Allen L., *Estadística I aplicada a los negocios y la economía*, México: McGraw-Hill, 2ª. edición, 2002, 154 pp.



SUAYED
UNA OPCIÓN
PARA TI

SITIOS DE INTERNET

- *Annabolika (seudónimo), Historia de las probabilidades, en el sitio, <http://www.slideshare.net/AnnaBolika/historia-de-las-probabilidades-presentation-603669>.*
- Sierra Cinos, José Luis y García Diz, Luis (Profesores de la Universidad Complutense de Madrid), *Introducción a la probabilidad*, del curso de especialización en Bioestadística, impartido por, España, en:

<http://nutriserver.com/Cursos/Bioestadistica/Probabilidad.html>.
- La Gacetilla matemática de España, *Probabilidad total. Teorema de Bayes*, en el sitio: <http://www.arrakis.es/~mcj/azar10.htm>.
- La Gacetilla matemática de España, *Tablas de contingencia*, en:
<http://www.arrakis.es/~mcj/azar11.htm>.



SUAYED
UNA OPCIÓN
PARA TI

- Luna Gándara Rita, *Teorema de Bayes*, (forma parte de los apuntes del curso de Probabilidad y Estadística del departamento de ingeniería industrial del Instituto Tecnológico de Chihuahua), México, en el sitio:

http://www.itch.edu.mx/academic/industrial/sabaticorita/_private/08Teorema%20de%20bayes.htm

- García Ben Marta, *Teoría de la probabilidad*, secciones 2.1 a 2.5 (del departamento de matemáticas), Facultad de Ciencias Exactas y Naturales de la Universidad de Buenos Aires, Argentina, en el sitio:

http://www.dm.uba.ar/materias/estadistica_Q/2008/1/EstadQuimProbabilidad.pdf

- *Independencia estocástica de sucesos*, página del proyecto CEACES de la Universidad de Valencia, España, en donde se presenta el concepto de independencia estocástica (probabilística) y sus implicaciones, en el sitio:

<http://www.uv.es/ceaces/base/probabilidad/independen.htm>



SUAYED UNA OPCIÓN PARA TI

UNIDAD 5

DISTRIBUCIONES DE PROBABILIDAD

APUNTES DIGITALES PLAN 2011

SUAYED UNA OPCIÓN PARA TI



OBJETIVO

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



SUAYED
UNA OPCIÓN
PARA TI

INTRODUCCIÓN

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

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



LO QUE SÉ

Según cifras publicadas por **Indexmundi**, la esperanza de vida al nacimiento en México en el año 2007 es de 75.63 años.

En este orden de ideas, es posible que para el año 2050 sea de 80 años. ¿Qué opinas de ello?



TEMARIO DETALLADO

- 5.1 Variables aleatorias, discretas y continuas.
- 5.2 Media y varianza de una distribución de probabilidad.
- 5.3 Distribuciones de probabilidad de variables discretas.
 - 5.3.1 Distribución binomial.
 - 5.3.2 Distribución de Poisson.
 - 5.3.3 La distribución de Poisson como aproximación de la distribución binomial.
 - 5.3.4 Distribución hipergeométrica.
 - 5.3.5 Distribución multinomial.
- 5.4 Distribuciones de probabilidad de variables continuas.
 - 5.4.1 Distribución normal.
 - 5.4.2 Distribución exponencial.
- 5.5 Ley de los grandes números.



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.





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:

$$\text{Ahorro} = \text{ingreso} - \text{gasto}$$

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.



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 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 corresponden 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, $P(X \leq 7)$ 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 $P(X \leq i)$ |
|----|---------------------------------------|--|
| 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

$$\text{Esperanza} = \mu = \sum x \mathbf{P}(X = x)$$

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.



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

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.



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.

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 ésta 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 (el alumno debe recordar el tema de reglas de conteo) por la probabilidad de éxito elevada al número de éxitos deseado y por la probabilidad de fracaso elevada al número de fracasos deseados.



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 ($n \geq k$), 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,

$$nCk = n! / [k! (n-k)!]$$

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

$$n! = (1)(2)(3)\dots(n)$$

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, indique usted 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 $P(x) = nC_x p^x q^{(n-x)}$, 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. Diga usted 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:

- ¿Cuál será la probabilidad de que exactamente dos puedan localizar este país?
- ¿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,

$$\begin{aligned} P(X \leq 2) &= P(2) + P(1) + P(0) = \\ &= {}_6C_2 (0.30)^2(0.70)^4 + {}_6C_1 (0.30)^1(0.70)^5 + {}_6C_0 \\ &(0.30)^0(0.70)^6 = \\ &= 15(0.09)(0.2401) + 6(0.30)(0.16807) + 0.1176 = \\ &= 0.661941 \end{aligned}$$



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:

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



| 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 325125 veces encontraremos un refrigerador con defecto, en otras 57375 ocasiones encontraremos dos refrigeradores con defecto y en 3375 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 | 614125 | 0 |
| 1 | 325125 | 325125 |
| 2 | 57375 | 114750 |
| 3 | 3375 | 10125 |
| Total | 1000000 | 450000 |



Luego,

$$\mu = 450\,000 / 1\,000\,000 = 0.45$$

Asimismo, podemos calcular la varianza:

$$\begin{aligned}\sigma^2 &= [614125 (0-0.45)^2 + 325125 (1-0.45)^2 + 57375 (2-0.45)^2 + 3375 (3-0.45)^2] / 100 \\ &= (124360.313 + 98350.3125 + 137843.438 + 29945.9375) / 100 \\ &= 0.3825\end{aligned}$$

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

- $\mu = 0.45$ 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

$$\mu = np \quad \sigma^2 = n$$



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:

$$nC_x = \frac{n!}{x! (n-x)!}$$

$$p^x = p^x$$

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

$$\text{Media} = np$$

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

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.

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 m². 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:

$$P_{x/\lambda} = \frac{\lambda^x \cdot e^{-\lambda}}{x!}$$

En esta fórmula, x representa el número de éxitos cuya probabilidad deseamos calcular; λ es el promedio de éxitos en un 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

$$40 - 400$$

$$\lambda - 25$$

$$\therefore \lambda = 2.5$$

$$x = 2$$

$$e = 2.71828$$

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

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

b) Más de dos errores en 40 páginas



Datos

$$40 - 400$$

$$\lambda - 40$$

$$\therefore \lambda = 4$$

$$x = 2$$

$$e = 2.71828$$

$$P_{0/4} = \frac{4^0 \cdot 2.71828^{-4}}{0!} = 0.018 = 1.8\%$$

$$P_{1/4} = \frac{4^1 \cdot 2.71828^{-4}}{1!} = 0.073 = 7.3\%$$

$$P_{2/4} = \frac{4^2 \cdot 2.71828^{-4}}{2!} = 0.146 = 14.6\%$$

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

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

c) Una página no tenga errores:

Datos

$$40 - 400$$

$$\lambda - 1$$

$$\therefore \lambda = 0.10$$

$$x = 2$$

$$e = 2.71828$$

$$P_{0/0.10} = \frac{0.10^0 \cdot 2.71828^{-0.10}}{0!} = 0.905 = 90.5\%$$



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

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

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.

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

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

$$P_{x/\mu} = \frac{\mu^x \cdot e^{-\mu}}{x!}$$

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

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

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



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

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.

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 una 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:

$$P(x) = \frac{\binom{C_x}{k} \binom{C_{n-x}}{N-k}}{N \binom{C_n}{n}}$$

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



$$\begin{array}{l} n=10 \\ k=18 \\ x=5 \end{array} \Rightarrow P(5) = \frac{{}^8C_5 \cdot {}^{17}C_5}{{}^{35}C_{10}} = 0.2888$$

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, x_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:

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

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 dicho 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:

$$N=14$$

$$p_1 = 0.40$$

$$p_2 = 0.35$$

$$p_3 = 0.25$$

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



$$x_1=10$$

$$x_2=1$$

$$x_3=4$$

5.4 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, imagínese 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 \int denota el proceso de integración.

Los valores que se obtienen de $P(X \leq a)$ para todos los valores posibles a , constituyen la **función de distribución acumulativa** de la variable aleatoria X , misma que se denota como 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 el terreno de la contaduría y la administración muchos de los problemas que habrán de enfrentarse hacen referencia a distribuciones de probabilidad muy conocidas, y por lo mismo distribuciones sobre las que se ha trabajado mucho, a grado tal que los valores de la probabilidad están ya tabulados. Uno de estas distribuciones es la distribución normal.



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 al alumno de ser un tanto complicada, pero no debe preocuparse 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.

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

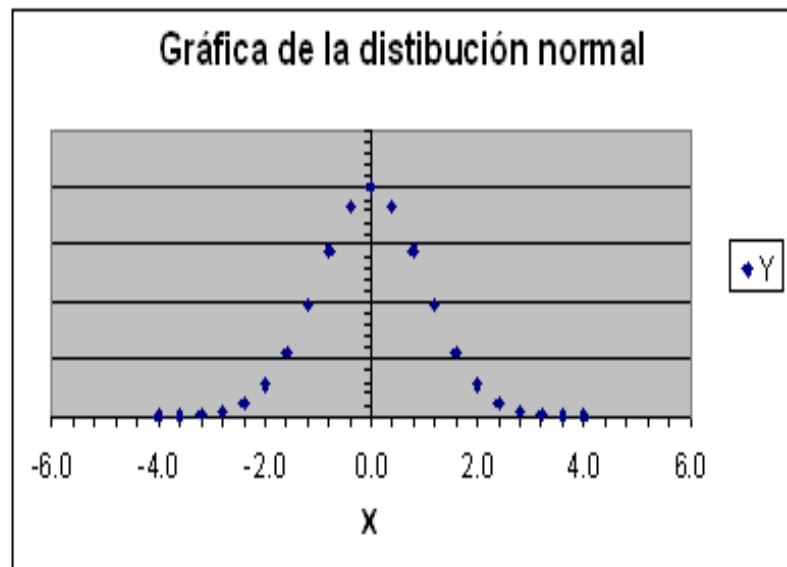
En esta función todos los términos son conocidos por el estudiante. La “y” es la ordenada de las coordenadas rectangulares cartesianas y representa la altura sobre el eje “x”; x es la abscisa en este sistema de coordenadas; $\pi = 3.14159$; “e” corresponde a la base de los logaritmos naturales que el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

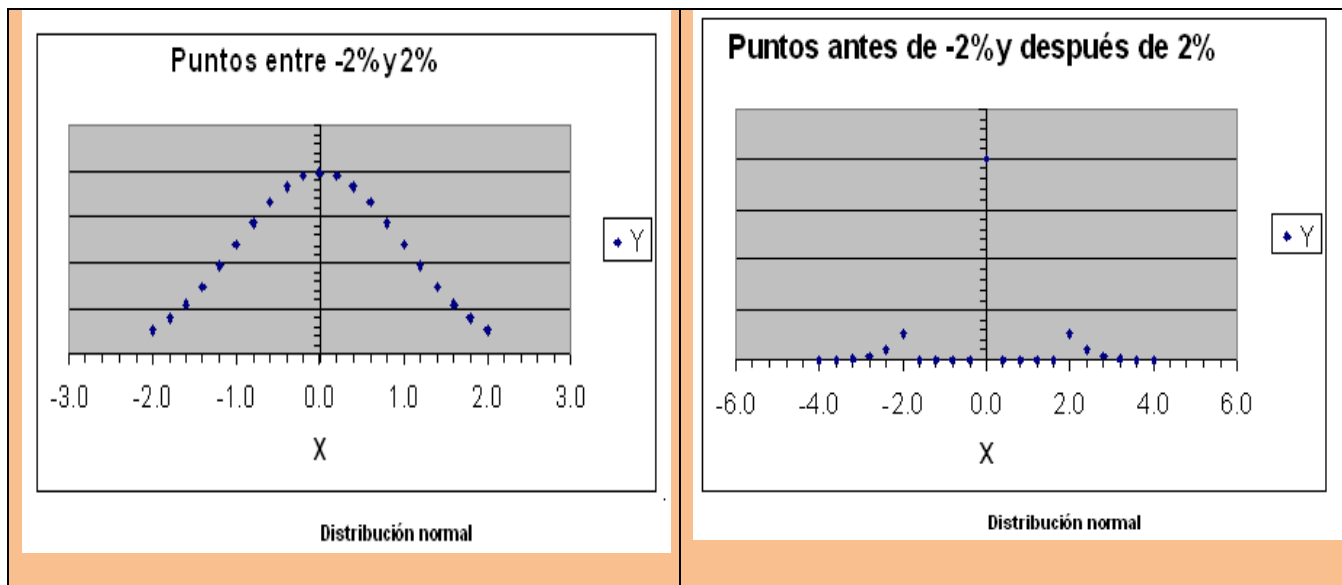
Es importante mencionar que el área que se encuentra entre la curva y el eje de las abscisas es igual a la unidad ó 100%. La curva normal es simétrica en relación con la media. Esto quiere decir que la parte de la



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 2000 clientes. Cada cliente debe en promedio \$7000 con una desviación estándar de \$1000. La distribución de los adeudos de los clientes es aproximadamente normal. Diga usted



cuantos clientes esperamos que tengan un adeudo entre \$7000 y \$8,500.

Solución:

Nos 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 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. Indique cuál será la probabilidad de que el suministro sea suficiente en una semana cualquiera si se surten:

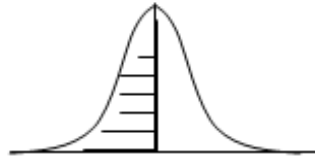
- a) 160 metros cúbicos.
- b) 180 metros cúbicos.
- c) 200 metros cúbicos.
- d) Indique asimismo cual será la probabilidad de que se acabe el agua si una semana cualquiera surten 190 metros cúbicos.

Solución:

- a) 160 metros cúbicos.

El valor de Z en este caso sería: $z = \frac{160 - 160}{20} = 0.0$ Esto nos

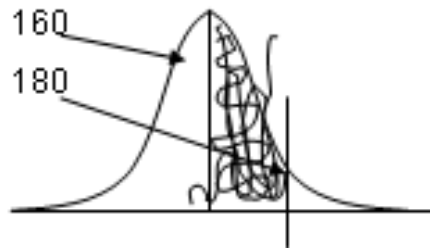
puede desconcertar un poco; sin embargo, nos podemos dar cuenta de que si se surten 160 metros cúbicos el agua alcanzará si el consumo es menor que esa cifra. La media está en 160. Por ello el agua alcanzará en toda el área de la curva que se muestra rayada. Es decir, toda la mitad izquierda de la curva. El área de cada una de las mitades de la curva es de 0.5, por tanto la probabilidad buscada es también de 0.5.



b) 180 metros cúbicos

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

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 dela curva (que por el diseño de la tabla no aparece). Ese valor, como ya se comentó es de .5. Por tanto, el valor buscado es de .5 más 0.34134. Por ello la probabilidad de que el agua alcance si se surten 180 metros cúbicos es de 0.84134, es decir, aproximadamente el 84%.



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

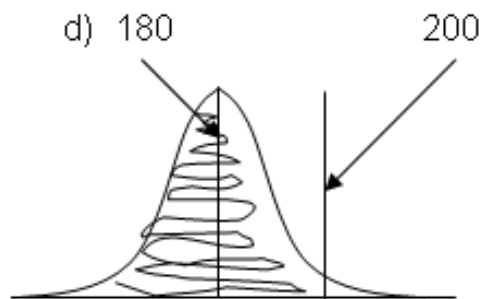
c) 200 metros cúbicos



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.

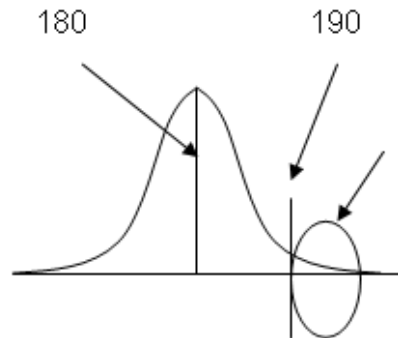
$$Z = \frac{200 - 160}{20} = 2$$

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

La probabilidad de que se termine el agua en estas condiciones se encuentra representada en la siguiente figura.



La probabilidad de que falte el agua está representada por el área en la cola de la distribución, después del 190. La tabla no nos da directamente ese valor. Para obtenerlo debemos de calcular Z para 190 y el valor del área entre la media y 190 restársela a .5 que es el área total de la parte derecha de la curva.

$$Z = (190 - 160) / 20 = 1.5.$$

El área para $Z = 1.5$ es de 0.43319. Por tanto, la probabilidad buscada es $0.5000 - 0.43319 = 0.06681$ o aproximadamente el 6.7%.

