Material de Autoformación e Innovación Docente

Estadística

Primero y Segundo Año de Bachillerato

11

Viceministerio de Ciencia y Tecnología

Ministerio de Educación

Viceministerio de Ciencia y Tecnología

Programa Cerrando la Brecha del Conocimiento

Subprograma Hacia la CYMA

Material de Autoformación e Innovación Docente

Para 1° y 2° año de Bachillerato

Versión Preliminar para Plan Piloto

Ministerio de Educación

Franzi Hasbún Barake
Secretario de Asuntos Estratégicos de la Presidencia de la República
Ministro de Educación Ad-honorem

Erlinda Hándal Vega
Viceministra de Ciencia y Tecnología

Héctor Jesús Samour Canán
Viceministro de Educación

William Ernesto Figueroa
Director Nacional de Ciencia y Tecnología

Xiomara Guadalupe Rodríguez Amaya
Gerente de Educación en Ciencia, Tecnología e Innovación

Oscar de Jesús Águila Chávez
Jefe de Educación Media en CTI (Coordinador de Matemática)

Carlos Ernesto Miranda Oliva
Jefe de Educación Básica en CTI (Coordinador de Ciencias Naturales)

Alba Idalia Córdova Cuellar
Autora

Oscar de Jesús Aguila Chávez
Revisión técnica y diagramación

Jorge Vargas Méndez
Revisión de texto

Sergio Armando Márquez
Diseño de Portada

Primera edición (Versión Preliminar para Plan Piloto).

Edificios A4, segundo nivel, Plan Maestro, Centro de Gobierno, Alameda Juan Pablo II y Calle Guadalupe, San Salvador, El Salvador, América Central. Teléfonos: + (503) 2510-4217, + (503) 2510-4218, + (503) 2510-4219, Correo electrónico: gecti@mined.gob.sv

Estimadas y estimados docentes:

El Plan Social Educativo “Vamos a la Escuela” 2009-2014 nos plantea el reto histórico de formar ciudadanas y ciudadanos salvadoreños con juicio crítico, capacidad reflexiva e investigativa, con habilidades y destrezas para la construcción colectiva de nuevos conocimientos, que les permitan transformar la realidad social y valorar y proteger el medio ambiente.

Nuestros niños, niñas y jóvenes desempeñarán en el futuro un rol importante en el desarrollo científico, tecnológico y económico del país; para ello requieren de una formación sólida e innovadora en todas las áreas curriculares, pero sobre todo en Matemática y en Ciencias Naturales; este proceso de formación debe iniciarse desde el Nivel de Parvularia, intensificándose en la Educación Básica y especializándose en el Nivel Medio y Superior. En la actualidad, es innegable que el impulso y desarrollo de la Ciencia y la Tecnología son dos aspectos determinantes en el desarrollo económico, social y humano de un país.

Para responder a este contexto, en el Viceministerio de Ciencia y Tecnología se han diseñado materiales de autoformación e innovación docente para las disciplinas de Matemática y Ciencias Naturales, para bachillerato. El propósito de éstos materiales es orientar al cuerpo docente para fundamentar mejor su práctica profesional, tanto en dominio de contenidos, como también en la implementación de metodologías y técnicas que permitan la innovación pedagógica, la indagación científica-escolar y sobre todo una construcción social del conocimiento, bajo el enfoque de Ciencia, Tecnología e Innovación (CTI), en aras de mejorar la calidad de la educación.

Los materiales, son para el equipo docente, para su profesionalización y autoformación permanente que le permita un buen dominio de las disciplinas que enseña. Los contenidos que se desarrollan en estos cuadernillos, han sido cuidadosamente seleccionados por su importancia pedagógica y por su riqueza científica. Es por eso que para el estudio de las lecciones incluidas en estos materiales, se requiere rigurosidad, creatividad, deseo y compromiso de innovar la práctica docente en el aula. Con el estudio de las lecciones (de manera individual o en equipo de docentes), se pueden derivar diversas sesiones de trabajo con el estudiantado para orientar el conocimiento de los temas clave o “pivotes” que son el fundamento de la alfabetización científica en Matemática y Ciencias Naturales.

La enseñanza de las Ciencias Naturales y la Matemática debe despertar la creatividad, siendo divertida, provocadora del pensamiento crítico y divergente, debe ilusionar a los niños y niñas con la posibilidad de conocer y comprender mejor la naturaleza y sus leyes. La indagación en Ciencias Naturales y la resolución de problemas en Matemática son enfoques que promueven la diversidad de secuencias didácticas y la realización de actividades de diferentes niveles cognitivos.

Esperamos que estos Materiales de Autoformación e Innovación Docente establezcan nuevos caminos para la enseñanza y aprendizaje de las Ciencias Naturales y Matemática, fundamentando de una mejor manera nuestra práctica docente. También esperamos que el contenido de estos materiales nos rete a aspirar a mejores niveles de rendimiento académico y de calidad educativa, en la comunidad educativa, como en nuestro país en general.

Apreciable docente, ponemos en sus manos estos Materiales de Autoformación e Innovación Docente, porque sabemos que está en sus manos la posibilidad y la enorme responsabilidad de mejorar el desempeño académico estudiantil, a través del desarrollo curricular en general, y particularmente de las Ciencias Naturales y Matemática.

Lic. Franzi Hasbún Barake
Secretario de Asuntos Estratégicos de la Presidencia de la República
Ministro de Educación Ad-honorem

Dr. Héctor Jesús Samour Canán
Viceministro de Educación

Dra. Erlinda Hándal Vega
Viceministra de Ciencia y Tecnología

Índice

¿Por qué material de autoformación e innovación docente?

Presentación

El Viceministerio de Ciencia y Tecnología a través de la Gerencia de Educación en Ciencia, Tecnología e Innovación (GECTI) y el sub programa "Hacia la CYMA" que se desarrolla durante el quinquenio 2009-2014, ejecuta el Proyecto de Enriquecimiento Curricular, en el área de Ciencias Naturales y Matemática, el cual tiene entre sus acciones la elaboración y entrega de materiales de autoformación e innovación docente para docentes.

Este material de autoformación e innovación para docentes, tiene como propósito fortalecer el desarrollo curricular de Matemática y Ciencias desde parvularia hasta educación media, introduciendo el enfoque Ciencia Tecnología e Innovación (CTI) como parte inherente y relevante del proceso de formación científica.

Con este propósito se han elaborado lecciones con temas pivotes considerados fundamentales para el conocimiento de los docentes, para obtener una fundamentación científica que permita fortalecer las capacidades de investigación, innovación docente, y de esta manera mejorar la calidad de la educación. Se busca que mediante la formación científica se logren cambios en las condiciones sociales y económicas que permitan a la población salvadoreña alcanzar una vida digna.

El enriquecimiento y profundización de los temas propuestos, tiene la posibilidad de ser plataforma de construcción de conocimiento bajo el enfoque de resolución de problemas, metodología mediante la cual se desarrollan competencias matemáticas necesarias, que debe tener cada estudiante para alcanzar sus propósitos de formación intelectual, como son: saber argumentar, cuantificar, analizar críticamente la información, representar y comunicar, resolver y enfrentarse a problemas, usar técnicas e instrumentos matemáticos y modelizar e integrar los conocimientos adquiridos.

La resolución de problemas en Matemática

Desde asegurar la subsistencia cotidiana, hasta abordar los más complejos desafíos derivados desde la Ciencia y la Tecnología, sin excepción todos resolvemos problemas. Lo vital de la actividad de resolución de problemas es evidente; en definitiva, todo el progreso científico y tecnológico[1], el bienestar y hasta la supervivencia de la especie humana dependen de esta habilidad. No debemos extrañarnos de que la misma se haya convertido en un nuevo objeto de estudio, atrayendo por igual la atención de profesionales de la psicología, ingeniería, física, química, biología, matemática, etc.

En Matemática debemos hacer algunos cuestionamientos que son fundamentales en el proceso metodológico de la resolución de problemas.

¿Cuál es la diferencia entre ejercicio y problema en Matemática? ¿Cuándo está el estudiantado resolviendo un ejercicio y cuándo un problema? ¿Cuál es el papel de un profesor en la enseñanza de la resolución de problemas?

Al analizar un ejercicio se puede deducir si se sabe resolver o no; Comúnmente se aplica un algoritmo elemental o complejo que los niños y niñas pueden conocer o ignorar, pero una vez encontrado este algoritmo, se aplica y se obtiene la solución.

Justamente, la exagerada proliferación de ejercicios en la clase de Matemática ha desarrollado y penetrado en el estudiantado como un síndrome generalizado. En cuanto se les plantea una tarea a realizar, tras una simple reflexión, tratan de obtener una solución muchas veces elemental, sin la apelación a conocimientos diversos.

En un problema no es siempre evidente el camino a seguir. Incluso puede haber muchos. Hay que apelar a conocimientos, no siempre de Matemática, relacionar saberes procedentes de campos diferentes, poner a punto nuevas relaciones. El papel de cada docente es proporcionar a la niñez la posibilidad de aprender hábitos de pensamiento adecuados para la resolución de problemas matemáticos y no matemáticos.

¿De qué les puede servir hacer un hueco en su mente en que quepan unos cuantos algoritmos, teoremas y propiedades relativas a entes con poco significado si luego van a dejarlos allí herméticamente acumulados? A la resolución de problemas se le ha llamado, con razón, el corazón de las matemáticas, pues ahí es donde se puede adquirir el verdadero sabor que ha traído y atrae a académicos de todas las épocas. Del enfrentamiento con problemas adecuados es de donde pueden resultar motivaciones, actitudes, hábitos, ideas y competencias para el desarrollo de herramientas, en una palabra, la vida propia de la Matemática[2].

---

[1] José Heber Nieto Said; Resolución de Problemas Matemáticos 2004.
[2] Miguel de Guzmán Ozámiz, (1936 - 2004) matemático español.

Obviamente la resolución de problemas tiene una clásica y bien conocida fase de formulación elaborada por el matemático húngaro George Polya[3] en 1945. Fase que consisten en comprender el problema, trazar un plan para resolverlo, poner en práctica el plan y comprobar el resultado.

Por supuesto hay que pensar que no sólo basta con conocer las fases y técnicas de resolución de problemas. Se pueden conocer muchos métodos pero no siempre cuál aplicar en un caso concreto.

Justamente hay que enseñar también a las niñas y niños, a utilizar las estrategias que conocen, con lo que nos encontramos en un nivel metacognitivo. Es ahí donde se sitúa la diferencia entre quienes resuelven problemas y los demás, entendiendo que este nivel es la capacidad que tienen de autoregular su propio aprendizaje, es decir, de planificar qué estrategias se han de utilizar en cada situación, aplicarlas, controlar el proceso, evaluarlo para detectar posibles fallos, y como consecuencia transferir todo ello a una nueva actuación[4].

Hay que tener presente que resulta difícil motivar. Sólo con proponer ejercicios no se puede conseguir que las niñas y niños sean capaces de investigar y descubrir nuevos conocimientos y relaciones entre las ciencias. Se recomienda establecer problemas en los que no sepan qué hacer en un primer intento, con esto conseguiremos atraer su atención y motivación, para que se impliquen en el proceso de resolución. Otro aspecto no menos importante a tener en cuenta es la manipulación de materiales para resolver problemas. Hemos de ser capaces de que las niñas y los niños visualicen el problema, utilizando materiales concretos, materiales que manipulen, pues la manipulación es un paso previo e imprescindible para la abstracción en las ciencias en general.

Descripción de la estructura de los cuadernillos

El cuadernillo de estadística de bachillerato es un material de apoyo para el docente, considerado Material de Autoformación e Innovación Docente que permite reorientar lecciones que se desarrollan desde distintos materiales a un entorno participativo y de investigación fundamentado en la resolución de problemas, donde el estudio de la Física, Química y Biología en conjunto con la Matemática fortalecen competencias conceptuales, procedimentales y actitudinales de la juventud salvadoreña. Se proponen diez temas que llamamos contenidos pivotes, que por su importancia en la formación de competencias matemáticas, forman parte del enriquecimiento curricular, profundizando tanto en la explicación de los contenidos, como haciendo propuestas de abordaje metodológico fundamentalmente en la resolución de problemas, con el propósito de que se puedan emular en el aula tanto maestros como alumnos puedan desarrollar habilidades intelectuales propias del pensamiento y del quehacer científico.

---

[3] George Pólya (1887-1985), matemático Húngaro, How to solve it, Pricenton University Press.
[4] Allan Schoenfeld (1985). Mathematical Problem Solving. New York: Academic Pres.