Búsqueda de Z cuando el área bajo la curva es conocida

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

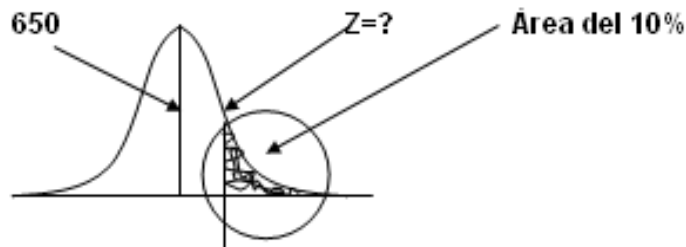


Ejemplo 4: Una universidad realiza un examen de admisión a 10,000 aspirantes para asignar los lugares disponibles. La calificación media de los estudiantes es de 650 puntos sobre 1000 y la desviación estándar es de 100 puntos; las calificaciones siguen una distribución normal. Indique usted qué calificación mínima deberá de tener un aspirante para ser admitido si:

- a) Se aceptará al 10% de los aspirantes con mejor calificación.
- b) Se aceptará al 5% de aspirantes con mejor calificación.

Solución:

- a) Si hacemos un pequeño esquema de la curva normal, los aspirantes aceptados representan el 10% del área que se acumula en la cola derecha de la distribución. El siguiente esquema nos dará una mejor idea.



El razonamiento que se hace es el siguiente:

Si el área que se busca es el 10% de la cola 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). Este es el valor de 0.39973 y se encuentra en el renglón donde aparece un valor para Z de 1.2 y en la columna de 0.08. Eso quiere decir que el valor de Z que más se aproxima es el de 1.28. No importa si al valor de la tabla le falta un poco o se pasa un poco; la idea es que sea el más cercano posible.

| z | 0 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 1.2 | 0.3849 | 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 a continuación se indica.

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.

- c) 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% ó 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.

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

$$15 - 60$$

$$\lambda - 5$$

$$\therefore \lambda = 1.25 \quad P T > t = e^{-\lambda}$$

$$p t > 5 = 2.71828^{-1.25} = 0.286 = 28.6\%$$

$$e = 2.71828$$

b) Primera llamada en 5 minutos:

$$P T \leq t = 1 - e^{-\lambda}$$

$$P T \leq 5 = 1 - 2.71828^{-1.25} = 1 - 0.287 = 0.713 = 71.3\%$$

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 |



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:

$$\begin{aligned}
 P(110 < S < 130) &= P(S= 111) + P(S= 112) + P(S= 113) + \dots + \\
 P(S= 129) &= \sum_{i=111}^{129} {}_{200}C_i (0.5)^i (0.5)^{200-i}
 \end{aligned}$$

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

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

En consecuencia,

$$\begin{aligned}
 P(111 \leq X \leq 129) &= P [(110.5 - 100) / 7.07 \leq Z \leq (129.5 - 100) / \\
 7.07] &= P (10.5 / 7.07 \leq Z \leq 29.5 / 7.07) \\
 &= P(1.49 \leq Z \leq 4.17) \\
 &= 0.5 - 0.4319 \\
 &= 0.0681
 \end{aligned}$$

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 | 0.0000 | 0.0100 | 0.0200 | 0.0300 | 0.0400 | 0.0500 | 0.0600 | 0.0700 | 0.0800 | 0.0900 |
|-----|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 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 DE LA UNIDAD

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.



GLOSARIO

Distribución binomial

Es la distribución de probabilidades en un experimento binomial, de modo que la variable aleatoria, discreta, se refiere al número de éxitos en n ensayos.

Distribución hipergeométrica

Es la distribución de probabilidades en un experimento en el que se extrae una muestra de n observaciones de una población que a su vez se puede dividir en dos grupos, K y L , en el primero de los cuales hay k observaciones y en el otro $N-k$, y en donde se desea determinar la probabilidad de que en la muestra haya x observaciones que pertenecen al grupo K . En este modelo la extracción de las observaciones no sigue un esquema con reemplazo por lo que la probabilidad de éxito no es constante.

Distribución multinomial

Es la distribución de probabilidades en un experimento en el que los resultados se pueden agrupar en K grupos mutuamente excluyentes, de modo que las variables aleatorias discretas correspondientes se refieren al número de casos que de la muestra caen en cada grupo.



Distribución normal

Describe el comportamiento estadístico de muchas situaciones del mundo real en donde:

- Hay simetría alrededor de un punto medio.
- Se miden poblaciones grandes.
- El 95% de los datos se concentra en un intervalo de longitud 1.96 veces la desviación estándar alrededor de la media.

Distribución de Poisson

Es la distribución de probabilidades en un experimento Poisson donde la variable aleatoria, discreta, se refiere al número de veces que ocurre un cierto resultado A.

Esperanza matemática

Es un atributo de las variables aleatorias y por lo tanto de la respectiva distribución de probabilidades que describe lo que ocurre en el centro de la distribución, aunque no siempre existe. En analogía con los indicadores de estadística descriptiva se puede decir que corresponde a la media.

Experimento binomial

Es un experimento que se caracteriza porque:

- Hay un número finito de ensayos.
- En cada ensayo sólo hay dos posibles resultados, a saber, éxito o fracaso.
- Los ensayos son independientes, de modo que la probabilidad de éxito permanece constante en cada ensayo.



Experimento Poisson

Es un experimento que se caracteriza porque:

- Un cierto resultado A puede ocurrir un número infinito pero numerable de veces a lo largo de un intervalo (de tiempo, longitud o área).
- La probabilidad de que ocurra el resultado A es proporcional al tamaño del intervalo.
- La probabilidad de que ocurra el resultado A en una fracción del intervalo es independiente de la probabilidad de que ocurra en otra fracción del intervalo.
- La probabilidad de que ocurra más de una vez el resultado A en una fracción muy pequeña del intervalo es prácticamente nula.
- Se conoce el valor de la tasa promedio de ocurrencia del resultado A.

Función de distribución

Es aquélla que acumula los valores de la función de probabilidad. Se denota como $F(x)$ y corresponde al valor de $P(X \leq x)$.

Función de probabilidad discreta

Es aquélla que asocia una probabilidad a cada posible valor de una variable aleatoria discreta.

Variable aleatoria

Es una aplicación que asocia un número real a cada elemento del espacio muestral de un experimento.



Variable aleatoria continua

Es una variable aleatoria que hace referencia a un experimento cuyos universo de resultados se encuentra en un intervalo del tipo (a,b) y la variable puede tomar cualesquier valor en ese intervalo.

Variable aleatoria discreta

Es una variable aleatoria que hace referencia a los distintos resultados de un experimento tomando un número finito o infinito numerable de posibles valores. Intuitivamente, una variable es discreta si deja huecos entre cada par de posibles valores.



ACTIVIDADES DE APRENDIZAJE

ACTIVIDAD 1

Considera 5 situaciones de tu vida cotidiana que en tu opinión dan lugar a experimentos aleatorios, esto es, situaciones en las que no puedes anticipar con toda certeza el resultado.

Establece para cada experimento el conjunto de posibles resultados. Designa la variable aleatoria correspondiente y clasifícala de acuerdo a los criterios que se han revisado en esta unidad.

ACTIVIDAD 2

Considera la siguiente situación.

Tres matrimonios, a los que conoceremos como A-B, M-N y P-Q, se han reunido para jugar canasta por una bolsa de \$30,000.00. El torneo es de parejas y con este propósito acuerdan que sean A y M quienes en ese orden seleccionen al azar compañero de juego.

La forma en que se determinan las parejas es la siguiente:

- Cada quien, excepto A, escribe su nombre en una papeleta e introduce



ésta en una urna.

- La persona cuya papeleta sea seleccionada por A, será la pareja de ésta.
- Si la papeleta seleccionada por A es la de M, entonces M hará pareja con A y ya no extraerá papeleta alguna. En este caso, la segunda papeleta será extraída por P.
- M, o P, según corresponda, extrae la segunda papeleta y de ser la propia, elimina ésta y procede a una nueva extracción.
- La tercera pareja queda automáticamente seleccionada.

Caracteriza la variable aleatoria que denota el número de parejas de juego formadas por matrimonios. Tal caracterización debe incluir el nombre de la variable, su tipo (discreto o continuo), su recorrido y su distribución de probabilidades.

ACTIVIDAD 3

Considera la siguiente situación.

En el puerto de Balankub hay una sociedad cooperativa de taxis que proporciona servicio desde varias bases a cualquier destino. En la cooperativa desean determinar el número de unidades que deben tener en promedio en la base del aeropuerto. Saben que todos los martes llegan, en el mismo vuelo, cuatro gerentes ejecutivos de cuatro diferentes empresas. Cada uno puede escoger, de manera independiente, ir a la terminal de autobuses si su destino final es A1, ir a la terminal del tren si su destino final es A2 o ir a la terminal del transbordador si su destino final es A3. Si en cierto mes hay cinco días martes, ¿cuál es la probabilidad de que por lo menos en tres de ellos los cuatro ejecutivos escojan el mismo destino?



ACTIVIDAD 4

Resuelve los siguientes problemas.

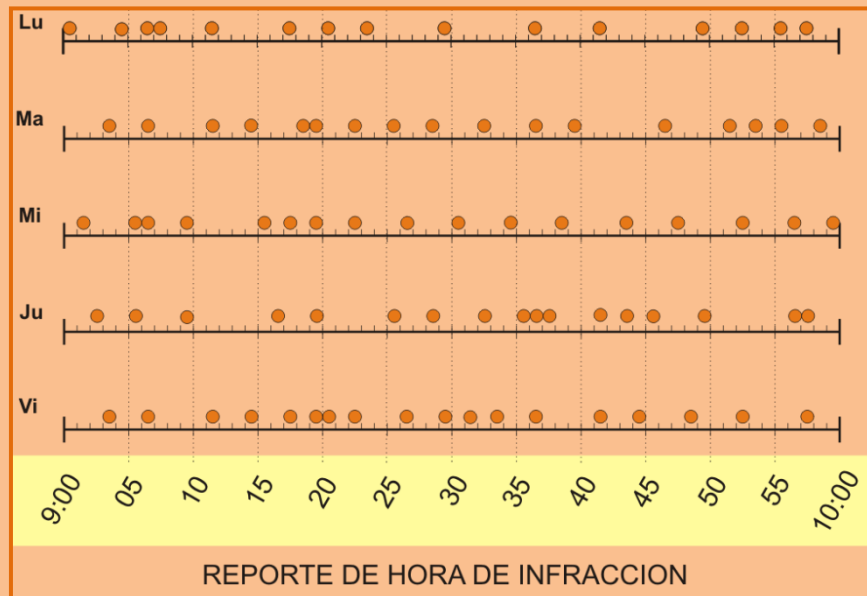
1. En un corporativo con 500 empleados se llevó a cabo una auditoria preliminar de documentos en el área de recursos humanos. Se detectó que en 8 de cada 30 expedientes falta el documento A, que en 6 de cada 24 expedientes falta el documento B y que en uno de cada 50 falta el documento C. Se considera como omisión grave que falte cualquiera de los tres documentos. Se desea saber cuál es la probabilidad de que en 400 expedientes no se detecte omisión alguna.

¿Qué modelo de distribución probabilística aplicarías?. Establece los parámetros del mismo y expresa la relación algebraica que permitiría calcular la probabilidad señalada. Si consideras que hay un modelo alternativo que daría un valor aproximado de la probabilidad, calcula los valores solicitados y compara los valores obtenidos.

2. La carretera que comunica las poblaciones de San Albano y San Miguel tiene un tramo recto de 4.2 kms en el que con frecuencia se registran accidentes por exceso de velocidad, por lo que las autoridades han decidido colocar una cámara-radar de velocidad que envía la información a la computadora de la oficina de tránsito, en la que se registra la hora, la velocidad y número de placa del vehículo. En aquéllos casos en que la velocidad excede el límite establecido se emite la multa correspondiente. Además, se genera un reporte en forma de cinta para mostrar la hora de la infracción



En la figura se muestran tales reportes para los últimos cinco días hábiles entre las 9:00 y las 10:00 am.



- Se desea determinar la probabilidad de que en un lapso de cinco minutos
 - k vehículos excedan la velocidad, con $k=0, 1, 2, 3, 4$ y 5
 - Como máximo tres vehículos excedan la velocidad
 - Como mínimo tres vehículos excedan la velocidad
- ¿Cuál es el valor esperado de vehículos que exceden la velocidad en un lapso de cinco minutos?
- ¿Cuál es el valor esperado de vehículos que exceden la velocidad en un lapso de una hora?



ACTIVIDAD 5

Se ha diseñado un sistema de alerta que emite una señal auditiva para avisar a dependientes invidentes cuando alguien ingresa a un establecimiento comercial. La duración de la señal es una variable aleatoria y se cree que la función de densidad respectiva es $f_X(x) = c$, en el intervalo $(0.2, 0.8)$. Se desea saber cuál debe ser el valor de c para que la función sea efectivamente una función de densidad.

Sugerencia:

Recuerda que para que una función sea de densidad, el área bajo la curva en todo el recorrido de la variable aleatoria debe ser igual a uno. Asigna un valor arbitrario a c (por ejemplo, $c = 1$) y elabora la grafica respectiva de $f(x) = c$. Observa la forma de la figura que se genera y establece de qué manera puedes calcularla. Luego, determina cuál debería ser el valor de c para que el área sea uno.

ACTIVIDAD 6

Contesta las siguientes preguntas.

La gerencia de recursos humanos de un corporativo aplica a un grupo de solicitantes de empleo una prueba de aptitud. La calificación promedio obtenida por los solicitantes es de 78 puntos con una desviación estándar de 13.

- a) ¿Cuál es la probabilidad de que si se selecciona al azar a uno de tales solicitantes, éste tenga una calificación



- superior a 85 puntos?
- menor a 75 puntos?
- entre 70 y 90 puntos?

b) ¿Entre qué valores se encuentra el 80% de la población que excluye al 10% más apto y al 10% menos apto?

c) ¿Cuál es la calificación máxima del 25% menos apto?

ACTIVIDAD 7

La gerencia de un banco está interesada en determinar la probabilidad de errores en las operaciones de depósito. Si se auditan 5 000 de estas operaciones, ¿cuál es la probabilidad de encontrar entre 10 y 15 operaciones con error?

- a. Si se sabe que la probabilidad de cometer un error es de 0.005.
- b. Si se sabe que la probabilidad de cometer un error es de 0.3.

Justifica el uso el uso de las distribuciones normal o de Poisson como aproximación a la distribución real.



ACTIVIDAD 8

Considera la siguiente situación:

En una ciudad de tamaño medio se han establecido 3 academias particulares, sin que desde el punto de vista de la investigación de mercados haya mayores diferencias entre ellas. Se calcula que para el siguiente año hay 800 alumnos que podrían matricularse de los cuales finalmente lo harán 450. Las autoridades de una de las academias desean saber cuál debe ser la capacidad de su escuela para que sea suficiente en el 80% de los cursos

Si tú fueses consultor, qué respuesta darías. Justifica ésta ampliamente.



CUESTIONARIO DE REFORZAMIENTO

1. Indique la diferencia entre las variables discretas y las variables continuas.
2. ¿A qué se refiere el nivel conceptual y el nivel operacional?
3. Exprese cuáles son las propiedades de una distribución binomial.
4. ¿En qué consiste una distribución de Poisson?
5. ¿En qué casos se utiliza una aproximación de la distribución de Poisson a la binomial?
6. ¿Qué es una distribución de probabilidad de variable continua?
7. Explique las características fundamentales y uso de la distribución normal.
8. Exprese la fórmula de la variable “z” parametrizada de una distribución normal.
9. Explique las características fundamentales y uso de la distribución exponencial.
10. Exprese la fórmula para obtener la probabilidad de éxito de un evento en una distribución exponencial.



LO QUE APRENDÍ

En esta unidad has entrado en contacto con conceptos tales como variable aleatoria, distribución de probabilidades y esperanza matemática. Con ellos puedes desarrollar una explicación e interpretación formal del concepto de esperanza de vida al nacimiento, entre otras muchas aplicaciones de las ideas que aquí se han presentado.

En particular puedes explicar qué aspectos inciden en el incremento en el valor de la esperanza de vida y discute con tus compañeros ¿Por qué crees que esto ocurre?. Discútelo con tu asesor.



EXAMEN DE AUTOEVALUACIÓN

I. Para cada una de las siguientes situaciones, determina si la variable aleatoria correspondiente es discreta o continua. Marca con una X en la columna que corresponda.

| Situación | Discreta | Continua |
|--|-----------------------|-----------------------|
| El número de movimientos requeridos para ganar un juego de Mahjong | <input type="radio"/> | <input type="radio"/> |
| El tiempo de vida útil de un foco | <input type="radio"/> | <input type="radio"/> |
| La tasa de cambio peso-dólar | <input type="radio"/> | <input type="radio"/> |
| El número de personas que usan un determinado torniquete en una estación del metro de la Cd. De México | <input type="radio"/> | <input type="radio"/> |
| La marca que obtiene un atleta en salto de longitud en un intento cualquiera | <input type="radio"/> | <input type="radio"/> |



II. Considera las situaciones que se presentan a continuación y contesta lo que se te pide. Anota tu respuesta en el cuadro de la derecha

Se arroja una moneda 6 veces. ¿Cuál es el número esperado de veces que saldrá cara?

Se arroja una moneda 4 veces. ¿Cuál es la varianza del número de veces que saldrá cruz?

Se arroja un dado 10 veces. ¿Cuál es el número esperado de veces que saldrá 3 o 4, si en el primer lanzamiento salió 2?