Las lecciones se estructuran normalmente en once partes, las cuales se detallan a continuación:

a. **Título.** Condensa la idea central de la lección. Se presenta como una idea clara y precisa del contenido.
b. **Introducción.** Presenta todos aquellos puntos relevantes que se tratarán en la lección, haciendo énfasis en las características (generalidades, importancia, usos, etc.) que se desarrollan.
c. **Objetivos de la lección.** Son las metas que se persiguen con la lección, es decir, lo que se pretende alcanzar con el desarrollo de la lección.
d. **Competencia matemática.** Son las habilidades y destrezas que el estudiante puede adquirir al finalizar la lección.
e. **Tiempo.** Es la duración aproximada para el desarrollo de la lección. El tiempo puede variar según la planificación didáctica de la clase.
f. **Contenido de la lección.** En esta parte se puede ver los elementos de contenido que componente la lección.
g. **Vocabulario.** En este apartado se encuentra un pequeño glosario de conceptos básicos del contenido de la lección. La elección de estos conceptos se ha realizado con la intención de que sirva de ayuda en el momento de leer el marco teórico de la lección.
h. **Presaberes.** Esta sección aborda los conceptos, proposiciones y toda la información relevante que se establece como marco de referencia de los tópicos a estudiar. La información se respalda en principios, leyes, clasificaciones, características, propiedades, etc. Se acompaña de ilustraciones, esquemas, modelos y otros con la intención de que el contenido quede lo más claro posible.
i. **Aplicando lo aprendido.** Las actividades de aplicación serán para contribuir al fortalecimiento del marco teórico, asimilando los conceptos de una manera práctica. Las actividades están encaminadas a forjar ideas que construyan la comprensión, el análisis y la resolución de problemas como eje fundamental; éstas se refieren a la ejecución de prácticas significativas de aprendizaje que van desde lo simple a lo complejo, desarrollándose con distintas alternativas de abordaje plasmando al menos tres alternativas de solución comentadas por el docente. Estas contienen estrategias de solución encaminadas a fortalecer la capacidad de razonamiento lógico.
j. **Diagrama de contenido.** En este se presenta la estructura de la lección que permite dar una idea general acerca del contenido abordado y su secuencia lógica
k. **Bibliografía**. Acá podemos encontrar los títulos de la literatura utilizada para enriquecer la lección.

Lección 1 y 2 **Primer año de Bachillerato** **Unidad III** **Tiempo: 16 horas clase**

Análisis tabular y Gráfico

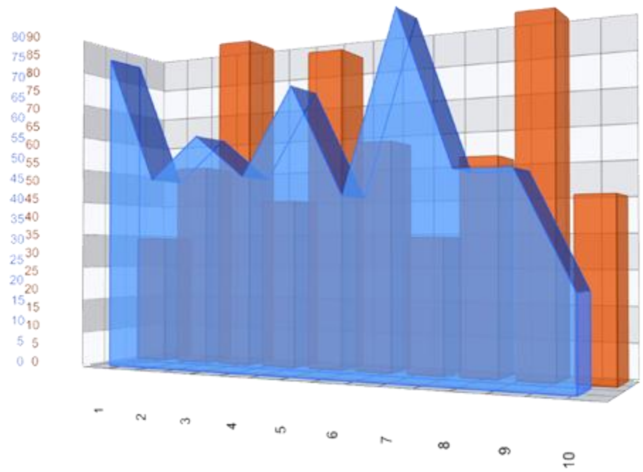


Figura 1. El Salvador. Matrimonios, por sexo, según grupos de edad: 2006.
Fuente:DIGESTYC

Introducción del Tema

Al registrar los resultados de un estudio observacional o experimental, se obtiene un número de observaciones que puede ser muy grande y su simple listado es de poca relevancia en el sentido interpretativo.

La información es importante para la toma de decisiones en muchos problemas. Para esto necesitamos un procesamiento adecuado de los datos, para que nos arrojen conclusiones certeras. En caso contrario, si no se aplica un buen procesamiento, es posible que en base a los resultados se tome una mala decisión.

El poder de una gráfica

Con ventas anuales cercanas a los $10,000 millones, y con alrededor de 50 millones de usuarios, el fármaco Lipitor de Pfizar se ha convertido en el medicamento de prescripción más redituable y más utilizado de la historia. Al inicio de su desarrollo, Lipitor se comparó con otros fármacos (Zocor, Mevacor, Lescol y Pravachol), en un proceso que implicó ensayos controlados. El resumen del informe incluyó una **gráfica** que mostraba una curva del Lipitor con un incremento más pronunciado que las curvas de los otros medicamentos, lo cual demostraba visualmente que Lipitor era más efectivo para reducir el colesterol que los otros fármacos.

Objetivos

- Comprender los conceptos básicos y terminología de la Estadística.
- Elaborar tablas de distribución de frecuencias, distinguiendo las más adecuadas para cada tipo de variable.
- Realizar la representación gráfica adecuada a cada caso de la información obtenida de una muestra o población e interpretarla.

Importancia

Cuando se realizan estudios estadísticos las herramientas que se utilizan para la organización y representación de la información son las tablas de distribuciones de frecuencias y los gráficos estadísticos. Las tablas permiten organizar y resumir la información de un conjunto grande de datos para tener una mejor compresión de ellos y permiten tener una base para la construcción de gráficos.

Los gráficos tienen doble utilidad; ya que pueden servir no sólo como sustituto a las tablas, sino que también constituyen por sí mismos una poderosa herramienta para el análisis de los datos, siendo en ocasiones el medio más efectivo no sólo para describir y resumir la información, sino también para analizarla. El propósito de un gráfico es ayudar a la comprensión y comunicación de la evidencia aportada por los datos respecto a una hipótesis en estudio. Un gráfico científico debe servir por tanto para representar la realidad.

Par Kelly, que en ese entonces era un alto ejecutivo de marketing de Pfizer, declaró: "Nunca olvidare cuando vi esa gráfica [....]". En ese momento pensé ¡caray!. Ahora sé de qué se trata ¡Podemos comunicar esto! La Food and Drug Administración de Estados Unidos aprobó el Lipitor y permitió a Pfizer incluir la gráfica con cada prescripción. El personal de ventas de la empresa también distribuyó la gráfica entre los médicos.

El acierto "una imagen vale más que mil palabras" se puede aplicar al ámbito de la estadística descriptiva diciendo que "un gráfico bien elaborado vale más que mil tablas de frecuencias".

Competencias a reforzar.

Organiza, gráfica e interpreta la información obtenida de una muestra o población.

Presaberes

- Conocimiento de las operaciones básicas.
- Concepto de ejes cartesianos.
- Calculo de valores porcentuales.
- Manejo de calculadora.