III. Indica cuáles de las siguientes aseveraciones son verdaderas y cuáles son falsas. Marca tus respuestas en las columnas del lado derecho según corresponda.

| | Falso | Verdadero |
|---|-----------------------|-----------------------|
| La media de la distribución binomial es $np(1-p)$ | <input type="radio"/> | <input type="radio"/> |
| En una distribución multinomial es necesario especificar k valores de probabilidad para poder aplicar el modelo | <input type="radio"/> | <input type="radio"/> |
| Se puede emplear el modelo binomial para aproximar el modelo Poisson si el valor de p es grande. | <input type="radio"/> | <input type="radio"/> |



| | | |
|---|-----------------------|-----------------------|
| En una distribución multinomial los K grupos deben ser mutuamente excluyentes | <input type="radio"/> | <input type="radio"/> |
| El modelo hipergeométrico es una generalización del modelo binomial | <input type="radio"/> | <input type="radio"/> |
| En el modelo Poisson la varianza está dada por $np(1-p)$ | <input type="radio"/> | <input type="radio"/> |
| Se puede emplear el modelo Poisson para aproximar el modelo binomial cuando los valores de n son pequeños | <input type="radio"/> | <input type="radio"/> |

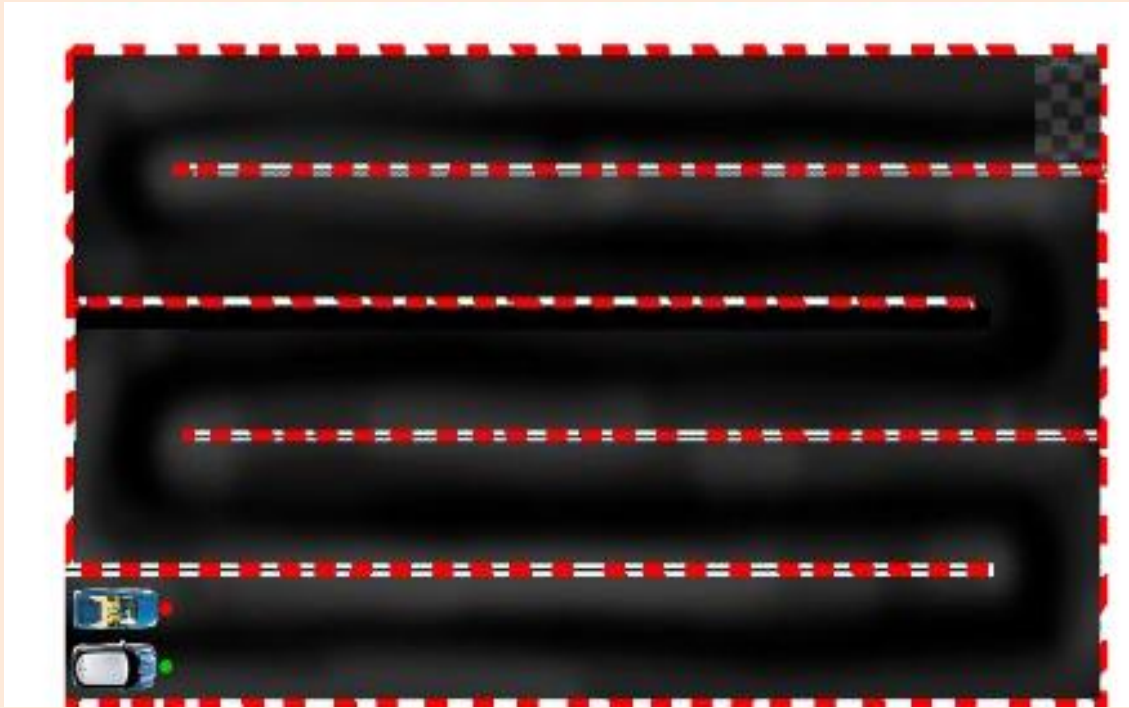
IV. Determina el valor de las probabilidades que se solicitan. Anota tu respuesta en el cuadro de la derecha.

Si X es una variable aleatoria con distribución normal con media 15 y desviación estándar 3, ¿cuál es el valor de:

| | |
|--------------------------|--|
| $P(X \geq 14)$? | |
| $P(X \leq 17)$? | |
| $P(13 \leq X \leq 14)$? | |
| $P(14 \leq X \leq 17)$? | |
| $P(X \geq 16)$? | |



V. Una vez que hemos estudiado esta unidad te invitamos a que participes en este Rally, para que compruebes el logro del objetivo que nos planteamos. Responde las siguientes preguntas



a

b

c

d

e

1. Las empresas mexicanas están aprovechando la condición de costo de mano de obra más barato para realizar trabajos en el extranjero; para ello utilizan los servicios de empresas de contratación locales para la resolución de todos los aspectos legales. Las encuestas realizadas por el Banco de Comercio Exterior indican que el 20% de las empresas mexicanas utilizan a este tipo de empresas. Si el banco selecciona al azar a un grupo de 15 empresas mexicanas. ¿Calcule la probabilidad de que exactamente cinco de ellas estén empleando a estas empresas locales?

a. 0.1032

b. 0.1058

d. 0.1035

e. 0.1038



c. 0.1028

a

b

c

d

e

2. De acuerdo con la situación anterior, calcule la probabilidad de que el número de empresas mexicanas que contratan empresas locales en el extranjero se ubique entre seis y nueve.

- a. 0.0651
- b. 0.0609
- c. 0.0631
- d. 0.0607
- e. 0.0609

a

b

c

d

e

3. El Banco Nacional de México sabe por su experiencia que durante los días lunes, entre las 9:00 y las 10:00, se presentan a la ventanilla de atención a clientes un promedio de 2.8 clientes cada 4 minutos, número que la cajera puede atender con eficiencia. Con el propósito de verificar si el número de cajeras es el adecuado, calcule la probabilidad de que se presente un total de cuatro clientes en un intervalo de cuatro minutos.

- a. 0.1568
- b. 0.1557
- c. 0.1535
- d. 0.1678
- e. 0.1456



a

b

c

d

e

4. Un auxiliar de contador puede cometer 1.2 errores por cada 200 declaraciones fiscales. ¿Calcule la probabilidad de que al seleccionar una de las declaraciones elaboradas por él no se encuentre algún error?

a. 0.9600

b. 0.9940

c. 0.3012

d. 0.5990

e. 0.7890

a

b

c

d

e

5. Un banco recibe en promedio a 3.2 clientes cada 4 minutos. ¿Cuál es la probabilidad de obtener exactamente 10 clientes en los próximos 8 minutos?

a. 0.0538

b. 0.0635

c. 0.0535

d. 0.0525

e. 0.0528



a

b

c

d

e

6. Una persona presentará el examen de conocimientos y dominio de la lengua inglesa, denominado GMAT; sus resultados tienen un valor medio de 494 puntos con desviación estándar de 100; La persona desea conocer la probabilidad de obtener 700 puntos. Considere que los resultados siguen una distribución normal estándar.

- a. 0.0255
- b. 0.0204
- c. 0.0197
- d. 0.0199
- e. 0.0193

a

b

c

d

e

7. Una empresa de seguros está considerando incluir entre los riesgos cubiertos, una enfermedad denominada Túnel Carpiano, la cual aparece en manos y muñecas, provocada por los esfuerzos realizados con estas partes del cuerpo durante tiempos prolongados. Se estima que el costo de tratamiento de estas afecciones es alrededor de \$30,000 pesos al año por trabajador lesionado, con una desviación estándar de \$9,000.00. La aseguradora supone que la afección está normalmente distribuida y desea estimar los costos en que puede incurrir. Calcule la probabilidad de que el costo de atención se encuentre entre \$15,000 y \$ 45,000.

- a. 0.9050
- b. 0.0950
- c. 0.9152
- d. 0.9070
- e. 0.9030



a

b

c

d

e

8. Una empresa de automóviles menciona en su publicidad que sustituirá por una unidad nueva los autos que presenten cualquier tipo de falla en el tren motriz durante los primeros 80,000 kilómetros. Si la empresa sabe que el valor medio del kilometraje sin fallas es de 80,000 kilómetros y la desviación estándar de 10 000 kilómetros, ¿cuál debería ser el kilometraje garantizado para no tener que reponer más del 10% de los autos?

- a. 65 600
- b. 68 300
- c. 67 200
- d. 68 450
- e. 63 600

a

b

c

d

e

9. Una distribución normal tiene una media de 4.9 y una desviación estándar de 1.2. ¿Qué porcentaje del área bajo la curva es mayor que 6?

- a. 0.1685
- b. 0.1814
- c. 0.1797
- d. 0.1788
- e. **0.1750**



a

b

c

d

e

10. Se aplica un examen de Matemáticas a 4000 estudiantes próximos a egresar del ciclo de educación media superior. Si en experiencias previas ha ocurrido que la calificación promedio es de 6.7 con una desviación estándar de 3.1, y bajo el supuesto de que las calificaciones se distribuyen de manera normal, ¿cuál es el número de estudiantes que podría esperarse en esta ocasión obtuviesen una calificación superior a 9.0?

- a. 229
- b. 918
- c. 230
- d. 770
- e. 3082



VI. El tiempo que le toma a un metrobus efectuar un recorrido completo entre las dos estaciones terminales tiene una distribución normal. Se sabe que 75% de las veces le lleva más de 57 minutos completar el trayecto mientras que 15% de las veces le lleva más de 65 minutos. Se desea saber cuáles son los parámetros de la distribución normal.

Para dar respuesta a esta interrogante te invitamos a descubrir la trayectoria correcta de pasos a seguir. Para ello, escoge de cada punto una opción.

1) Con la información que se proporciona podemos establecer dos afirmaciones probabilísticas:

Opción a

$$P(X \geq 57) = 0.75$$

$$P(X \geq 65) = 0.15$$

Opción b

$$P(X \geq 57) = 0.75$$

$$P(X \geq 65) = 0.15$$

Opción c

$$P(X \geq 57) = 0.75$$

$$P(X \geq 65) = 0.15$$

2) Sean μ y σ los parámetros media y desviación estándar respectivamente. Como μ no puede ser cero, sabemos que es necesario estandarizar. Las afirmaciones probabilísticas toman la siguiente forma:

Opción a

$$P\left(\frac{X - \mu}{\sigma} \leq \frac{57 - \mu}{\sigma}\right) = 0.75$$

$$P\left(\frac{X - \mu}{\sigma} \leq \frac{65 - \mu}{\sigma}\right) = 0.15$$

Opción b

$$P\left(\frac{X - \mu}{\sigma} \geq \frac{57 - \mu}{\sigma}\right) = 0.25$$

$$P\left(\frac{X - \mu}{\sigma} \geq \frac{65 - \mu}{\sigma}\right) = 0.85$$

Opción c

$$P\left(\frac{X - \mu}{\sigma} \geq \frac{57 - \mu}{\sigma}\right) = 0.75$$

$$P\left(\frac{X - \mu}{\sigma} \geq \frac{65 - \mu}{\sigma}\right) = 0.15$$



3) Se determina en tablas el valor de las abscisas para las cuales se cumplen los valores de probabilidad indicados.

Opción a

$$\frac{57 - \mu}{\sigma} = -0.675$$

$$\frac{65 - \mu}{\sigma} = 0.385$$

Opción b

$$\frac{57 - \mu}{\sigma} = -0.675$$

$$\frac{65 - \mu}{\sigma} = 0.385$$

Opción c

$$\frac{57 - \mu}{\sigma} = -0.675$$

$$\frac{65 - \mu}{\sigma} = 0.385$$

4) Se genera un par de ecuaciones simultáneas con dos incógnitas:

Opción a

$$57 = -\mu - 0.675\sigma$$

$$65 = \mu + 0.385\sigma$$

Opción b

$$57 = \mu - 0.675\sigma$$

$$65 = \mu + 0.385\sigma$$

Opción c

$$57 = \mu + 0.675\sigma$$

$$65 = -\mu - 0.385\sigma$$

5) Al resolver el sistema se obtienen los valores de μ y σ .

- a) ¿Cuál es la trayectoria correcta de pasos que se deben seguir?
- b) ¿Cuáles son los valores de los parámetros?



MESOGRAFÍA

BIBLIOGRAFÍA RECOMENDADA

| Autor | Capítulo | Páginas |
|--------------------------------|--|-----------|
| 1. Anderson, Sweeney, Williams | 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 | 218-229 |



| | | |
|-----------------------------------|--|------------|
| | Sección 6.2 Distribución de probabilidad normal | |
| | Sección 6.3 Distribución de probabilidad exponencial | 232-234 |
| 2. Berenson, Levine y Krehbiel. | 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, | 6. Formación de hipótesis Sección: Definición conceptual o constitutiva. | 145-146 |
| 4. Levin y Rubin. | 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 |
| 5. Lind, Marchal, Wathen. | 6. Distribuciones discretas de de probabilidad. Secciones: ¿Qué es una distribución de probabilidad?. | 181 - 183 |



| | |
|--|-----------|
| | |
| 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 |



BIBLIOGRAFÍA BÁSICA

1. Anderson, David R., Sweeney, Dennis J., Williams, Thomas A. (2005), *Estadística para administración y economía*, 8a edición, México, International Thomson Editores, 888 páginas más apéndices.
2. Berenson, Mark L., David M. Levine, y Timothy C. Krehbiel, (2001), *Estadística para administración*, 2ª edición, México, Prentice Hall, 734 páginas.
3. Hernández Sampieri, R., C. Fernández Collado, Lucio P Baptista, (2006), *Metodología de la investigación*, 4ª edición, México: McGraw Hill Interamericana, 850 páginas.
4. Levin, Richard I. y David S. Rubin, (2004), *Estadística para administración y economía*, 7a. Edición, México, Pearson Educación Prentice Hall, 826 páginas más anexos.
5. 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.



BIBLIOGRAFÍA COMPLEMENTARIA

1. Bowerman Bruce, *Pronósticos, series de tiempo y regresión; un enfoque aplicado*, México: Cengage Learning, 4ª edición, 2007, 720 pp.
2. Mendenhall William, *Introducción a la probabilidad y estadística*, México: Cengage Learning, 13ª edición, 2010, 776 pp.
3. Webster Allen L., *Estadística I aplicada a los negocios y la economía*, México: McGraw-Hill, 2ª. edición, 2002, 154 pp.



SITIOS DE INTERNET

- García Cebrián María José, *La distribución normal*, material que forma parte del proyecto Descartes 2D del ministerio de educación, política social y deporte del gobierno de España, en el sitio:

http://descartes.cnice.mec.es/materiales_didacticos/distribuciones_probabilidad/dis_normal.htm

- García Cebrián, José María, *Otras aplicaciones de la distribución normal*, publicado como parte del proyecto Descartes 2D del ministerio de educación, política social y deporte del gobierno de España, en el sitio:

http://descartes.cnice.mec.es/materiales_didacticos/distribuciones_probabilidad/aplic_normal.htm

- Larios O. Víctor, (profesor de la Universidad Autónoma de Querétaro), *Unidad 5. Distribuciones de probabilidad*, del hipertexto Estadística, en el sitio:

<http://www.uaq.mx/matematicas/estadisticas/xu4-5.html#t1>



- Luna Gándara MC Rita (profesora de la licenciatura en ingeniería industrial del Instituto Tecnológico de Chihuahua), *Unidad V. Distribuciones de probabilidad continuas*, apuntes del curso de probabilidad y estadística, en el sitio:

http://www.itch.edu.mx/academic/industrial/sabaticorita/_private/01UNIDAD%20%20V.htm.

- Hospital Universitario Ramón y Cajal, *Variable aleatoria*, hasta la sección Inducción a la probabilidad, en el sitio:

http://www.hrc.es/bioest/estadis_21.html

- Rodríguez Mayté (profesora del curso de Estadística aplicada a las ciencias sociales II de la licenciatura de sociología de la Universidad Autónoma de Madrid), *Variables aleatorias*, en el capítulo 3, secciones 3.1 y 3.2, p. 25-27, en el sitio:

http://www.uam.es/personal_pdi/ciencias/mayter/docencia/sociologia/apuntes.pdf

- Rodríguez Mayté (profesora del curso de Estadística aplicada a las ciencias sociales II de la licenciatura en sociología de la Universidad Autónoma de Madrid) *Variables aleatorias discretas*, en el capítulo 4, secciones 4.1, 4.2 y 4.4, págs. 34 a 41, en el sitio:

http://www.uam.es/personal_pdi/ciencias/mayter/docencia/sociologia/apuntes.pdf



SUAYED
UNA OPCIÓN
PARA TI

UNIDAD 6

NÚMEROS ÍNDICES

APUNTES DIGITALES PLAN 2011

SUAYED
UNA OPCIÓN
PARA TI



OBJETIVO ESPECÍFICO DE LA UNIDAD

El alumno conocerá los métodos para calcular e interpretar los números índice.



INTRODUCCIÓN

Los gobiernos y otras entidades publican diversas clases de índices. Estos están elaborados con el propósito de presentar de manera sencilla el **comportamiento** de alguna o algunas **variables de interés**. El alumno seguramente habrá escuchado mencionar el Índice Nacional de Precios al Consumidor, el Índice de Precios y Cotizaciones (IPC) de la Bolsa Mexicana de Valores o el Dow Jones que se relaciona con el mercado de valores de Nueva York. Todos estos son precisamente números índice.