DEFINICIONES BÁSICAS

Cuando se hacen estudios se reúnen datos de una pequeña parte de un grupo más grande, para aprender o investigar algo acerca de este último. Una meta común e importante de las investigaciones estadísticas es aprender de un grupo grande examinando los datos de algunos miembros. Es en dicho contexto, que los términos de población y muestra adquieren importancia.

Una de las razones por las cuales se decide tomar una muestra en lugar de la población entera es para reducir costos y/o tiempo, pero además puede realizarse cuando es sumamente difícil accesar a ciertos elementos de la población.

La ***población*** es el conjunto total o completo de todos los individuos, objetos o medidas que poseen algunas características comunes observables en un lugar y un momento determinado que se desean estudiar. El conjunto es completo porque incluye a todos los elementos a estudiar; mientras que la ***muestra*** es un subconjunto de elementos seleccionados fielmente representativos de la población.

La cantidad de elementos de la población se simboliza con la letra N y la cantidad de elementos de la muestra con la letra n.

Las ventajas de estudiar una población a partir de sus muestras son principalmente:

Coste reducido:

Si los datos que buscamos los podemos obtener a partir de una pequeña parte del total de la población, los gastos de recogida y tratamiento de los datos serán menores. Por ejemplo, cuando se realizan encuestas electorales, es más barato preguntar a 4,000 personas su intención de voto, que a 30.000.000.

Mayor rapidez: Estamos acostumbrados a ver cómo con los resultados del escrutinio de las primeras mesas electorales, se obtiene una aproximación bastante buena del resultado final de unas elecciones, muchas horas antes de que el recuento final de votos haya finalizado.

Más posibilidades:

Para hacer cierto tipo de estudios, por ejemplo el de duración de cierto tipo de foco, no es posible en la práctica destruirlos todos para conocer su vida media, ya que no quedaría nada que vender. Es mejor destruir sólo una pequeña parte de ellas y sacar conclusiones sobre las demás.

Existen múltiples técnicas estadísticas para extraer una muestra de la población, depende del problema se decide elegir una de ellas, ya que ésta servirá para hacer inferencias (generalizaciones) sobre toda la población. Cuando calculamos una medida numérica sobre la población completa le llamamos ***parámetro*** mientras que si calculamos una medida numérica sobre la muestra le llamamos ***estadístico***.

Ejemplo 1: Si se considera como población a todos los alumnos de bachillerato de un Instituto Nacional, la edad promedio de ellos, la proporción de estudiantes del sexo masculino, el gasto medio por mes de todos los estudiantes, son valores numéricos que los describen a todos ellos y se les denomina parámetros; supongamos que se toma una muestra representativa de la población descrita anteriormente (75 alumnos) y se calculan los mismos valores numéricos para estos 75 alumnos; a dichos valores se les denomina estadísticos.

Para diferenciar los parámetros de los estadísticos, se utilizan algunas notaciones como las siguientes:

Tabla 1: Diferencia entre parámetros y estadísticos

Parámetros (Población)	Estadísticos (Muestra)
Media: μ	Media: $\overline{X}$
Varianza: σ^2	Varianza: s^2
Desviación típica: σ	Desviación típica: s
Proporción: P (porcentaje)	Proporción: p (porcentaje)

Si un estadístico se usa para deducir un parámetro también se le suele llamar **estimador**.

Cuando se realiza una investigación estadística se recopilan datos, y estos datos se recogen o guardan en lo que se conoce como variables. Las ***variables*** son las características de interés que posee cada uno de los elementos individuales de una población o muestra.

Estas variables se organizan de forma ordenada y se almacenan en archivos; ya que posteriormente será posible operar, aplicar transformaciones y los análisis estadísticos de interés.

Las variables pueden contener datos numéricos o datos categóricos no cuantificables numéricamente.

Las variables que contienen datos numéricos se conocen como ***Variables Cuantitativas***; y son las que se refieren a cantidades y se registran o expresan en forma numérica, además las operaciones aritméticas, como sumar y obtener promedios, si son significativas para este tipo de variable.

Por ejemplo: la altura o peso de una persona, número de hijos de una familia, el salario que cobran los empleados, el número de artículos producidos en una semana, distancia recorrida por un automóvil, la nota obtenida en la asignatura de matemática etc.

Las variables cuantitativas pueden ser: Discretas o Continuas.

Las ***variables cuantitativas Discretas:*** son las que toman una cantidad finita numerable de valores aislados, es decir, entre cada dos valores consecutivos no se puede intercalar ningún otro valor de la variable. Por ejemplo, el número de empleados de una empresa, el número de artículos producidos, el número de empresas de la competencia, etc.

Las ***variables cuantitativas Continuas:*** son las que entre dos valores consecutivos puede tomar infinito número de valores; es decir, entre uno y otro valor de la variable existen infinitas posibilidades intermedias. Por ejemplo, el peso, la temperatura, el tiempo, el salario, la fuerza física, la longitud y el peso de un objeto, etc.

Las variables que contienen datos categóricos no cuantificables se conocen como ***Variables Cualitativas***; y son aquellas que no permiten construir una serie numérica definida; los atributos o características que toman son distintas modalidades observadas cualitativamente. Por ejemplo, la profesión, el estado civil, sexo, religión, color de ojos, color de cabello, etc. En el caso de la variable Estado Civil toma las modalidades de soltero, casado, divorciado, viudo.

Las variables cualitativas pueden ser: Nominales u Ordinales.

Las ***variables cualitativas Nominales:*** son aquellas que describen las características directamente por su contenido y solamente indican diferencias en las modalidades por sus clases; pero sin ningún tipo de ordenamiento.

Las ***variables cualitativas Ordinales:*** son aquellas que describen sus modalidades por el orden que ocupan, desde la primera a la última. Por ejemplo, clase social (baja-media-alta), nivel educativo (primaria-secundaria-bachillerato-universitario), tamaños de los vehículos (pequeño-mediano-grande), etc.

TABLA DE DISTRIBUCIÓN DE FRECUENCIAS.

Las distribuciones de frecuencias son la herramienta más sencilla y más utilizadas y eficaz cuando estamos rodeados de grandes cantidades de datos, que no nos dicen nada si no hacemos más que enumerarlos. Al organizar estos datos en una distribución de frecuencia ya nos proporcionan diversas ideas.

El fin principal de la organización de los datos en una distribución de frecuencias es, usualmente uno de los siguientes.

- Resumir conjuntos grandes de datos.
- Lograr cierta comprensión sobre la naturaleza de los datos.
- Tener una base para la construcción de gráficas importantes.
- Dejar bien visible la distribución de la variable estudiada e identificar su forma.

- Analizar, controlar y mostrar las capacidades de los procesos que derivan sus datos, tanto cualitativa como cuantitativamente.
- Ayuda a determinar el promedio, la desviación estándar, los coeficientes de asimetría y curtosis, así como otras medidas características de una distribución.
- Probar a qué tipo de distribución matemática se puede acoplar estadísticamente la distribución empírica de los datos relativos a la variable estudiada.

La distribución de frecuencia es una disposición tabular de datos estadísticos, ordenados ascendente o descendentemente, con la frecuencia de cada dato. Las distribuciones de frecuencias pueden ser para datos no agrupados y para datos agrupados o de intervalos de clase.

DISTRIBUCIÓN DE FRECUENCIA NO AGRUPADAS

Es aquella distribución que indica las frecuencias con que aparecen los datos estadísticos, desde el menor de ellos hasta el mayor de ese conjunto sin que se haya hecho ninguna modificación al tamaño de las unidades originales. En estas distribuciones los valores de cada variable han sido solamente reagrupados, siguiendo un orden lógico con sus respectivas frecuencias.

Este tipo de distribución es utilizada cuando los datos provienen de una variable cuantitativa (con muchos valores repetidos) o variable cualitativa; ya que se puede disponer a cada valor o categoría de la variable con su respectiva frecuencia.

Dada una variable X con valores $x_1, x_2, \ldots, x_n$; aparece una serie de conceptos generales que se definen a continuación.

Frecuencia absoluta (n_i): Se denomina frecuencia absoluta del valor x_i de la variable X, al número de veces que se repite x_i en la serie de datos.

Frecuencia relativa (fi): Se denomina frecuencia relativa del valor x_i de la variable X, a la proporción de observaciones de la variable que toman el valor x_i; y se define por el cociente de la frecuencia absoluta del valor x_i y el número total de valores de la variable (n). Es decir $f_i = \frac{n_i}{n}$

Frecuencia absoluta acumulada (N_i): Se denomina frecuencia absoluta acumulada del valor x_i a la suma de las frecuencias absolutas de los valores de la variable X anteriores o iguales a x_i. Y se define como: $N_i = \sum_{k=1}^{i} n_k$ i ≤k ó $N_i = N_{i-1} + n_i$

Frecuencia relativa acumulada (Fi): Se denomina frecuencia relativa acumulada del valor x_i de la variable X, a la proporción de observaciones de la variable que toman valores, menores o iguales que x_i. Y se define como el cociente de la frecuencia absoluta acumulada y el número total de valores de la variable (n). Es decir: $F_i = \frac{N_i}{n}$.

PROPIEDADES

Frecuencias Absolutas

- La suma de las frecuencias absolutas es igual al número total de valores. Es decir:

$$\sum_{i=1}^{k} n_i = n$$

- Las frecuencias absolutas son positivas y menores o iguales a n. Es decir: $0 \le n_i \le n$

Frecuencias Relativas

- La suma de todas las frecuencias relativas es igual a 1. Es decir: $\sum_{i=1}^{k} f_i = 1$
- Las frecuencias relativas son positivas y menores o iguales a 1. Es decir: $0 \le n_i \le 1$

Frecuencias Absolutas Acumuladas

- El valor de la última frecuencia absoluta acumulada coincide con el total de los valores de la variable (n). Es decir: $N_i = n_1 + n_2 + \ldots + n_i = \sum_{k=1}^{i} n_k = N_{i-1} + n_i = n$

Frecuencias Relativas Acumuladas

- El valor de la última frecuencia relativa acumulada es igual a la unidad. Es decir:

$$F_i = f_1 + f_2 + \ldots + f_i = \sum_{k=1}^{i} f_k = F_{i-1} + f_i = 1$$

La distribución de frecuencias de una variable suele presentarse ordenadamente mediante la tabla de frecuencias siguiente:

Tabla 2: Distribución de Frecuencias no Agrupada

Valores de la Variable x_i	Frecuencias Absolutas n_i	Frecuencias Relativas $f_i = \frac{n_i}{n}$	Frecuencias Absolutas Acumuladas N_i	Frecuencias Relativas Acumuladas $F_i = \frac{N_i}{n}$
x_1	n_1	f_1	N_1	$F_1=f_1$
x_2	n_2	f_2	N_2	$F_2=F_1+f_2$
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
x_i	n_i	f_i	N_i	$F_I=F_{i-1}+F_i$
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
x_k	n_k	f_k	$N_k=n$	$F_k=1$
	$\sum_{i=1}^{k} n_i = n$	$\sum_{i=1}^{k} f_i = 1$		

PASOS PARA ELABORAR LA TABLA DE FRECUENCIAS NO AGRUPADAS

1. Ordenar los datos de menor a mayor; si hay datos repetidos se deben considerar tantas veces como aparezcan.
2. Formar la tabla con sus respectivos encabezados en las columnas.
3. Observar que valores se encuentran en los datos y escribir en la primera columna los valores sin repetirse de menor a mayor.
4. Contar cuantas veces se repite cada valor en los datos y escribir esa cantidad en la columna dos.
5. Calcular las demás columnas por medio de las fórmulas.

Ejemplo 2: A los alumnos de primer año de bachillerato general del instituto nacional de San Bartolo se les pregunto: ¿cuál es el número de miembros de su familia que trabajan?. Los resultados obtenidos de esta pregunta se presentan a continuación: 2 1 2 2 1 2 4 2 1 1 2 3 2 1 1 1 3 4 2 2 2 2 1 2 1 1 1 3 2 2 3 2 3 1 2 4 2 1 4 1 1 3 4 3 2 2 2 1 3 3

a) Identifique la variable en estudio y su tipo.
b) Organice los datos en una tabla de distribución de frecuencia.
c) ¿Cuántos alumnos tienen tres miembros de su familia que trabajan?
d) ¿Qué porcentaje de alumnos tienen dos miembros de su familia que trabajan?
e) ¿Qué porcentaje de alumnos tienen tres o menos miembros de su familia que trabajan?
f) ¿Cuántos alumnos tienen menos de tres miembros de su familia que trabajan?

Solución a)

Variable en estudio: Número de miembros que trabaja.

Tipo: Cuantitativa discreta, porque toma sólo valores enteros.

Solución b)

Elaboración de tabla de distribución de frecuencia.