Estos indicadores son muy útiles para los profesionales de la contaduría y la administración, pues son elementos de juicio para la toma de decisiones. Es importante mencionar que un solo número índice nos arroja muy poca información. Por ejemplo, si alguien nos dice que el IPC de la Bolsa Mexicana de Valores cerró hoy a 4900 puntos, no revela una información que pueda ser usada para tomar decisiones. Lo importante es saber cómo se ha comportado el índice a lo largo de los días; es decir, saber si el valor del índice ha aumentado o si por el contrario ha disminuido.

Así, la información de los índices nos es útil en cuanto podemos ver su comportamiento en el tiempo. Podemos decir que un **índice** conforma una **serie de tiempo**. Una serie de tiempos es un conjunto de datos recopilados y utilizados en orden cronológico. De esta manera, el estudio de un índice a través del tiempo nos proporciona una idea de la dinámica de los fenómenos que el propio índice contempla.



LO QUE SE

De acuerdo con cifras del Banco de México, la inflación (acumulada) en el periodo 2000-2006 fue como se muestra en la siguiente tabla:

| Año | Inflación |
|------|-----------|
| 2001 | 4.4 |
| 2002 | 5.7 |
| 2003 | 3.98 |
| 2004 | 5.19 |
| 2005 | 3.33 |
| 2006 | 4.05 |

¿Consideras que tasas de inflación como estas pueden afectar tu vida?

Para mayor información puedes consultar el documento *Devaluación-Inflación. México-USA 1970-2008* publicado por Manuel Aguirre Botello en MéxicoMáxico en <http://www.mexicomaxico.org/>



TEMARIO DETALLADO

- 6.1. Número índice simple.
- 6.2. Índices de precios agregados.
 - 6.2.1 Índice de Lapeyres.
 - 6.2.2 Índice de Paasche.
- 6.3. Principales índices de precios.
 - 6.3.1 Índice de precios al consumidor.
 - 6.3.2 Índice de precios al productor.
 - 6.3.3 Índice de precios y cotizaciones (IPC).
 - 6.3.4 Índice Dow Jones.
- 6.4. Deflación de una serie.



6.1 Número índice simple

En primer término, y como antecedente, vamos a mencionar algunos aspectos relacionados con la construcción de un índice.

Dado que lo importante de un índice es observar su comportamiento en el tiempo, la elección del periodo que va a servir de base es muy importante. Vamos a suponer que deseamos desarrollar un índice que refleje la ocupación de los hoteles y que definimos nuestro índice de ocupación como:

Porcentaje de ocupación de 1980=100.

Si decimos que, en el contexto de la ocupación hotelera, el año de 1980 fue bueno, entonces podemos suponer que la mayoría de los otros años se verá “mal” en comparación, pues tendrán valores menores a cien. En cambio, si el año de 1980 fue muy malo, la mayoría de los años se verá “bien” pues tendrá valores mayores a cien. Si deseamos construir un índice que no resulte engañoso, debemos elegir un periodo base que no sea ni exageradamente bueno ni exageradamente malo. Usualmente el índice del periodo base es igual a 100.



Otra consideración que debemos hacer al construir un número índice es que la razón básica de su utilización es la de resumir circunstancias, a veces muy complejas, en un solo número que sea fácil de comprender y de manejar. Por ello debemos de tomar la decisión de lo que queremos reflejar en él.

Si deseamos reflejar la fluctuación en la cantidad de bienes o servicios seleccionados (cantidad de automóviles, cantidad de consultas médicas, etc.) debemos construir un índice de cantidad.

Si lo que deseamos es reflejar los cambios en el valor total de un grupo de bienes o servicios, crearemos un índice de valor. Por ejemplo, el índice del valor total de automóviles vendidos, o de servicios médicos proporcionados en un año.

Cuando usamos un índice para reflejar un solo bien, estaremos construyendo un **índice simple**. En cambio, si conjuntamos varios bienes en el mismo índice, estaremos trabajando un **índice agregado**. A veces no existe un único índice que satisfaga nuestras necesidades, pero sí dos o tres (o más) de ellos que contemplan la información que necesitamos. En ese caso podemos conjuntar estos índices para formar un **índice compuesto**.

Existen diversos tipos de índices útiles. Explicaremos algunos de ellos.

Índice de cantidad

Este índice mide los cambios en las unidades de un bien de acuerdo con su origen, destino, utilización, etc. Todo ello a través del tiempo. Como lo que nos importa es medir la variación en la cantidad, mantenemos



constantes los precios de los bienes para luego calcular el valor de lo consumido o producido en los periodos considerados

La expresión que en seguida aparece nos muestra el manejo formal de este tipo de índices.

$$IQ = \left(\frac{\sum P_b Q_i}{\sum P_b Q_b} \right) 100$$

En donde Q_i es la cantidad del bien en el periodo que se desea obtener, Q_b es la cantidad del bien que se toma como base y P_b es el precio del año base.

Si sólo deseamos comparar cantidades de un solo bien, como en el caso de un índice simple, la expresión anterior se simplifica para quedar

$$IQ = \left(\frac{Q_i}{Q_b} \right) 100$$

A continuación se presenta un ejemplo que nos permitirá una mejor comprensión de los índices de cantidad.

Ejemplo 1. Una empresa que vende autobuses de pasajeros ha decidido establecer un índice de cantidades vendidas. Se acordó como periodo base el mes de junio de 2001. En la primera columna se muestran las unidades vendidas; en la segunda, el índice propiamente dicho.



| VENTAS | (UNIDADES) | ÍNDICE |
|---------------|------------|--------|
| Marzo 01 | 93 | 109.41 |
| Abril 01 | 81 | 95.29 |
| Mayo 01 | 78 | 91.76 |
| Junio 01 | 85 | 100.00 |
| Julio 01 | 90 | 105.88 |
| Agosto 01 | 94 | 110.59 |
| Septiembre 01 | 84 | 98.82 |
| Octubre 01 | 89 | 104.71 |
| Noviembre 01 | 92 | 108.24 |

Por ejemplo, para el mes de abril, el índice se calcula tomando el cociente del número de unidades correspondientes al mes de abril entre las de junio y multiplicando por cien

Índice de valor

Este índice mide en unidades monetarias (pesos, dólares, etc.) el valor (ya sea de costo o de precio de venta, según el caso) de un conjunto de bienes y/o servicios. Es importante subrayar que en este tipo de índices se toma en cuenta el cambio en el valor de cada bien que se incluye en el índice, por lo que para su cálculo, mantenemos constantes las cantidades de los bienes para luego calcular el valor de lo consumido o producido en los periodos considerados

La expresión que en seguida aparece nos muestra el manejo formal de este tipo de índices.

$$IV = \left(\frac{\sum P_i Q_b}{\sum P_b Q_b} \right) 100$$



En donde Q_b es la cantidad del bien en el periodo base, en tanto que P_i y P_b son los precios del periodo de referencia y del periodo base respectivamente.

Si sólo deseamos comparar precios o valores de un solo bien, como en el caso de un índice simple, la expresión anterior se simplifica para

$$IQ = \left(\frac{P_i}{P_b} \right) 100$$

quedar

6.2 Índices de precios agregados

Como ya se ha anticipado en el tema anterior, un índice agregado es aquel que agrupa o resume el comportamiento de varios bienes o servicios. Consideremos el siguiente ejemplo, a manera de repaso del concepto de índice agregado.

Ejemplo 1. Una librería adquirió 5 títulos en el mes de mayo. Los datos se muestran en la siguiente tabla



| Título | Qb | Pb | Pb Qb |
|-------------|-----------------------|-------------------------|---------------|
| | Cantidad del mes base | Costo unitario mes base | |
| 1 | 20 | 104 | 2,080 |
| 2 | 50 | 89 | 4,450 |
| 3 | 20 | 215 | 4,300 |
| 4 | 40 | 100 | 4,000 |
| 5 | 35 | 155 | 5,425 |
| SUMA | | | 20,255 |

Al mes siguiente adquirió los mismos títulos, pero a precios distintos, como se muestra a continuación:

| Título | Qb | Pi | Pi Qb |
|-------------|-----------------------|------------------------------|---------------|
| | Cantidad del mes base | Costo unitario mes corriente | |
| 1 | 20 | 107 | 2,140 |
| 2 | 50 | 95 | 4,750 |
| 3 | 20 | 215 | 4,300 |
| 4 | 40 | 115 | 4,600 |
| 5 | 35 | 172 | 6,020 |
| SUMA | | | 21,810 |

Con los datos anteriores ya podemos obtener el índice de valor.

$$IV = \left(\frac{\sum P_i Q_b}{\sum P_b Q_b} \right) 100$$



Recordemos que:

Por lo tanto, Índice de Valor = $21810 / 20255 \cdot 100 = 108.68$

Índice compuesto

Se puede presentar la posibilidad de que no exista un índice ya publicado por alguna agencia de gobierno o institución de investigación privada que satisfaga las necesidades particulares de una empresa. Pensemos, sin embargo, que existen dos o más índices que sí están publicados y que satisfacen, por partes, sus necesidades de información.

Vamos a suponer que una empresa tiene el 70% de sus negocios en la Ciudad de México y el 30% restante en la ciudad de Puebla. Para tomar decisiones acertadas, el gerente de esta empresa necesita un índice de precios que combine ponderadamente los índices agregados de precios de ambas ciudades.

Cuando hablamos de combinación ponderada nos referimos a darle un peso (la ponderación) a cada uno de los elementos que se van a combinar. Este peso representa de alguna manera, la importancia relativa que se asigna a cada elemento, de manera tal que la suma de las ponderaciones sea uno o 100%. A continuación se ilustra cómo aplicar esto.

Supongamos que:

- a) Índice agregado de precios al consumidor de la ciudad de México en abril del año en curso: 214.382



- b) Índice agregado de precios al consumidor de la ciudad de Puebla en abril del año en curso: 208.214

$$\text{Índice compuesto} = (0.70) 214.382 + (0.30) 208.214$$

← Índice de la Cd. de Puebla
Peso del índice De la Cd. de Puebla
Índice de la Cd. de México
Peso del Índice de la Cd. de México

$$\text{Índice compuesto} = 150.067 + 62.464 = 212.531$$

En este caso, las ponderaciones se han asignado en función del peso relativo que tienen los diferentes negocios de la empresa según la ciudad.

El valor que hemos obtenido corresponde al valor del índice compuesto y, como hemos visto, lo podemos generar mediante la siguiente expresión:

$$IC = \sum I_i P_i$$

En donde:

IC Es el índice compuesto

li Es cada uno de los índices agregados que queremos que tomen parte del índice compuesto.

Pi Es el peso o proporción que se da a ese índice en el propio índice compuesto.

No está demás volver a señalar que la suma de todos los pesos debe ser igual a la unidad, es decir $\sum P_i = 1$

Para comprender mejor estas ideas, consideremos el siguiente ejemplo:

Ejemplo 1. El gerente de la empresa que tenía negocios en Puebla y en la Ciudad de México, abrió recientemente una sucursal en Querétaro. Ahora sus negocios se reparten de la siguiente manera: 50% en la Cd. de México; 30% en Puebla y 20% en Querétaro. Los índices de precios de las tres ciudades se detallan a continuación para el mes de junio del año base.

| | I_i | P_i | $I_i P_i$ |
|-------------------------|---------|-------------|----------------|
| I.P.C. Cd México | 221.310 | 0.50 | 110.655 |
| I.P.C Puebla | 215.240 | 0.30 | 64.572 |
| I.P.C Querétaro | 218.700 | 0.20 | 43.740 |
| S U M A | | 1.00 | 218.967 |

El índice compuesto es ahora: $IC = 218.967$

Índices ponderados

Aún cuando los índices agregados incorporan mayor información que los índices simples, pueden no ser relevantes porque llegan a estar influidos por las unidades de medida; por ello, se necesita un medio de «ponderar» adecuadamente los artículos según su importancia relativa.



Entre los **índices compuestos** ponderados que más se utilizan se encuentran los que se refieren a las variaciones de precios. Los más importantes son los de

Laspeyres y de Paasche.

La característica común a estos índices y a la mayoría de los índices de precios es que utilizan como coeficientes de ponderación los valores que resultan del producto de un precio por una cantidad.

Estas ponderaciones se establecen para tener en cuenta las cantidades vendidas de cada producto. Lo anterior proporciona un reflejo más exacto del costo verdadero de la canasta típica del consumidor.

Índice de Laspeyres.

Utiliza las cantidades vendidas en el año base dentro del ponderador y permite comparaciones más significativas con el tiempo.

Su fórmula es la siguiente:

$$L = \frac{\sum p_n q_o}{\sum p_o q_o} \times 100$$

en donde: L es el índice de precios de Laspeyres.

P_n es el precio actual.

P_o es el precio base.

q_0 es la cantidad vendida en el periodo base.

En el siguiente ejemplo se ilustra este índice:

Una empresa posee una planta empacadora de carne y sus 3 principales productos tienen las siguientes cantidades de precio y venta en los últimos 3 años:

| Artículo | Unidad | Precio (\$) / unidad | | | Cantidad vendida | | |
|----------|--------|----------------------|------|------|------------------|------|------|
| | | 2003 | 2004 | 2005 | 2003 | 2004 | 2005 |
| Res | kilo | 30 | 33 | 45 | 250 | 320 | 350 |
| Cerdo | kilo | 20 | 22 | 21 | 150 | 200 | 225 |
| Ternera | kilo | 40 | 45 | 36.4 | 80 | 90 | 70 |

Si tomamos como año base el año 2003, y calculamos los productos precio en año corriente por cantidad en año base, tendríamos:

| Artículo | Precio (\$) / unidad | | | Cantidad año base | Precio x cantidad ($p_n \times q_0$) | | |
|--------------|----------------------|------|------|-------------------|--|---------------|---------------|
| | 2003 | 2004 | 2005 | | 2003 | 2004 | 2005 |
| Res | 30 | 33 | 45 | 250 | 7,500 | 8,250 | 11,250 |
| Cerdo | 20 | 22 | 21 | 150 | 3,000 | 3,300 | 3,150 |
| Ternera | 40 | 45 | 36.4 | 80 | 3,200 | 3,600 | 2,912 |
| TOTAL | | | | | 13,700 | 15,150 | 17,312 |

El índice para el año 2003 será:



$$L_{2003} = \frac{\sum p_{2003} q_{2003}}{\sum p_{2003} q_{2003}} \times 100 = \frac{13,700}{13,700} = 100$$

Esto era de esperarse, puesto que es el propio año base.

El índice para el año 2004 utiliza los precios del año de referencia (2004)

y las cantidades en el año base (2003). Se tendría entonces:

$$L_{2004} = \frac{\sum p_{2004} q_{2003}}{\sum p_{2003} q_{2003}} \times 100 = \frac{15,150}{13,700} = 110.58$$

El índice para el año 2005 utiliza los precios del año de referencia (2005)

y las cantidades en el año base (2003) para el numerador:

$$L_{2005} = \frac{\sum p_{2005} q_{2003}}{\sum p_{2003} q_{2003}} \times 100 = \frac{17,312}{13,700} = 126.36$$

Estos valores indican que del año 2003 al 2004, el precio de la canasta para estos 3 artículos se incrementó en un 10.58%. Se tendrían que gastar \$110.58 en 2004 para adquirir lo que en 2003 costaba \$100.00.

También indica que del año 2003 al 2005, el precio de la canasta para estos 3 artículos se incrementó en un 26.36%. Se gastarían \$126.36 en 2005 para adquirir lo que en 2003 se compraba con \$100.00.

Se hace notar que el denominador es el mismo para cualquier año, ya que el índice de Laspeyres siempre utiliza cantidades del periodo base.

Índice de Paasche



Incorpora dentro del ponderador las cantidades vendidas en cada uno de los años de referencia y tiene la ventaja de que el índice se basa en los patrones de comportamiento del consumidor común. A medida que los consumidores cambian sus hábitos de compra, los gastos que efectúan se reflejan directamente en el índice.