- Ordenando los datos de menor a mayor: 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 4 4 4 4 4
- Tamaño de la población: 50 (n=50).
- Los valores que toma la variable son: 1, 2, 3 y 4.
- Elaborar la estructura de la tabla de distribución de frecuencia.

Valores de la Variable x_i	Frecuencias Absolutas n_i	Frecuencias Relativas f_i	Frecuencias Absolutas Acumuladas N_i	Frecuencias Relativas Acumuladas F_i
1	16	0.32	16	0.32
2	20	0.40	36	0.72
3	9	0.18	45	0.90
4	5	0.10	50	1.00
	n=50	1		

- Escribir en la primera columna los valores que toma la variable: 1, 2, 3 y 4.
- Contar el número de veces que aparece cada valor en la muestra y escribir ese número en la segunda columna.
- Calculo de frecuencias relativas :

$$f_1 = \frac{n_1}{n} = \frac{16}{50} = 0.32 \qquad f_2 = \frac{n_2}{n} = \frac{20}{50} = 0.4$$

$$f_3 = \frac{n_3}{n} = \frac{9}{50} = 0.18 \qquad f_4 = \frac{n_4}{n} = \frac{5}{50} = 0.1$$

- Cálculo de las frecuencias absolutas acumuladas :

$$N_1 = n_1 = 16$$
$$N_2 = N_1 + n_2 = 16 + 20 = 36$$
$$N_3 = N_2 + n_3 = 36 + 9 = 45$$
$$N_4 = N_3 + n_4 = 45 + 5 = 50$$

- Cálculo de frecuencias relativas acumuladas:

$$F_1 = f_1 = 0.32$$
$$F_2 = F_1 + f_2 = 0.32 + 0.40 = 0.72$$
$$F_3 = F_2 + f_3 = 0.72 + 0.18 = 0.90$$
$$F_4 = F_3 + f_4 = 0.90 + 0.10 = 1.00$$

Solución c) ¿Cuántos alumnos tienen tres miembros de su familia que trabajan?

Se observa el valor de la variable x_3=3, y luego el valor que le corresponde en la columna de frecuencia absoluta es n_3 =9; por lo tanto, son 9 alumnos que tienen tres miembros de su familia que trabaja.

Solución d) ¿Qué porcentaje de alumnos tienen dos miembros de su familia que trabajan?

Se observa el valor de la variable x_2 = 2, y luego el valor que le corresponde en la columna de la frecuencia relativa es f_2=0.40*100=40%; por lo tanto, el 40% de los alumnos tienen dos miembros de su familia que trabajan.

Solución e) ¿Qué porcentaje de alumnos tienen tres o menos miembros de su familia que trabajan?

Se observa el valor de la variable $x_i \leq 3$; es decir, x_1 =1, x_2=2 y x_3=3, y luego el valor que le corresponde a x_3=3 en la columna de las frecuencias relativas acumuladas es F_3=0.90*100=90%; por lo tanto el porcentaje de alumnos que tienen tres o menos miembros de su familia que trabajan es el 90%.

Solución f) ¿Cuántos alumnos tienen menos de tres miembros de su familia que trabajan?

Se observa el valor de la variable $x_i<3$; es decir x_1 =1 y x_2=2, luego el valor que le corresponde a x_2=2 en la columna de las frecuencias absolutas acumuladas es F_2=36; por lo tanto el número de alumnos que tienen menos de tres miembros de su familia que trabajan es 36 alumnos.

DISTRIBUCIÓN DE FRECUENCIA AGRUPADAS

En estas distribuciones los valores observados generalmente aparecen agrupados en intervalos o clases $[L_{i-1},L_i]$ debido al elevado número de observaciones, y , por tanto, las frecuencias absolutas correspondientes a cada intervalo se obtiene como la cantidad de valores de la variable que contiene.

Este tipo de distribuciones se asocia, fundamentalmente, con variables cuantitativas y especialmente a variables continuas, aunque, en algunos casos también es aplicable a variables discretas, especialmente en aquellas situaciones en que la variable toma muchos valores diferentes y existe mucha variabilidad, de forma que si éstos no se agruparan, la tabla resultaría demasiado extensa y la función de síntesis de la misma se perdería.

Dada una variable X con valores $x_1, x_2, \ldots, x_n$; aparece una serie de conceptos generales que se definen a continuación.

En general los intervalos se definen de la siguiente forma:

$[l_{i-1}, l_i] = \{x: l_{i-1} \leq X \leq l_i\}$

$[l_{i-1}, l_i) = \{x: l_{i-1} \leq X < l_i\}$

$(l_{i-1}, l_i] = \{x: l_{i-1} < X \leq l_i\}$

Dónde:

l_{i-1}: Límite inferior del intervalo i-ésimo.

l_i : Límite superior del intervalo i-ésimo.

Pero la forma más usual de definirlos es:

$[l_{i-1}, l_i) = \{x: l_{i-1} \leq X < l_i\}$

Número de clases o intervalos (k).

El número de clases en que se dividen los datos de la variable no debe ser excesivo, puesto que pueden aparecer irregularidades accidentales si hay pocas observaciones en algunas clases. Por el contrario, si se elige un número reducido, se producirá una pérdida importante de información. Existen muchos criterios para determinar el número de clases, de los cuales se mencionan:

- Tomar k entre 5 y 16.
- Reglas empíricas de Sturges :

  $$k = 1 + 3.3 * \log(n) \quad \text{ó} \quad k = \frac{3}{2} + \frac{\log(n)}{\log(2)}$$

- $k \approx \sqrt{n}$, si la cantidad de datos es pequeña.
- Usar k tal que: $2^k = n$

En cualquiera de los casos se debe tomar la parte entera como valor de k.

Rango de los Datos (R): Representa el intervalo de dispersión de los datos de la muestra o población. Y se define como la diferencia entre el dato mayor y el dato menor. R = Max – Min

Amplitud del intervalo/clase i-esima (a_i): Representa el margen o rango de valores que incluyen los límites de cada clase. En general se define como la diferencia entre dos límites de clase inferiores consecutivos.

En la forma usual de representación de los intervalos se define como: $a_i = l_i - l_{i-1}$. Este valor puede definirse igual o distinto para todos los intervalos; pero se recomienda que sea igual para que la información organizada en la tabla no se vea distorsionada.