Su expresión es la siguiente:

$$P = \frac{\sum p_n q_n}{\sum p_o q_n} \times 100$$

en donde: P es el índice de precios de Laspeyres.

p_n es el precio actual.

p_o es el precio base.

q_n es la cantidad vendida en el periodo actual

Utilizaremos los mismos datos del caso anterior para calcular el índice de Paasche:

| Artículo | Unidad | Precio (\$) / unidad | | | Cantidad vendida | | |
|----------|--------|----------------------|------|------|------------------|------|------|
| | | 2003 | 2004 | 2005 | 2003 | 2004 | 2005 |
| Res | kilo | 30 | 33 | 45 | 250 | 320 | 350 |
| Cerdo | kilo | 20 | 22 | 21 | 150 | 200 | 225 |
| Ternera | kilo | 40 | 45 | 36.4 | 80 | 90 | 70 |



Con ello podemos formar los productos necesarios para calcular el índice:

| Artículo | Precio x Cantidad | | | | |
|----------|-------------------|----------------|----------------|----------------|----------------|
| | $P_{03}Q_{03}$ | $P_{04}Q_{04}$ | $P_{05}Q_{05}$ | $P_{03}Q_{04}$ | $P_{03}Q_{05}$ |
| Res | 7,500 | 10,560 | 15,750 | 9,600 | 10,500 |
| Cerdo | 3,000 | 4,400 | 4,725 | 4,000 | 4,500 |
| Tenera | 3,200 | 4,050 | 2,548 | 3,600 | 2,800 |
| Suma | 13,700 | 19,010 | 23,023 | 17,200 | 17,800 |

El índice para el año 2003 será:

$$P_{2003} = \frac{\sum P_{03}q_{03}}{\sum P_{03}q_{03}} \times 100 = \frac{13700}{13700} = 100$$

El índice para el año 2004 será:

$$P_{2004} = \frac{\sum P_{04}q_{04}}{\sum P_{03}q_{04}} \times 100 = \frac{19010}{17200} = 110.5$$

El índice para el año 2005 será:

$$P_{2005} = \frac{\sum P_{05}q_{05}}{\sum P_{03}q_{05}} \times 100 = \frac{23023}{17800} = 129.3$$

Este índice indica que del año 2003 al 2004, el precio de la canasta para estos 3 artículos se incrementó en un 10.5%. Se gastaría \$110.50 en 2004 para adquirir lo que con \$100.00 se compraba en 2003. También



indica que del año 2003 al 2005, el precio de la canasta para estos 3 artículos se incrementó en un 29.3%. Se gastaría \$129.30 en 2005 para comprar lo que en 2003 costaba \$100.00

El índice de Laspeyres tiende a sobreponderar los bienes cuyos precios se incrementan. Lo anterior ocurre porque el incremento en el precio tiende a reducir las cantidades vendidas, pero tal reducción en las cantidades compradas no se refleja en el índice de Laspeyres debido a que éste utiliza las cantidades del año base.

6.3 Principales índices de precios

Dentro de los principales índices de precios se tienen:

- El índice nacional de precios al consumidor
- Índice de precios productor
- El índice de precios y cotizaciones de la Bolsa de Valores
- Los índices Dow Jones

A continuación presentamos sus características generales.

Índice nacional de precios al consumidor

El índice de precios al consumidor es, en términos generales, un indicador que se construye para analizar la evolución de los precios al consumidor de un conjunto de bienes y servicios. A esto es lo que



generalmente se conoce en los medios noticiosos como el índice del costo de la vida.

En el caso de México se le conoce como el Índice nacional de precios al consumidor (INPC) y es elaborado por el Banco de México desde 1927, año en el que se investigaron 16 artículos alimenticios en la ciudad de México. Actualmente, se investigan y registran cada mes más de 170,000 precios para más de 300 tipos de productos en 46 ciudades del país y para su cálculo se sigue el modelo de Laspeyres.

En el caso de Estados Unidos el índice se publica mensualmente en Estados Unidos por la Dirección de Estadísticas Laborales (Bureau of Labor Statistics).

Los valores del índice de precios al consumidor se expresan como promedios anuales y mensuales.

Este índice se puede usar de múltiples maneras. Un uso común es para medir “el poder adquisitivo del consumidor” o el de la moneda.

El INPC se utiliza también para medir el ingreso “real”, que es el ingreso ajustado para cambios en los precios. De este modo, dividir el salario neto entre el valor corriente del INPC en cualquier año revelará el ingreso real para ese año. También es posible hacer una comparación entre años. Otra aplicación de uso generalizado de este índice tiene que ver con las denominadas cláusulas de “aumentos graduales” en los contratos colectivos de trabajo que ligan los aumentos salariales al índice nacional de precios al consumidor (INPC). Más aún, las cuotas al



Seguro Social y algunas otras incluyen cláusulas relativas a cambios que toman en consideración el nivel del INPC.

Todos los índices de precios al consumidor son índices agregados. Para entender como se trabaja un índice de precios al consumidor, haremos uso de un ejemplo sencillo simplificado.

Ejemplo 1: Supongamos que vivimos en una pequeña población que tiene un solo almacén general para surtir a los pobladores de todos los bienes, tanto de primera necesidad como suntuarios. En esta población deseamos comenzar a llevar un índice de precios al consumidor. Lo primero que necesitamos hacer es definir una canasta de bienes y servicios que la población en general consume o compra con regularidad. La conformación de esta canasta y el peso de cada artículo es fuente de polémica en todos los casos. En nuestro caso supondremos que ya nos pusimos de acuerdo y que la canasta de consumo mensual familiar aparece en la siguiente tabla.

| ARTÍCULO | PESO |
|---------------------------|------------|
| Frijol | 10 Kg |
| Maíz | 10 Kg |
| Jitomate | 4 Kg |
| Carne de res (bistec) | 2 Kg |
| Pollo (entero sin cabeza) | 3 Kg |
| Zapatos | 1/3 par * |
| Pantalón | ¼ unidad * |
| Televisión color 21" | 1/60 * |
| Refrigerador mediano | 1/20* |
| Automóvil compacto | 1/60* |
| Gasolina | 100 litros |

Nota: Los artículos marcados con (*) son los bienes que llamamos “de consumo duradero” ya que no se acaban en un mes, como sí sería el caso del maíz o del pollo. En este caso estaremos indicando que una televisión dura 5 años (60 meses) y que un refrigerador dura 10 años (120 meses), por lo que le asignamos a cada mes una parte proporcional de su valor.

El siguiente paso para construir nuestro índice será investigar en el almacén general el precio de cada uno de los artículos considerados para elaborar con ellos una tabla como la que se muestra a continuación. En la misma tabla aparecen los precios mensuales de los artículos; después, el cálculo del índice correspondiente.

Enseguida de la tabla aparece la explicación de cómo construimos el índice.

| ARTÍCULO | MES BASE | | | MES SIGUIENTE | |
|---------------------------|---------------|---------------------|--------------------|---------------------|--------------------|
| | PESO Ó UNIDAD | PRECIO UNITARI O \$ | COSTO POR ARTÍCULO | PRECIO UNITARI O \$ | COSTO POR ARTÍCULO |
| Frijol | 10 Kg | 2.20 | 22.00 | 2.30 | 23.00 |
| Maíz | 10 Kg | 16.00 | 160.00 | 16.00 | 160.00 |
| Jitomate | 4 Kg | 8.00 | 32.00 | 8.50 | 34.00 |
| Carne de res (bistec) | 2 Kg | 42.00 | 84.00 | 4.00 | 88.00 |
| Pollo (entero sin cabeza) | 3 Kg. | 30.00 | 90.00 | 28.00 | 84.00 |
| Zapatos | 1/3 par * | 300.00 | 100.00 | 300.00 | 100.00 |

| | | | | | |
|-------------------------|---------------|--------------|-------------------|-----------|--------------------|
| Pantalón | ¼ unidad * | 200.00 | 50.00 | 210.00 | 52.50 |
| Televisión color 21" | 1/60 * | 2,100.00 | 35.00 | 2,050.00 | 34.17 |
| Refrigerador mediano | 1/20* | 3,000.00 | 25.00 | 3,000.00 | 25.00 |
| Automóvil compacto | 1/60* | 40,000.00 | 666.67 | 40,000.00 | 666.67 |
| Gasolina | 100 litros | 6.00 | 600.00 | 6.05 | 605.00 |
| | | Total | \$1,864.67 | | \$ 1,872.34 |

Nota: En el caso del automóvil se trata del precio estimado del éste menos el valor de rescate, es decir, la pérdida de valor que sufre mientras su dueño lo utiliza.

El valor de nuestros artículos (con su peso correspondiente) en el mes de base fue de \$1,864.67 y en el mes siguiente fue de \$1,872.34. De acuerdo con la metodología de construcción del índice que ya vimos, la cantidad \$ 1,864.67 conforman el 100% de nuestro periodo base. Entonces nuestro índice del mes siguiente es:

$$INPC = \frac{1872.34}{1864.67} \times 100 = 100.41$$

Podemos criticar razonadamente nuestro índice. En este sentido, podemos argumentar que la ponderación del medio de transporte (automóvil y gasolina) es de dos tercios del total, además de que faltan muchos alimentos y el costo de la vivienda y la educación están completamente ausentes.



Con este sencillo ejemplo se puede uno percatar de lo complicado que es construir un índice de esta naturaleza. La complicación aumenta si consideramos que existen diversas calidades para un mismo producto (por ejemplo, pantalones de trabajo y de vestir o diversas calidades de carne o jitomate).

El conocimiento continuo de los principales **indicadores económicos** es de gran utilidad en todo momento para las empresas o negocios. También son una de las principales herramientas de predicción económica para cualquier gobierno.

Índice de precios al productor

Este es un índice elaborado, en el caso de nuestro país, por el Banco de México. En términos generales tiene como objetivo medir el cambio en los precios de una canasta fija de bienes y servicios que se producen en el país. Sus resultados se publican de manera mensual.

En torno a este índice es importante señalar que el concepto de precio productor se refiere al precio fijado por el productor a quien primero adquiere su producto, y por lo tanto no se refiere al valor de la producción o su costo.

Estos precios se recaban directamente del productor principalmente bajo el concepto “libre a bordo” ya que no incluyen impuestos al consumo, costos de transporte y márgenes de comercialización.



En este sentido uno de sus principales usos es medir la inflación desde la perspectiva de la oferta.

Índice de precios y cotizaciones (IPC)

Este es un índice elaborado diariamente por la Bolsa Mexicana de Valores con el propósito de medir el cambio en el nivel general de precios de las acciones que forman el mercado respectivo.

Actualmente, en su cálculo se integra información de una muestra de 35 emisoras, las cuales se seleccionan bimestralmente tomando en cuenta aspectos como número de operaciones y razón entre el monto operado y el monto suscrito, buscando con ello una muestra balanceada y debidamente ponderada.

Se puede decir, en términos generales que su valor es un indicador para el inversionista de los rendimientos que se obtienen al colocar dinero en el mercado accionario.

Índice Dow Jones

Bajo este nombre se tienen más de 130,000 indicadores económicos, todos ellos etiquetados como Dow Jones al ser elaborados por la empresa Dow Jones Indexes. De ellos, uno de los más famosos es el índice bursátil Dow Jones, que es un índice semejante al de precios y cotizaciones de la Bolsa Mexicana de Valores, por cuanto mide el cambio en el nivel general de precios de las acciones, pero del mercado accionario norteamericano.



Los registros de este índice constituyen la serie de indicadores bursátiles mas antigua y conocida ya que existe desde finales del siglo XIX. Debe su nombre a la iniciativa de dos editores financieros: Charles H. Dow y Edward D. Jones

Actualmente, en el cálculo del índice se integra información de 30 emisoras.

6.4 Deflación de una serie

Estrictamente hablando, la deflación es una disminución sostenida en el nivel de precios de los bienes y servicios. Es, así, un concepto opuesto al de inflación.

Ambos conceptos se refieren, en este sentido, a la variación que experimenta el dinero por cuanto medio de pago.

Todos cuantos hemos pagado por un conjunto de bienes y servicios a lo largo de un periodo de tiempo nos hemos percatado, por lo menos, de que con la misma cantidad de dinero cada vez compramos cantidades menores de esos bienes y servicios por efecto de la inflación. Asimismo nos percatamos de que aún si se aumentan los sueldos, compramos cantidades menores de esos bienes y servicios, de igual manera, por efecto de la inflación.

Si los precios bajasen de manera sostenida nos percataríamos de que adquiriríamos una cantidad mayor de bienes y servicios por efecto de la deflación.



En general, a este proceso de cambio en las cantidades de bienes y servicios que se pueden adquirir se le conoce como cambios en el poder adquisitivo.

Medir estos cambios lleva a comparar magnitudes referidas a instantes de tiempo distintos y esto hace necesario distinguir dos conceptos.

Precios corrientes. Se refiere a mediciones de cantidades con los precios del periodo que corre.

Precios constantes. Se refiere a mediciones de cantidades con los precios de un periodo base.

Los precios constantes se pueden obtener a partir de los precios corrientes mediante un proceso que se conoce como deflactación de series, el cual consiste en dividir cada precio corriente entre un índice explicado por la tasa de inflación.

La expresión respectiva se muestra a continuación:

$$VPCons = \frac{VPCorr}{(1 + \text{tasa inf lación})^n},$$

donde n es el número de periodos que hay entre el periodo base y el corriente.

Consideremos el siguiente ejemplo.



Ejemplo 1. Una empresa dedicada a la fabricación de muebles para oficina facturó en el año de 2008 ventas por \$4 567,000. Las cifras correspondientes a 2009 y 2010 son, respectivamente, \$4 872,345 y \$5 103,576. Aparentemente, hay un incremento en las ventas, sin embargo, el dueño sabe que durante ese periodo se registro una tasa de inflación significativa que afecta las comparaciones. Suponiendo que la tasa de inflación para ese periodo fue de 3.60% anual, ¿qué se puede decir respecto del volumen de ventas?

Solución:

Construyamos primero una tabla donde incorporemos la serie de valores.

| Año | Ventas | Inflación |
|------|-----------|-----------------------|
| 2008 | 4'567,000 | |
| 2009 | 4 872,345 | De 2008 a 2009, 3.60% |
| 2010 | 5 103,576 | De 2009 a 2010, 3.60% |

Los valores registrados de las ventas corresponden a precios corrientes. El incremento en el monto de las ventas se debe, por lo menos en parte, al efecto del incremento general de los precios, expresado por una tasa de inflación anual de 3.6%. Esto lleva a la duda de si realmente las ventas han crecido.

Por lo tanto, procedemos a valorar las cantidades (las ventas) según las cifras del año base (2008), esto es, procedemos a obtener los precios constantes.



Tendríamos:

- Para 2009 las ventas a precios constantes serían

$$4\,872,345 / 1.036 = 4\,703,035.71$$

- Para 2010, éstas serían:

$$5\,103,576 / (1.036)^2 = 4\,755,049.87$$

Finalmente podemos estructurar el siguiente cuadro

| Año | Ventas | |
|------|--------------------|--------------------|
| | Precios corrientes | Precios constantes |
| 2008 | 4'567,000 | 4'567,000.00 |
| 2009 | 4 872,345 | 4 703,035.71 |
| 2010 | 5 103,576 | 4 755,049.87 |

Ahora sí podemos afirmar que hay un incremento real en el monto de las ventas.



RESUMEN DE LA UNIDAD

En esta unidad se revisan los indicadores por medio de los cuales se puede presentar información resumen de los cambios relativos sufridos por precios y cantidades entre dos periodos. En este sentido se introduce el concepto de número índice como un elemento que permite hacer comparaciones respecto de un punto de referencia o base del índice. Dentro de tales indicadores se incluyen los índices simples, los compuestos y los ponderados, y como parte de éstos últimos, los de Laspeyres y Paasche. Finalmente, se presenta el índice nacional de precios al consumidor, número índice elaborado por el Banco de México para medir el incremento en los precios de una canasta de productos y servicios de consumo generalizado.



GLOSARIO

Deflactar

Operación por la cual se ajusta o reduce a precios constantes una cantidad monetaria, descontando con este propósito el efecto de la inflación.

Índice agregado

Es un indicador que se usa para determinar las variaciones en el tiempo de las cantidades o precios de varios bienes y/o servicios.

Índice compuesto

Es un indicador que se obtiene al ponderar otros índices con la condición de que las ponderaciones sumen uno (ó 100%).

Índice simple

Es un indicador que se usa para determinar las variaciones en el tiempo de las cantidades o precios de un solo bien y/o servicio.

Inflación

Es el incremento en el precio de los bienes y servicios como consecuencia de la pérdida de valor del dinero.

Número índice

Es un estadístico que asigna un número único a un conjunto de datos estadísticos a efecto de cuantificar tendencias y/o cambios de precio o cantidad entre dos periodos.

Periodo base

Es el periodo contra el cual se efectúan comparaciones de cambios relativos. El valor del índice respectivo es 100.



ACTIVIDADES DE APRENDIZAJE

ACTIVIDAD 1

Reflexiona sobre cinco situaciones cotidianas de tu vida en torno a cada una de las cuales puedas construir un índice simple, bien sea de valor o de cantidad. Explica con detalle el significado o utilidad de cada uno.

ACTIVIDAD 2

Una planta industrial ha desarrollado un catálogo de actividades laborales a través de la cual calcula parte de la nómina. Por el momento desea conocer un índice de cantidad para las horas trabajadas. Los datos disponibles son:

| Actividad | Precio base (\$/hora) | Horas trabajadas | |
|-----------|--------------------------|------------------|-------|
| | | marzo | junio |
| 1 | 15 | 729 | 842 |
| 2 | 18 | 632 | 615 |
| 3 | 27 | 153 | 179 |
| 4 | 22 | 426 | 316 |
| 5 | 14 | 519 | 418 |
| 6 | 16 | 650 | 750 |
| 7 | 19 | 512 | 562 |

Determina el valor de dicho índice.



ACTIVIDAD 3

En la tabla siguiente se muestran los datos relativos a destinos turísticos, número de viajeros y costo de transporte desde una ciudad del interior de la república para dos años distintos.

| Destino | 2000 | | 2007 | |
|----------|--------------------|-------|--------------------|-------|
| | Número de viajeros | Costo | Número de viajeros | Costo |
| Cancún | 4522 | 5250 | 3595 | 3170 |
| Acapulco | 8729 | 2730 | 10578 | 3270 |
| Mazatlán | 1545 | 3120 | 3264 | 3720 |
| Huatulco | 893 | 3250 | 960 | 3900 |

Tomando el año 2000 como año base, determina el valor de los índices de Laspeyres y de Paasche.



ACTIVIDAD 4

Construye una canasta familiar de 20 bienes y servicios. Define para cada uno de ellos el consumo mensual promedio que se registra en tu familia.

Por otro lado, realiza un ejercicio de memoria y registra los precios de esos bienes y servicios para los últimos seis meses.

Con estos datos, construye tu índice familiar de precios de bienes y servicios para los últimos seis meses.

No olvides incluir todos los cuadros que hayas elaborado.

ACTIVIDAD 5

Estudia el documento *Preguntas frecuentes (INPC)*, publicado por el Banco de México

A partir de esta lectura, elabora un cuadro sinóptico sobre el tema Números índice e INPC.



ACTIVIDAD 6

Investiga cuáles han sido las tasas de inflación anuales para el caso de México en los últimos 10 años y además, para esos mismos años, los salarios mínimos generales. Con esos datos, deflacta la serie de salarios mínimos. Redacta además una conclusión a partir de tus resultados.

ACTIVIDAD 7

Con el propósito de adquirir un cartucho para su impresora, Juan acude todos los años, en el mes de mayo, a un almacén que vende a mayoreo, medio mayoreo y menudeo. En su visita al almacén aprovecha para adquirir algunos otros productos. Un día, movido por la curiosidad, se decide a calcular en qué medida se han incrementado los precios, por lo que busca entre sus papeles y encuentra las notas de los tres últimos años, con los cuales elabora la siguiente tabla:

| Producto | Presentación | 2005 | 2006 | 2007 |
|----------------------|---------------------------------|------------|------------|------------|
| Frijol | (Pq con 3 bolsas de un kilo) | 45 | 48 | 51 |
| Leche | (caja con 10 litros) | 67 | 69 | 70 |
| Aceite | (Pq con 4 botellas de un litro) | 52 | 60 | 64 |
| Cartucho p/impresora | (caja con un cartucho) | 430 | 452 | 490 |
| TOTAL | | 594 | 629 | 675 |

Decide dividir el total pagado en 2006 y en 2007 entre el total pagado en 2005. De acuerdo con sus cálculos, los precios cambiaron 5.89% de 2005 a 2006 y 13.64% de 2005 a 2007. Jugando con las cifras piensa que el hecho de que haya



adquirido algunos productos a medio mayoreo puede estar influyendo en sus resultados, por lo que decide realizar los cálculos considerando precios por unidad (por kilo y por litro), de modo que elabora una segunda tabla, misma que se muestra a continuación:

| Producto | Presentación | 2005 | 2006 | 2007 |
|----------------------|------------------------|------|------|------|
| Frijol | (un kilo) | 15 | 16 | 17 |
| Leche | (un litro) | 6.7 | 6.9 | 7 |
| Aceite | (un litro) | 13 | 15 | 16 |
| Cartucho p/impresora | (caja con un cartucho) | 430 | 452 | 490 |

Cuando realiza sus cálculos se encuentra con que los precios cambiaron 5.42% de 2005 a 2006 y 14.05% de 2005 a 2007.

Juan confirma que algo tiene que ver la presentación de los distintos artículos, pero además se percata que entre ambos mecanismos de cálculo hay una diferencia clara en los resultados ya que con el segundo procedimiento los precios cambiaron menos entre 2005 y 2006 y más entre 2005 y 2007 en comparación a lo que ocurre con el primer procedimiento.

Escribe un breve informe que explique el porqué de estas diferencias



CUESTIONARIO DE REFORZAMIENTO

1. ¿Cuáles son los principales elementos para construir un numero índice?
2. ¿Cuál es la utilidad de trabajar con números índice?
3. Haga referencia de la utilidad de conocer y seguir el comportamiento del índice de una bolsa de valores.
4. ¿Cuáles son las características de un índice compuesto?
5. Indique la diferencia entre un índice de cantidad y un índice de valor.
6. Explique la diferencia que existe entre un índice agregado y un índice simple.
7. ¿A qué se refiere el nivel conceptual y el nivel operacional?
8. ¿Qué es un índice ponderado y su utilización?
9. ¿Cuál es la utilidad del índice de Laspeyres?
10. ¿En qué consiste el Índice Nacional de Precios al Consumidor y qué beneficios proporciona su conocimiento?



LO QUE APRENDÍ

Al inicio de esta unidad te planteamos la siguiente situación:

De acuerdo con cifras del Banco de México, la inflación (acumulada) en el periodo 2000-2006 fue como se muestra en la siguiente tabla:

| Año | Inflación |
|------|-----------|
| 2001 | 4.4 |
| 2002 | 5.7 |
| 2003 | 3.98 |
| 2004 | 5.19 |
| 2005 | 3.33 |
| 2006 | 4.05 |

¿Consideras que tasas de inflación como estas pueden afectar tu vida?

Con los conocimientos que has adquirido al estudiar esta unidad ya puedes aplicar el índice de precios al consumidor así como el concepto de inflación. También puedes ya interpretar y explicar sus implicaciones. Por ejemplo, puedes analizar si el valor del índice publicado por el Banco de México refleja el incremento en los precios de los bienes y servicios que tú consumes cotidianamente, o si lo subestima o sobrestima. Expón tus conclusiones respecto de este tema tomando como eje de tus comentarios tu capacidad de ahorro y comenta con tu asesor.



EXAMEN DE AUTOEVALUACIÓN

I. El precio de la tonelada de maíz fluctúa de acuerdo con las condiciones del mercado internacional. A continuación se listan los precios por tonelada al cierre de algunos meses. Si el mes base es junio, determina el índice simple de precios para los diferentes meses, escribe la respuesta en el recuadro correspondiente.

| Mes | \$ / tonelada | Índice |
|---------|---------------|--------|
| Febrero | 1210 | |
| Marzo | 1180 | |
| Abril | 1040 | |
| Mayo | 1314 | |
| Junio * | 1200* | |
| Julio | 1190 | |
| Agosto | 1220 | |

* mes base



II. Un estudio indicó que para transportarse de su casa a la escuela, a la biblioteca u otros lugares, el estudiante universitario promedio utiliza los transportes que abajo se detallan. Con estos datos deseamos crear un índice del costo de transporte.

| | Precios unitarios (periodo base) | Costo total |
|-----------------------|-------------------------------------|-------------|
| 8 viajes en metro | \$1.50 | \$12.00 |
| 6 viajes en colectivo | \$2.50 | \$15.00 |
| 4 viajes en autobús | \$2.00 | \$ 8.00 |
| 2 viajes en taxi | \$10.00 | \$20.00 |

Seis meses después se actualizan los precios unitarios y se obtiene la siguiente información.

| | | Viajes semana* | Costo total |
|-----------|--------|----------------|-------------|
| Metro | \$2.00 | 8 | \$16.00 |
| Colectivo | \$2.00 | 6 | \$12.00 |
| Autobús | \$3.00 | 4 | \$12.00 |
| Taxi | \$9.00 | 9 | \$18.00 |

* De la tabla anterior

Determina los siguientes valores:

| | |
|--|--|
| El costo de transporte en la semana o periodo base fue: | |
| El costo del transporte en la semana o periodo actual fue: | |
| El índice de este periodo es: | |



III. Se desea determinar el índice de precios para el conjunto de bienes y servicios que se muestra en la siguiente tabla. Anota las cifras en los cuadros en blanco.

| ARTÍCULO | MES BASE | | | MES SIGUIENTE | |
|---------------------------|---------------|--------------------|--------------------|--------------------|--------------------|
| | PESO Ó UNIDAD | PRECIO UNITARIO \$ | COSTO POR ARTICULO | PRECIO UNITARIO \$ | COSTO POR ARTÍCULO |
| Frijol | 2 | 14 | | 14.5 | |
| Maíz | 1 | 8 | | 11 | |
| Jitomate | 1 | 17 | | 24 | |
| Carne de res (bistec) | 2 Kg | 84 | | 82 | |
| Pollo (entero sin cabeza) | 3 Kg. | 42 | | 44 | |
| Zapatos | 1/3 par * | 450 | | 500 | |
| Televisión color 21" | 1/60 * | 3,500.00 | | 3,500.00 | |
| Refrigerador mediano | 1/20* | 7,500.00 | | 8,000.00 | |
| Transporte | 80 viajes | 2.5 | | 3 | |
| Total | | | | | |

Índice de precios=



IV. Relaciona las siguientes columnas, anotando en el paréntesis el número que corresponda de las situaciones que se presentan en la columna izquierda.

| | | |
|---|-----|----------------------------------|
| (1) Se le puede considerar un indicador de precios del "lado de la demanda" | () | Índice de Laspeyres |
| (2) Es un primer indicador de la evolución futura de la inflación | () | Índice simple |
| (3) Es un indicador del cambio en el precio de un solo bien o servicio | () | Índice compuesto |
| (4) Rasgo básico de índices ponderados de precios | () | Índice de precios productor |
| (5) Reflejan el verdadero poder de compra de los salarios | () | Deflación |
| (6) Mide las variaciones en los precios pivotando sobre los consumos de un periodo base | () | Salarios reales |
| (7) Incluye dos o mas índices simples | () | Índice de precios al consumidor |
| (8) Proceso por el cual se hace posible comparar dos o mas series de precios | () | Inflación |
| (9) Nombre con el que se conoce al proceso sostenido de aumento en los precios | () | La suma de los ponderadores es 1 |



MESOGRAFÍA

BIBLIOGRAFÍA RECOMENDADA

| Autor | Capítulo | Páginas |
|--------------------------------|--|----------|
| 1. Anderson, Sweeney, Williams | 17. Indicadores Sección 17.1 Relativos de precios | 733 |
| | 17.2 Índices agregados de precios | 733-736 |
| | 17.3 Cálculo de un índice agregado de precios a partir de relativos de precios | 737-738 |
| | 17.4 Algunos índices importantes de precios | 742 -744 |
| 2. Levin y Rubin. | 16. Números índice Sección 16.1 Definición de número índice. | 720-723 |
| | 16.2 Índice de agregados no ponderados | 723- 727 |
| | 16.3 Índice de agregados ponderados. | 727- 734 |
| | 16.4 Métodos de promedio de | 735-740 |



| | | |
|---------------------------------|--|-----------|
| | relativos. | |
| | Sección 16.5 Índices de cantidad y de valor. | 740-744 |
| 3. Lind, Marchal, Wathen. | 15. Números índice | 570 - 573 |
| | Números índice simples. | |
| | ¿Por qué convertir datos en índices? | 573 |
| | Elaboración de números índice | 573 - 575 |
| | 15. Números índice | 575 - 577 |
| | Índices no ponderados. | |
| | Índices ponderados. | 577 - 581 |
| | Índice de precios al consumidor. | 588 - 591 |
| | Índices para fines especiales | 583 - 587 |
| Índice de precios al consumidor | 588-591 | |



BIBLIOGRAFÍA BÁSICA

1. Anderson, David R.; Sweeney, Dennis J.; Williams, Thomas A., (2005), *Estadística para administración y economía*, 8a. edición, México, International Thomson editores, 888 páginas más apéndices.
2. Levin, Richard I. y David S. Rubin, (2004), *Estadística para administración y economía*, 7a. Edición, México, Pearson Educación Prentice Hall, 826 páginas más anexos.
3. 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.



BIBLIOGRAFÍA COMPLEMENTARÍA

1. Bowerman Bruce, *Pronósticos, series de tiempo y regresión; un enfoque aplicado*, México: Cengage Learning, 4ª edición, 2007, 720 pp.
2. Mendenhall William, *Introducción a la probabilidad y estadística*, México: Cengage Learning, 13ª edición, 2010, 776 pp.
3. Webster Allen L., *Estadística I aplicada a los negocios y la economía*, México: McGraw-Hill, 2ª. edición, 2002, 154 pp.



SITIOS DE INTERNET

- Banco de México, *Preguntas frecuentes (INPC)*, en el sitio:
<http://www.banxico.org.mx/polmoneinflacion/didactico/preguntasfrecuentes/PreguntasFrecuentesINPC.html>
- de Ribeiro, Washington; Rodríguez, Eloy y Cervini, María (de la Escuela de administración de la facultad de ciencias económicas y de administración de la Universidad de la República), *Números Índice: conceptos y aspectos generales*, págs. 30-34 del capítulo 3, del hipertexto Indicadores de la actividad económica, en el sitio:
<http://www.fing.edu.uy/catedras/economia/descriptiva.doc>
- Facultad de Ciencias Económicas y Empresariales de la Universidad de Valladolid, *Deflación de series monetarias*, de la serie de materiales escritos y multimedia para la docencia de la Estadística orientada a la Economía y a la Empresa.
www2.eco.uva.es/estadmed/datos/indices/indices6.htm
- Guillermo López Dumrauf, *Tasas de ajuste de interés, índices y coeficientes*, de la Universidad del Cema, Argentina, en el sitio:
http://www.cema.edu.ar/~ql24/Slides/Indices_y_coeficientes_de_ajuste.pdf



RESPUESTAS A LOS EXÁMENES DE AUTOEVALUACIÓN

UNIDAD 1

Solución:

I.

| | |
|-----------------------------------|-------|
| Ingreso total de los casados | 19800 |
| Ingreso promedio | 3300 |
| Total de hijos de hombres casados | 6 |
| Promedio | 1.5 |
| Total de hijos de mujeres casadas | 6 |
| Promedio | 3 |
| Ingreso total de los solteros | 12800 |
| Ingreso promedio | 3200 |

Solución:

1. Inferencia
2. Conocimiento
3. Información
4. Datos
5. Muestra
6. Descriptiva
7. Parámetro

Unidad 2



Solución:

1. (N)
2. (N)
3. (R)
4. (I)
5. (R)
6. (N)
7. (I)
8. (R)
9. (O)

Solución

Distribución Geográfica de Empleados por Sector

| Sector | Ubicación | | Total |
|---------------|-------------|-------------|-------------|
| | México | Sudamérica | |
| Agroindustria | 458 | 59 | 517 |
| Agropecuario | 345 | 712 | 1057 |
| Transporte | 125 | 41 | 166 |
| Minería | 326 | 115 | 441 |
| Servicios | 1578 | 321 | 1899 |
| Total | 2832 | 1248 | 4080 |

Fuente: Informe Labores. 2005

Solución:

| | |
|----------------------------------|-------------|
| Suma de calificaciones | 36.3 |
| Media de calificaciones | 4.54 |
| Observación donde cae la mediana | 4.5 |
| Mediana | 4.2 |

Solución:

| | |
|----------------------------------|--------------|
| Suma de desviaciones cuadráticas | 15.34 |
| Varianza | 2.19 |
| Desviación estándar | 1.48 |
| Coeficiente de variación | 0.326 |

Solución:



| | | |
|------|-------|-------|
| 15.3 | 15.34 | |
| 2.1 | 2.19 | |
| 1.4 | 1.48 | |
| .3 | .32 | .326 |
| 0.3 | 0.32 | 0.326 |

Solución:

Inciso (c)

UNIDAD 3

Solución:

a) multiplicación

b) 8

Solución: tabla 2

a) 512

b) 336

UNIDAD 4

Solución:

1. (V)
2. (V)
3. (F)
4. (V)
5. (F)
6. (V)
7. (F)

Solución:

| | |
|--------------------|---------|
| ¿Sea roja? | 7 / 15 |
| ¿No sea verde? | 10 / 15 |
| ¿Sea Roja o verde? | 12 / 15 |



| | |
|-----------------------|--------|
| ¿Sea azul y verde? | 0 |
| ¿No sea azul ni roja? | 5 / 15 |

Solución:

| | |
|------------------------------|-----|
| $P(A)$ | 0.6 |
| $P(B^c)$ | 0.4 |
| $P(A \cap B)$ | 0.4 |
| $P(A - (A \cap B))$ | 0.2 |
| $P(A^c \cap B^c)$ | 0.2 |
| $P((A \cap B) \cap B)$ | 0.4 |
| $P(A^c \cap B)$ | 0.2 |
| $P((A \cup B) - (A \cap B))$ | 0.4 |
| $P(A^c \cup B^c)$ | 0.6 |

Solución:

- a. 0.6
- b. 0.2
- c. 0.33
- d. 0.57



ANEXO 1

MANUAL DE ESTADÍSTICA

David Ruiz Muñoz

Capítulo I: HISTORIA DE LA ESTADÍSTICA

Los comienzos de la estadística pueden ser hallados en el antiguo Egipto, cuyos faraones lograron recopilar, hacia el año 3050 antes de Cristo, prolijos datos relativos a la población y la riqueza del país. De acuerdo al historiador griego Heródoto, dicho registro de riqueza y población se hizo con el objetivo de preparar la construcción de las pirámides. En el mismo Egipto, Ramsés II hizo un censo de las tierras con el objeto de verificar un nuevo reparto

En el antiguo Israel la Biblia da referencias, en el libro de los Números, de los datos estadísticos obtenidos en dos recuentos de la población hebrea. El rey David por otra parte, ordenó a Joab, general del ejército hacer un censo de Israel con la finalidad de conocer el número de la población.