Para efectos de cálculo de la amplitud se obtiene como: $a = \frac{R}{k}$

Marca de clase o punto medio del intervalo/clase i- ésima (c_i): se considera como una forma abreviada de representar un intervalo mediante uno de sus puntos; y está definida por el punto medio de la clase, y se puede obtener de tres formas:

1. Por la semisuma de los valore extremos del intervalo: $c_i = \frac{l_i + l_{i-1}}{2}$
2. Se obtiene la primera marca de clase por el método anterior y si la amplitud (a) es constante, se le suma a la primera marca de clase obtenida y así sucesivamente.
3. Se divide la amplitud de cada intervalo (a) por dos y se le suma al límite inferior del intervalo o se le resta al límite superior del intervalo.

Densidad de frecuencia del intervalo/clase i-ésima (h_i): Se denomina densidad de frecuencia del intervalo $[L_{i-1}, L_i)$, a la proporción de observaciones por unidad de amplitud, y se define : $h_i = \frac{f_i}{a_i}$ ó $h_i = \frac{n_i}{a_i}$. Se utiliza cuando los intervalos son de diferente amplitud.

Frecuencia absoluta del intervalo/clase i- ésima (n_i): Se denomina frecuencia absoluta del intervalo $[L_{i-1},L_i)$, a la cantidad de x_i incluidos en el intervalo o clase i-ésima.

Frecuencia relativa del intervalo/clase i- ésima (fi): Se denomina frecuencia relativa del intervalo del intervalo $[L_{i-1},L_i]$, a la proporción de x_i incluidos en el intervalo o clase i-ésima. Y se define como el cociente de la frecuencia absoluta del intervalo/clase i-ésima y el número total de valores de la variable (n). Es decir

$$f_i = \frac{n_i}{n}$$

Frecuencia absoluta acumulada (N_i): Se denomina frecuencia absoluta acumulada del intervalo $[L_{i-1},L_i)$, a la suma de las frecuencias absolutas de ese intervalo y todos los intervalos anteriores.

Frecuencia relativa acumulada (Fi): Se denomina frecuencia relativa acumulada del intervalo $[L_{i-1},L_i)$, a la suma de las frecuencias relativas de ese intervalo y todos los intervalos anteriores.

La distribución de frecuencias de una variable suele presentarse ordenadamente mediante la tabla de frecuencias siguiente:

Tabla 3. Distribución de Frecuencias Agrupadas.

Clases $[L_{i-1},L_i)$	Marca de Clases C_i	Frecuencias Absolutas n_i	Frecuencias Relativas $f_i = \frac{n_i}{n}$	Frecuencias Absolutas Acumuladas N_i	Frecuencias Relativas Acumuladas $F_i = \frac{N_i}{n}$
$[L_0 , L_1)$	C_1	n_1	$f_1 = \frac{n_1}{n}$	$N_1 = n_1$	$F_1 = \frac{N_1}{n}$
$[L_1 , L_2)$	C_2	n_2	$f_2 = \frac{n_2}{n}$	$N_2 = n_1 + n_2$	$F_2 = \frac{N_2}{n}$
.		.	.	.	.
.		.	.	.	.
.		.	.	.	.
$[L_{i-1},L_i)$	C_i	n_i	$f_i = \frac{n_i}{n}$ fi	$N_i = n_{i-1} + n_i$	$F_i = \frac{N_i}{n}$
.		.	.	.	.
.		.	.	.	.
.		.	.	.	.
$[L_{k-1},L_k)$	C_k	n_k	$f_k = \frac{n_k}{n}$ fk	$N_k = \sum_{i=1}^{k} n_i = n$	$F_k = \frac{N_k}{n} = 1$
		$\sum_{i=1}^{k} n_i = n$	$\sum_{i=1}^{k} f_i = 1$		

PASOS PARA ELABORAR LA TABLA DE FRECUENCIAS AGRUPADA.

1. Ordenar los datos de menor a mayor; si hay datos repetidos se deben considerar tantas veces como aparezcan.
2. Determinar el rango de los datos.
3. Definir la cantidad de clases a utilizar.
4. Obtener la amplitud que tendrán las clases.
5. Encontrar la marca de clase o punto medio de clase.
6. Determinar la frecuencia absoluta de cada clase.
7. Calcular las demás columnas por medio de las formulas.

Ejemplo 3: Se tienen los datos de las temperaturas que se registraron en el mes de marzo del año 2011 en San Salvador y han sido reportadas por la estación meteorológica: 786630 (MSSS) con sede en Ilopango.

27.3, 25.8, 25.0, 25.0, 24.6, 25.0, 24.3, 25.2, 24.9, 23.8, 23.0, 24.3, 25.1, 25.1, 24.8, 23.8, 26.1, 27.4, 25.8, 25.8, 25.7, 24.8, 24.7, 23.8, 24.2, 24.1, 24.7, 25.8, 26.2, 26.5, 25.8

Determine:

a) Variable en estudio y su tipo.
b) Elabore la tabla de distribución de frecuencia correspondiente.
c) ¿Cuantos días se tuvo una temperatura en el mes de marzo menor a los 27° C?.
d) ¿Que porcentajes de días del mes de marzo se tuvo una temperatura mayor o igual que 24.6 ° C?.
e) ¿Cuantos días se tuvo en el mes de marzo una temperatura mayor que 23.8 ° C, pero menor de 27.0 C?.

Solución a) La variable en estudio es la temperatura registradas en el mes de marzo y es de tipo cuantitativa continua.

Solución b) Elaboración de Tabla de Frecuencia.

- Ordenar los datos de menor a mayor.

23.0 , 23.8 , 23.8 , 23.8 , 24.1 , 24.2 , 24.3 ,24.3 , 24.6 , 24.7 , 24.7 ,24.7 ,24.8 , 24.8 ,24.9 , 25.0, 25.0 , 25.1 ,25.1 ,25.2 ,25.7 ,25.8 , 25.8 , 25.8 , 25.8 , 25.8 , 26.1 , 26.2 , 26.5 , 27.3 , 27.4

Tamaño de la muestra: 31 (n=31 todos los días del mes de marzo)

- Cálculo del rango (R):
  $R = Max - Min = 27.4 - 23.0 = 4.4$
- Cálculo del número de clases: como la cantidad de datos es pequeña.
  $k = \sqrt{31} = 5.567 \approx 6$
- Determinar la amplitud de la clase(A):
  $A = \dfrac{R}{k} = \dfrac{4.4}{6} = 0.733 \approx 0.8$
- Determinación de los límites de las clases:

Límite inferior de la clase= Límite inferior de la clase + amplitud de la clase.