También los chinos efectuaron censos hace más de cuarenta siglos. Los griegos efectuaron censos periódicamente con fines tributarios, sociales (división de tierras) y militares (cálculo de recursos y hombres disponibles). La investigación histórica revela que se realizaron 69 censos para calcular los impuestos, determinar los derechos de voto y ponderar la potencia guerrera.

Pero fueron los romanos, maestros de la organización política, quienes mejor supieron emplear los recursos de la estadística. Cada cinco años realizaban un censo de la población y sus funcionarios públicos tenían la obligación de anotar nacimientos, defunciones y matrimonios, sin olvidar los recuentos periódicos del ganado y de las riquezas contenidas en las tierras conquistadas. Para el nacimiento de Cristo sucedía uno de estos empadronamientos de la población bajo la autoridad del imperio.

Durante los mil años siguientes a la caída del imperio Romano se realizaron muy pocas operaciones Estadísticas, con la notable excepción de las relaciones de tierras pertenecientes a la Iglesia, compiladas por Pipino el Breve en el 758 y por Carlomagno en el 762 DC. Durante el siglo IX se realizaron en Francia algunos censos parciales de siervos. En Inglaterra, Guillermo el Conquistador recopiló el Domesday Book o libro del Gran Catastro para el año 1086, un documento de la propiedad, extensión y valor de las tierras de Inglaterra. Esa obra fue el primer compendio estadístico de Inglaterra.

Aunque Carlomagno, en Francia; y Guillermo el Conquistador, en Inglaterra, trataron de revivir la técnica romana, los métodos estadísticos permanecieron casi olvidados durante la Edad Media.



Durante los siglos XV, XVI, y XVII, hombres como Leonardo de Vinci, Nicolás Copérnico, Galileo, Neper, William Harvey, Sir Francis Bacon y René Descartes, hicieron grandes operaciones al método científico, de tal forma que cuando se crearon los Estados Nacionales y surgió como fuerza el comercio internacional existía ya un método capaz de aplicarse a los datos económicos.

Para el año 1532 empezaron a registrarse en Inglaterra las defunciones debido al temor que Enrique VII tenía por la peste. Más o menos por la misma época, en Francia la ley exigió a los clérigos registrar los bautismos, fallecimientos y matrimonios. Durante un brote de peste que apareció a fines de la década de 1500, el gobierno inglés comenzó a publicar estadísticas semanales de los decesos. Esa costumbre continuó muchos años, y en 1632 estos Bills of Mortality (Cuentas de Mortalidad) contenían los nacimientos y fallecimientos por sexo. En 1662, el capitán John Graunt usó documentos que abarcaban treinta años y efectuó predicciones sobre el número de personas que morirían de varias enfermedades y sobre las proporciones de nacimientos de varones y mujeres que cabría esperar. El trabajo de Graunt, condensado en su obra *Natural and Political Observations...Made upon the Bills of Mortality* (Observaciones Políticas y Naturales. Hechas a partir de las Cuentas de Mortalidad), fue un esfuerzo innovador en el análisis estadístico.

Por el año 1540 el alemán Sebastián Muster realizó una compilación estadística de los recursos nacionales, comprensiva de datos sobre organización política, instrucciones sociales, comercio y poderío militar. Durante el siglo XVII aportó indicaciones más concretas de métodos de observación y análisis cuantitativo y amplió los campos de la inferencia y la teoría Estadística.



Los eruditos del siglo XVII demostraron especial interés por la Estadística Demográfica como resultado de la especulación sobre si la población aumentaba, decrecía o permanecía estática

En los tiempos modernos tales métodos fueron resucitados por algunos reyes que necesitaban conocer las riquezas monetarias y el potencial humano de sus respectivos países. El primer empleo de los datos estadísticos para fines ajenos a la política tuvo lugar en 1691 y estuvo a cargo de Gaspar Neumann, un profesor alemán que vivía en Breslau. Este investigador se propuso destruir la antigua creencia popular de que en los años terminados en siete moría más gente que en los restantes, y para lograrlo hurgó pacientemente en los archivos parroquiales de la ciudad. Después de revisar miles de partidas de defunción pudo demostrar que en tales años no fallecían más personas que en los demás. Los procedimientos de Neumann fueron conocidos por el astrónomo inglés Halley, descubridor del cometa que lleva su nombre, quien los aplicó al estudio de la vida humana. Sus cálculos sirvieron de base para las tablas de mortalidad que hoy utilizan todas las compañías de seguros.

Durante el siglo XVII y principios del XVIII, matemáticos como Bernoulli, Francis Maseres, Lagrange y Laplace desarrollaron la teoría de probabilidades. No obstante durante cierto tiempo, la teoría de las probabilidades limitó su aplicación a los juegos de azar y hasta el siglo XVIII no comenzó a aplicarse a los grandes problemas científicos.



Godofredo Achenwall, profesor de la Universidad de Gotinga, acuñó en 1760 la palabra estadística, que extrajo del término italiano statista (estadista). Creía, y con sobrada razón, que los datos de la nueva ciencia serían el aliado más eficaz del gobernante consciente. La raíz remota de la palabra se halla, por otra parte, en el término latino status, que significa estado o situación; Esta etimología aumenta el valor intrínseco de la palabra, por cuanto la estadística revela el sentido cuantitativo de las más variadas situaciones.

Jacques Quételet es quien aplica las Estadísticas a las ciencias sociales. Este interpretó la teoría de la probabilidad para su uso en las ciencias sociales y resolver la aplicación del principio de promedios y de la variabilidad a los fenómenos sociales. Quételet fue el primero en realizar la aplicación práctica de todo el método Estadístico, entonces conocido, a las diversas ramas de la ciencia.

Entretanto, en el período del 1800 al 1820 se desarrollaron dos conceptos matemáticos fundamentales para la teoría Estadística; la teoría de los errores de observación, aportada por Laplace y Gauss; y la teoría de los mínimos cuadrados desarrollada por Laplace, Gauss y Legendre. A finales del siglo XIX, Sir Francis Gaston ideó el método conocido por Correlación, que tenía por objeto medir la influencia relativa de los factores sobre las variables. De aquí partió el desarrollo del coeficiente de correlación creado por Karl Pearson y otros cultivadores de la ciencia biométrica como J. Pease Norton, R. H. Hooker y G. Udny Yule, que efectuaron amplios estudios sobre la medida de las relaciones.



Los progresos más recientes en el campo de la Estadística se refieren al ulterior desarrollo del cálculo de probabilidades, particularmente en la rama denominada indeterminismo o relatividad, se ha demostrado que el determinismo fue reconocido en la Física como resultado de las investigaciones atómicas y que este principio se juzga aplicable tanto a las ciencias sociales como a las físicas.

Tomado de Biblioteca Virtual

<http://www.eumed.net/cursecon/libreria/drm/1a.htm>



ANEXO 2

Los censos y los impuestos

Una de las bases de la prosperidad del Imperio fue su inmensa capacidad recaudatoria de impuestos y tributos. Ya durante la República, una vez vencidas las amenazas cartaginesa y macedonia, y sojuzgado el reino Selúcida, los romanos tuvieron por primera vez a su disposición una ingente cantidad de contribuyentes (voluntarios o forzosos) al Tesoro de Roma.



En la medida que las conquistas romanas fueron extendiéndose, los ingresos por impuestos siguieron creciendo, al igual que lo hacía la complejidad de recaudarlos. Evidentemente, no era lo mismo recolectar los tributos de una provincia que los de muchas, así como tampoco debía serlo cuando se trataba de regiones tan distantes como Britania, Bitinia o Egipto. Una de las soluciones que adoptaron los romanos, haciendo gala de su habitual pragmatismo, fue la de subcontratar a los temidos publicanos. Estos no sólo eran agentes de recaudación, sino que llegaban mucho más allá. El Estado romano les vendía la exclusiva de recaudar los impuestos de una



determinada región, a cambio de una cantidad fija. Los publicanos después podían llegar a recaudar más, no estando obligados a liquidar el excedente.

Para Roma este tipo de acuerdo era muy conveniente, ya que le permitía anticipar los ingresos y conocer de antemano la cantidad exacta que se iba a recaudar. Como todo empresario o político sabrá, esta solución es buena para quien quiera equilibrar un presupuesto. Sin embargo, podía llevar también a que los publicanos, en su afán por amortizar el dinero entregado al Estado, o simplemente por instinto rapaz, estrangularan fiscalmente a la población de las provincias. Esto llevó a no pocas revueltas y levantamientos por los desmanes de estos recaudadores fiscales subcontractados. Se reguló su actividad a finales de la república, pero si nos atenemos a lo que nos explica este texto del S. III dC, esta cuestión nunca quedó bien resuelta.

Lo que aumentó la catástrofe nacional y el luto general fue el nuevo censo impuesto a las provincias y a las ciudades. Los censores estaban diseminados por doquier y lo ponían todo en desorden como si se tratara de cosas de enemigos o de prisioneros o de esclavos. Se medían los campos en masa, se numeraban los árboles y las vides, se registraban los animales de todo tipo. Se usaba también el mismo procedimiento con las personas: los habitantes de las ciudades y de los campos eran reunidos masivamente en una plaza; todas las plazas regurgitaban grupos de siervos y de esclavos, porque cada uno debía estar presente con sus hijos y siervos. Por todas partes se oía un sonido de lamentos y de golpes; los hijos eran incitados contra los padres, y los siervos, hasta ese momento fidelísimos, contra sus patronos, las mujeres contra los



maridos, para que denunciaran la verdadera entidad de sus ingresos. Si la cosa no salía bien, ellos mismos eran sometidos a la tortura hasta que, bajo la presión del dolor, confesaban y denunciaban incluso lo que no tenían. No se aceptaba ninguna excusa de edad o de salud. Se hacía comparecer incluso a los enfermos y a los decrepitos; se estimaba la edad de cada uno, a los niños se les añadía y a los ancianos se les quitaban arbitrariamente los años. Por todas partes reinaba un llanto y una tristeza grandísima. No se daba fe a las declaraciones de los censores y se enviaba siempre a otros nuevos para que encontraran alguna cosa más; y si tampoco estos últimos lo conseguían, se añadían otras tasas para no dar la impresión de haberlos enviado en vano. Entretanto disminuían los animales domésticos, los hombres morían. Pero eso no impedía que hubiera que pagar los impuestos incluso por los muertos. En pocas palabras, ya no se podía ni vivir ni morir sin pagar impuestos. Sólo quedaban los mendigos, de los que era inútil esperar algún tributo: su miseria e infelicidad los hacía, por lo menos, inmunes a cualquier injusticia. Pero este hombre de tan gran corazón [*se refiere aquí al emperador Galeno, que había ordenado este censo: 293-310 d.C.*] también tuvo piedad de ellos y encontró el modo de liberarlos de su miseria: hizo reunir a todos, cargarlos en una barca y echarlos al mar.

Lactancio, De mort. Pers. 23. (finales del siglo III .d.C)

Traducción por cortesía de Arturo Herrera.



Gestión del Conocimiento

DIFERENCIA ENTRE DATO, INFORMACIÓN Y CONOCIMIENTO

Antes de nada, debemos comprender en qué se diferencia el conocimiento de los datos y de la información. En una conversación informal, los tres términos suelen utilizarse indistintamente y esto puede llevar a una interpretación libre del concepto de conocimiento. Quizás la forma más sencilla de diferenciar los términos sea pensar que los datos están localizados en el mundo y el conocimiento está localizado en agentes de cualquier tipo, mientras que la información adopta un papel mediador entre ambos.

Un agente no equivale a un ser humano. Podría tratarse de un animal, una máquina o una organización constituida por otros agentes a su vez.

DATO, INFORMACIÓN Y CONOCIMIENTO SEGÚN DAVENPORT Y PRUSAK (1999)



Dato:

Un dato es un conjunto discreto, de factores objetivos sobre un hecho real. Dentro de un contexto empresarial, el concepto de dato es definido como un registro de transacciones. Un dato no dice nada sobre el porqué de las cosas, y por sí mismo tiene poca o ninguna relevancia o propósito.

Las organizaciones actuales normalmente almacenan datos mediante el uso de tecnologías. Desde un punto de vista cuantitativo, las empresas evalúan la gestión de los datos en términos de coste, velocidad y capacidad.

Todas las organizaciones necesitan datos y algunos sectores son totalmente dependientes de ellos. Bancos, compañías de seguros, agencias gubernamentales y la Seguridad Social son ejemplos obvios. En este tipo de organizaciones la buena gestión de los datos es esencial para su funcionamiento, ya que operan con millones de transacciones diarias. Pero en general, para la mayoría de las empresas tener muchos datos no siempre es bueno. Las organizaciones almacenan datos sin sentido. Realmente esta actitud no tiene sentido por dos razones. La primera es que demasiados datos hacen más complicado identificar aquellos que son relevantes. Segundo, y todavía más importante, es que los datos no tienen significado en sí mismos.

Los datos describen únicamente una parte de lo que pasa en la realidad y no proporcionan juicios de valor o interpretaciones, y por lo tanto no son orientativos para la acción. La toma de decisiones se basará en datos, pero estos nunca dirán lo que hacer. Los datos no dicen nada acerca de lo que es importante o no. A pesar de todo, los datos son



importantes para las organizaciones, ya que son la base para la creación de información.



Información:

Como han hecho muchos investigadores que han estudiado el concepto de información, lo describiremos como un mensaje, normalmente bajo la forma de un documento o algún tipo de comunicación audible o visible. Como cualquier mensaje, tiene un emisor y un receptor. La información es capaz de cambiar la forma en que el receptor percibe algo, es capaz de impactar sobre sus juicios de valor y comportamientos. Tiene que informar; son datos que marcan la diferencia. La palabra “informar” significa originalmente “dar forma a “ y la información es capaz de formar a la persona que la consigue, proporcionando ciertas diferencias en su interior o exterior. Por lo tanto, estrictamente hablando, es el receptor, y no el emisor, el que decide si el mensaje que ha recibido es realmente información, es decir, si realmente le informa. Un informe lleno de tablas inconexas, puede ser considerado información por el que lo escribe, pero a su vez puede ser juzgado como “ruido” por el que lo recibe.

La información se mueve entorno a las organizaciones a través de redes formales e informales. Las redes formales tienen una infraestructura visible y definida: cables, buzones de correo electrónico, direcciones,... Los mensajes que estas redes proporcionan incluyen e-mail, servicio de entrega de paquetes, y transmisiones a través de Internet. Las redes informales son invisibles. Se hacen a medida. Un ejemplo de este tipo de red es cuando alguien te manda una nota o una copia de un artículo con las siglas “FYI” (For Your Information).



A diferencia de los datos, la información tiene significado (relevancia y propósito). No sólo puede formar potencialmente al que la recibe, sino que esta organizada para algún propósito. Los datos se convierten en información cuando su creador les añade significado. Transformamos datos en información añadiéndoles valor en varios sentidos. Hay varios métodos:

- Contextualizando: sabemos para qué propósito se generaron los datos.
- Categorizando: conocemos las unidades de análisis de los componentes principales de los datos.
- Calculando: los datos pueden haber sido analizados matemática o estadísticamente.
- Corrigiendo: los errores se han eliminado de los datos.
- Condensando: los datos se han podido resumir de forma más concisa.

Los ordenadores nos pueden ayudar a añadir valor y transformar datos en información, pero es muy difícil que nos puedan ayudar a analizar el contexto de dicha información. Un problema muy común es confundir la información (o el conocimiento) con la tecnología que la soporta. Desde la televisión a Internet, es importante tener en cuenta que el medio no es el mensaje. Lo que se intercambia es más importante que el medio que se usa para hacerlo. Muchas veces se comenta que tener un teléfono no garantiza mantener conversaciones brillantes. En definitiva, que actualmente tengamos acceso a más tecnologías de la información no implica que hayamos mejorado nuestro nivel de información.





Conocimiento:

La mayoría de la gente tiene la sensación intuitiva de que el conocimiento es algo más amplio, más profundo y más rico que los datos y la información.

Vamos a intentar realizar una primera definición de conocimiento que nos permita comunicar que queremos decir cuando hablamos de conocimiento dentro de las organizaciones.

Para Davenport y Prusak (1999) el conocimiento es una mezcla de experiencia, valores, información y “saber hacer” que sirve como marco para la incorporación de nuevas experiencias e información, y es útil para la acción. Se origina y aplica en la mente de los conocedores. En las organizaciones con frecuencia no sólo se encuentra dentro de documentos o almacenes de datos, sino que también esta en rutinas organizativas, procesos, prácticas, y normas.

Lo que inmediatamente deja claro la definición es que ese conocimiento no es simple. Es una mezcla de varios elementos; es un flujo al mismo tiempo que tiene una estructura formalizada; es intuitivo y difícil de captar en palabras o de entender plenamente de forma lógica. El conocimiento existe dentro de las personas, como parte de la complejidad humana y de nuestra impredecibilidad. Aunque solemos pensar en activos definibles y concretos, los activos de conocimiento son mucho más difíciles de manejar. El conocimiento puede ser visto como un proceso (flujo) o como un stock.



El conocimiento se deriva de la información, así como la información se deriva de los datos. Para que la información se convierte en conocimiento, las personas debe hacer prácticamente todo el trabajo. Esta transformación se produce gracias a:

- Comparación.
- Consecuencias.
- Conexiones.
- Conversación.

Estas actividades de creación de conocimiento tienen lugar dentro y entre personas. Al igual que encontramos datos en registros, e información en mensajes, podemos obtenemos conocimiento de individuos, grupos de conocimiento, o incluso en rutinas organizativas.

Nota: Información facilitada por Juan Carrión



ANEXO 4

Operacionalización de Variables

SONIA INÉS BETANCUR LÓPEZ

Enfermera Docente Departamento de Salud Pública

Facultad de Ciencias para la Salud

Universidad de Caldas

Puesto que todo investigador durante el proceso de elaboración de un proyecto debe plantearse cuál o cuáles serán las variables o características del objeto de estudio contenidas en las hipótesis que deberá evaluar en la realidad, es decir, someter a “prueba empírica” a través de la medición; se pretende con este artículo describir y ejemplificar el proceso de operacionalización de una variable, para hacer más comprensible la terminología y los conceptos dispersos en la literatura de investigación. Una variable es operacionalizada con el fin de convertir un concepto abstracto en uno empírico, susceptible de ser medido a través de la aplicación de un instrumento. Dicho proceso tiene su importancia en la posibilidad que un investigador poco experimentado pueda tener la seguridad de no perderse o cometer errores que son frecuentes en un proceso investigativo, cuando no existe relación entre la variable y la forma en que se decidió medirla, perdiendo así **LA VALIDEZ** (grado en que la medición empírica representa la medición conceptual). La precisión para definir los términos tiene la ventaja de comunicar con exactitud los resultados.



Algunas variables son tan concretas, o de igual significado en el ámbito mundial, que no requieren operacionalización, por ejemplo: el sexo de los individuos, los colores del semáforo como señal de tránsito, la ubicación o estructura de órganos en el cuerpo humano, entre otros. La operacionalización de variables se hace necesario por el grado de dificultad que representa la definición de algunas de ellas dado que corresponden a conceptos abstractos tales como: autoridad, calidad de vida, educación, riesgos biológicos, bienestar del enfermo y otros, que pueden tomar diferentes significados según el tipo de estudio y el interés o disciplina del investigador que lo realiza. Este proceso está descrito en algunos textos, solamente con algunos de los aspectos aquí considerados, lo que quiere decir, que no siempre será absolutamente necesario realizar todo el ejercicio que se plantea y será decisión del investigador, acorde con sus habilidades, llevarlo a cabo parcial o totalmente.

No obstante, es importante conocer el significado y la utilidad de cada columna

en el esquema para decidir, en cada caso particular, cuáles de ellas se hacen imprescindibles.

Con fines didácticos se explica cada una de las columnas del cuadro que hacen parte del proceso de operacionalización de una variable.



| VARIABLE | TIPO DE VARIABLE | OPERACIONALIZACIÓN | CATEGORIZACIÓN O DIMENSIONES | DEFINICIÓN | INDICADOR | NIVEL DE MEDICIÓN | UNIDAD DE MEDIDA | INDICE | VALOR |
|----------|------------------|--------------------|------------------------------|------------|-----------|-------------------|------------------|--------|-------|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

1. VARIABLE

- Una variable es una característica que se va a medir.
- Es una propiedad, un atributo que puede darse o no en ciertos sujetos o fenómenos en estudio, así como también con mayor o menor grado de presencialidad en los mismos y por tanto con susceptibilidad de medición.
- Su misma palabra define que “debe admitir rangos de variación”.
- Es el conjunto de valores que constituyen una clasificación.
- Debe traducirse del nivel conceptual (abstracto) al nivel operativo (concreto), es decir que sea observable y medible.
- Se deriva de la unidad de análisis y están contenidas en las hipótesis y en el título del estudio.

2. TIPO DE VARIABLE

Hace referencia a conceptos clasificatorios de las variables que puede ser de distinto orden a saber:

- Según el nivel de medición: nominal, ordinal, de intervalo y de razón (se explican en el numeral 7).
- Según el tipo de estudio: en estudios de investigación donde se supone la determinación de una o más variables sobre otra, los estudios son de relación causa-efecto, y en ellos las variables son denominadas: independiente, que representa la causa eventual, dependiente o de criterio, que representa el efecto posible, e interviniente aquella que representa una tercera variable que actúa entre la independiente y la dependiente y que puede ayudar a una mejor comprensión de dicha



relación. Ejemplo: en un estudio donde se trata de probar la influencia de los medios de comunicación con un mayor nivel de instrucción de los individuos, se consideraría como variable dependiente (vd) el mayor nivel de instrucción, como variable independiente, la exposición a los medios de comunicación (vi) y sería una variable interviniente (vt) el interés particular de los individuos por ciertos programas de los medios de comunicación.

- Según el origen de la variable: activa, cuando el investigador la crea o la diseña y, atributiva o preexistente cuando ya está establecida o existe.
- Según el número de valores que representa: continua, representa valores de manera progresiva y admite fraccionamiento como la edad y, categórica o discreta cuando sólo toma algunos valores discretos o sea que no admite fraccionamiento tales como el género, la raza, el número de hijos o de embarazos; si la variable sólo toma dos valores como el sexo se denomina categórica dicotómica, pero si toma más de dos valores se denominará politómica.
- Según el control de la variable por parte del investigador: la variable que tiene efecto sobre la variable dependiente requiere que sea controlada por el investigador, por ejemplo, el número de cigarrillos que consume por día un fumador y su relación con la aparición prematura de la patología pulmonar, en este caso la variable se denomina controlable o controlada. Cuando en el diseño o en el análisis la variable no se considera, será una variable no controlada.



3. OPERACIONALIZACIÓN O DEFINICIÓN OPERACIONAL

- Explica como se define el concepto específicamente en el estudio planteado, que puede diferir de su definición etimológica.
- Equivale a hacer que la variable sea mensurable a través de la concreción de su significado, y está muy relacionada con una adecuada revisión de la literatura.
- Puede omitirse cuando la definición es obvia y compartida.

4. CATEGORIZACIÓN O DIMENSIONES

Cuando el concepto tiene varias dimensiones o clasificaciones o categorías, éstas deben especificarse en el estudio; tal es el caso de la variable “recursos», que puede hacer referencia a “recursos técnicos, financieros, ambientales, humanos entre otros».

5. DEFINICIÓN DE LAS CATEGORIAS O DIMENSIONES

Cada una de las dimensiones, categorías o clasificaciones debe ser definida conceptual y etimológicamente.

6. INDICADOR

- Es la señal que permite identificar las características de las variables.
- Se da con respecto a un punto de referencia.
- Son señales comparativas con respecto a contextos o a sí mismas.
- Su expresión matemática se nutre de la estadística, la epidemiología y la economía.
- Se expresa en razones, proporciones, tasas e índices.
- Permite hacer “medible” la variable.
- Son ejemplos de indicadores: indicadores económicos (la UVR, el dólar, la libra de café, el gramo de oro).



- Indicadores de pobreza (las migraciones, los desplazados, el desempleo, los asentamientos suburbanos).
- Indicadores de calidad de vida (tasa de natalidad, de mortalidad, de fecundidad, de esperanza de vida).
- Indicadores de desarrollo (el PIB: precio interno bruto, tasa de desempleo, la inflación, el IPC: índice de precios al consumidor). Otros indicadores pueden ser construidos por el investigador.

7. NIVEL DE MEDICIÓN

La medición de una variable se refiere a su posibilidad de cuantificación o cualificación, y éstas se clasifican según el nivel o capacidad en que permite ser medido el objeto en estudio. Según el tipo de operaciones matemáticas que se puedan realizar con los números asignados al medir la variable, se distinguen cuatro niveles de medición estadística, como son:

NOMINAL: este nivel sólo permite clasificar, es decir, la única relación existente entre los objetos a los cuales se les ha asignado un número es una relación de equivalencia. Por ejemplo, si en la variable sexo se ha asignado el numeral 1 para designar a los hombres y el 2, para referirse a las mujeres, quiere decir que todos los miembros a los que se les asigne el numeral 1 son hombres, o sea, tienen una condición equivalente. La relación de equivalencia es reflexiva ($a=a$), es simétrica (si $a=b$ entonces $b=a$) y es transitiva (si $a=b$ y $b=c$ entonces $a=c$), de acuerdo con estas propiedades las técnicas estadísticas posibles de usar con la escala nominal son la moda y el cálculo de frecuencias también se pueden usar medidas no paramétricas como el chi cuadrado y la expresión binomial; en cuanto a medidas de asociación se puede usar el coeficiente de contingencia, Es necesario recalcar que los



números asignados a las diferentes categorías de la variable cualitativa sirven para almacenamiento de datos, pero por ser de asignación arbitraria no indica que se trate de variables cuantitativas.

ORDINAL: permite clasificar además ordenar, es decir, establecer una secuencia lógica que mide la intensidad del atributo. Por ejemplo, al medir el grado de satisfacción frente a un servicio de salud, se pueden establecer escalas tales como: satisfacción plena, satisfacción media, poca satisfacción, o insatisfacción; esta escala difiere de la meramente nominal que permite establecer un orden o graduación entre las observaciones. Las técnicas estadísticas apropiadas para las mediciones ordinales son: la mediana para describir las tendencias centrales, los coeficientes de Spearman, de Kendall y Gamma, para correlaciones y pruebas no paramétricas como Wilcoxon, Kolmorov-Smirnov, entre otras para pruebas de hipótesis. Al igual que el nivel nominal, los números asignados sólo indican un orden o rango entre los objetos y en ningún momento indican relación numérica, tal como el ejemplo anterior si el grado de satisfacción plena se le asigna el número 4 y 2 al grado de poca satisfacción, no indica esto que quien marcó el número 4 esté el doble de satisfecho que quien marcó el número 2. La escala ordinal además de poseer las propiedades de la relación de equivalencia del nivel nominal posee también la relación mayor que, expresada en términos como más satisfecho, más estable, de mayor tamaño, de mayor preferencia, más peligroso, más útil, de mayor riesgo etc. Todas las escalas socio-económicas pertenecen al nivel ordinal de medición, ya que las distancias entre clases sociales o estratos económicos no son iguales, si lo fueran pertenecerían al nivel intervalar.



INTERVALAR o NUMÉRICA: permite clasificar y ordenar pero además los intervalos son iguales, o sea, que en este nivel de medición no solo es posible ordenar las escalas sino que es posible conocer las distancias o grados que separan unas de otras. La escala intervalar tiene las mismas propiedades formales de las escalas nominales y ordinales, es decir, las relaciones de equivalencia y de mayor que; además, se le agrega la propiedad de poder determinar la razón que existe entre dos intervalos, en este caso existe una distancia numéricamente igual entre los objetos 2 y 3 que entre los objetos 3 y 4, porque en ambos la razón equivale a 1. En una escala de este nivel el punto cero y la unidad de medición son arbitrarios, como en el caso de la temperatura en que el grado cero no implica ausencia de temperatura, sino que se designó el cero en forma arbitraria. Entre las operaciones matemáticas correspondientes a esta escala pertenecen pruebas de la estadística paramétrica tales como la media aritmética, la desviación estándar, la correlación de Pearson, la T de Student, el Chi cuadrado, entre otras.

DE RAZÓN O PROPORCIÓN: posee las propiedades anteriores como clasificar, ordenar; los intervalos son iguales y además, existe el cero absoluto o verdadero”, lo que quiere decir que si un objeto que se está midiendo tiene el valor cero, ese objeto no posee la propiedad o atributo que se está midiendo. Esta escala constituye el nivel más alto de medición y admite para su análisis estadístico todas las técnicas y pruebas de los niveles anteriores, pero además admite la media geométrica, el cálculo del coeficiente de variación y las pruebas que requieran del conocimiento del punto cero de la escala.



Tabla 1: Nivel de Medición de las Variables

| ESCALA | TIPO DE VARIABLE | PROPIEDADES MATEMATICAS | PRUEBA ESTADÍSTICA | TÉCNICA ESTADISTICA |
|---------------------------|-----------------------|--|------------------------------|--|
| Nominal | Cualitativa Discreta | - De equivalencia | No paramétrica | Moda, cálculo de frecuencias, chi cuadrado, expresión binomial, coeficiente de contingencia |
| Ordinal | Cualitativa Discreta | - De equivalencia Mayor que | No paramétrica | Las anteriores y se adiciona: la mediana (tendencia central). coeficientes de Spearman. kendall, gamma, percentiles. |
| Intervalo | Cuantitativa Continua | De equivalencia Mayor que - Razón entre dos intervalos calculable | No paramétrica y paramétrica | Las anteriores y se adiciona: media aritmética, desviación estándar, correlación de pearson, correlación múltiple |
| Razón o proporción | Cuantitativa Continua | De equivalencia - Mayor que - Razón entre dos intervalos calculable - Razón entre dos valores de la escala calculable | No paramétrica y paramétrica | Las anteriores y se adiciona: Media geométrica, coeficientes de variación y otras. |



8. UNIDAD DE MEDIDA

- Se refiere a la respuesta que se espera en la medición planeada.
- Puede ser cuantitativa: en Kilos, en metros, en litros, en porcentajes, en proporciones, en tasas.
- Puede ser cualitativa: en grados de satisfacción (mucho, regular, poco), en calificaciones (excelente, regular, insuficiente), en grado de acuerdo (si y no) o (muy de acuerdo, en acuerdo, en desacuerdo) etc.

9. ÍNDICE

Es la expresión del indicador por ejemplo:

- Índice ocupacional: porcentaje de camas ocupadas.
- Índice de desempleo: porcentaje de desempleados.
- Índice de transición demográfica: porcentaje de atraso o avance de una sección del país.

10. VALOR

- Es el resultado o número de resultados posibles que se obtiene de una variable.
- Cuando una variable puede medirse a través de varios indicadores, algunos de ellos pueden tener mayor valor que otros y por tanto se hace necesario explicitarlo. Por ejemplo: la variable “calidad docente” puede medirse a través de: la hoja de vida del docente, el grado de capacitación, o sea. el número de títulos académicos, un examen de conocimientos o una prueba pedagógica: pero es posible que se le asigne un mayor valor porcentual a la hoja de vida y al grado de capacitación que a las dos restantes.
- Una columna 11 correspondería al ítem o pregunta que daría respuesta o mediría la variable. Si se quisiera agregar más, podría considerarse en la columna 12, la fuente de recolección de información

que haría referencia al individuo, sitio o instrumento de donde se tomaría el dato (historia clínica del Hospital de Caldas. usuario del servicio de salud, registros del Dane) y en una la columna 13, la técnica de recolección del dato (observación, entrevista, grupo focal).

Tabla 2: Ejemplo de Operacionalización de Variables

| 1. VARIABLE | 2. TIPO DE VARIABLE | 3. OPERACIONALIZACIÓN | 4. CATEGORÍAS o DIMENSIONES | 5. DEFINICIÓN |
|--|----------------------|---|---|--|
| Aprovechamiento de los recursos en un servicio hospitalario | Cuantitativa | Forma como se utiliza cada uno de los recursos institucionales para brindar al usuario un servicio de calidad | Aprovechamiento de los recursos físicos Aprovechamiento de los recursos humanos Aprovechamiento de los recursos financieros | 1. Uso del número de camas disponibles para atender el número de pacientes solicitantes en mimes (oferta/demanda) |
| 6. INDICADOR | 7. Nivel de medición | 8. Unidad de medida | 9. Índice | 10. Valor |
| Porcentaje ocupacional de camas <input type="checkbox"/> Promedio de estancia hospitalaria <input type="checkbox"/> Giro-cama= #de egresos/#de camas | De razón | % • # de días • # decimal | Índice ocupacional <input type="checkbox"/> Índice de hospitalización Índice de utilización | Se pueden tomar los 3 valores con igual nivel de importancia o tomar solo 1 valor. De esta manera en el Instrumento de recolección de la información, la pregunta hará referencia a uno o más de los indicadores seleccionados |



BIBLIOGRAFÍA

PARDO DE V., Graciela; CEDEÑO C., Marlene. Investigación en salud. Factores sociales. Bogotá: Interamericana. 1997.

PINEDA, Ej3. DE ALVARADO, E.L. y DE CANALES, F.H. Metodología de la investigación en salud. Washington:OMS, OPS. 2 ed. 1994.

POLIT Y HUNGLER. Investigación científica en ciencias de la salud. 4 ed. Barcelona: Interamericana

TAMAYO y TAMAYO, M. El proyecto de investigación. Serie: aprender a investigar. Bogotá: ICFES-ICESI. Módulo 2.

TOVAR, María Clara. Variables y su medición. Módulos de investigación. Universidad del Valle.

VARKEVISSER, Corlien y otros. Diseño y realización de proyectos de investigación sobre sistemas de salud. Serie de capacitación en investigación en sistemas de salud. Vol.2, parte 1. Módulo 8, s.a. 1995.