Límite inferior de la primera clase:

23 (valor mínimo de los datos)

Límite inferior de la segunda clase:

$23 + 0.8 = 23.8$

Límite inferior de la primera clase:

$23.8 + 0.8 = 24.6$

Límite inferior de la primera clase:

$24.6 + 0.8 = 25.4$

Límite inferior de la primera clase:

$25.4 + 0.8 = 26.2$

Límite inferior de la primera clase:

$26.2 + 0.8 = 27.0$

Límite inferior de la primera clase:

$27.0 + 0.8 = 27.8$

Como los intervalos se define de la forma: $[l_{i-1}, l_i) = \{ x: l_{i-1} \leq X < l_i\}$, las clases quedan definidas como:

23.0 - 23.8, 23.8 - 24.6, 24.6 - 25.4,

25.4 - 26.2, 26.2 - 27.0, 27.0 - 27.8.

- Llenado de la tabla.
  - Escribir el encabezado de las columnas.
  - Colocar los límites de los intervalos en la columna 1.
  - Determinar la marca de clase por medio de la fórmula: $c_i = \frac{l_i + l_{i-1}}{2}$ donde, $c_i = \frac{23.0 + 23.8}{2} = 23.4$, las siguientes se obtienen sumándoles la amplitud (0.8) a cada una de ellas. Y colocarlas en la columna 2.
- Determinar las frecuencias absolutas (ni), para ello se cuenta la cantidad de valores que contiene cada clase.

Primer clase: [23.0 - 23.8), se encuentra solo el valor de 23.0, ya que el 23.8 pertenece al siguientes intervalo, entonces $n_1 = 1$.

Segunda clase: [23.8 - 24.6), los valores que se encuentra son: 23.8, 23.8, 23.8, 24.1, 24.2, 24.3 ,24.3, entonces $n_2 = 7$,…… Colocarlos en la columna 3.

Las siguientes columnas se encuentran de la misma forma como se explicó en el ejemplo 1.

La tabla que se obtiene de los cálculos anteriores queda como se presenta a continuación.

Clases $[L_{i-1}, L_i)$	Marca de Clases C_i	Frecuencias Absolutas n_i	Frecuencias Relativas $f_i = \frac{n_i}{n}$	Frecuencias Absolutas Acumuladas N_i	Frecuencias Relativas Acumuladas $F_i = \frac{N_i}{n}$
[23.0 - 23.8)	23.4	1	0.032	1	0.032
[23.8 - 24.6)	24.2	7	0.226	8	0.258
[24.6 - 25.4)	25.0	12	0.387	20	0.675
[25.4 - 26.2)	25.8	7	0.226	27	0.871
[26.2 - 27.0)	26.6	2	0.0645	29	0.9355
[27.0 - 27.8)	27.4	2	0.0645	31	1
		n=31	1		

Tabla 4 Propuesta de solución.

Solución c) ¿Cuántos días se tuvo una temperatura en el mes de marzo menor a los 27 º C?

Como lo que se pide es cantidad de días se debe encontrar a partir de las frecuencias absolutas.

Forma 1: Sumar los valores de las frecuencias absolutas que corresponden a dicha condición, $n_1 + n_2 + n_3 + n_4 + n_5 = 1 + 7 + 12 + 7 + 2 = 29$ días.

Forma 2: Por medio de los valores de las frecuencias absolutas acumuladas que corresponden al valor de $N_5 = 29$ días, ya que es la cantidad de días que corresponde al valor de la temperatura menor de 27 ° C. Por lo tanto son 29 días del mes de marzo que se tuvo una temperatura menor de 27.0 ° C.

Solución d): ¿Qué porcentajes de días del mes de marzo se tuvo una temperatura mayor o igual que 24.6 ° C?

Como lo que se pide es porcentaje de días, se debe encontrar a través de las frecuencias relativas que corresponde a la suma de $f_3 + f_4 + f_5 + f_6 = 0.387 + 0.226 + 0.0645 + 0.0645 = 0.742*100 = 74.2\%$, lo que significa que el 74.2% de los 31 días del mes de marzo se tuvo una temperatura mayor o igual que 24,6C.

Solución e): ¿Cuántos días se tuvo en el mes de marzo una temperatura mayor que 23.8 ° C, pero menor de 27.0 ° C?

Como lo que se pide es cantidad de días se debe encontrar a partir de las frecuencias absolutas, que corresponde a dicho intervalo; es decir $n_2 + n_3 + n_4 + n_5 = 7 + 12 + 7 + 2 = 28$ días. Por lo tanto, son 28 días que se tuvo en el mes de marzo una temperatura mayor que 23.8 ° C, pero menor de 27.0 ° C.

No existen normas establecidas para determinar cuándo es apropiado utilizar datos agrupados o datos no agrupados; sin embargo, se sugiere que cuando el número total de datos (n) es igual o superior a 50 y además el rango o recorrido de la serie de datos es mayor de 20, entonces, se utilizará *la* ***distribución de frecuencia para datos agrupados***, también se utilizará este tipo de distribución cuando se requiera elaborar gráficos como histograma, el polígono de frecuencia o la ojiva.

APLICANDO LO APRENDIDO

1. El gobierno de El Salvador desea averiguar si el número medio de hijos por familia ha descendido respecto de la década anterior. Para ello ha encuestado a 50 familias respecto al número de hijos, y ha obtenido los siguientes datos:

2	4	2	3	1	2	4	2	3	0
2	2	2	3	2	6	2	3	2	2
3	2	3	3	4	3	3	4	5	2
0	3	2	1	2	3	2	2	3	1
4	2	3	2	4	3	3	2	2	1

Determinar:

a) ¿Cuál es la población objeto de estudio? R/Conjunto de familias de El Salvador
b) ¿Qué variable estamos estudiando? R/ Número de hijos por familia
c) ¿Qué tipo de variable es? R/ es discreta ya que el número de hijos solo puede tomar determinados valores enteros (es imposible tener medio o un cuarto de hijo).
d) Construir la tabla de frecuencias.

Podemos ver que el número de hijos, toma los valores existentes entre 0 hijos, los que menos y, 6 hijos los que más; de esta manera se tiene:

Valores de la Variable x_i	Frecuencias Absolutas n_i	Frecuencias Absolutas Acumuladas N_i	Frecuencias Relativas f_i	Frecuencias Relativas Acumuladas F_i
0	2	2	0.04	0.04
1	4	6	0.08	0.12
2	21	27	0.42	0.54
3	15	42	0.30	0.84
4	6	48	0.12	0.96
5	1	49	0.02	0.98
6	1	50	0.02	1
	n=50		1	

e) ¿Cuál es el número de familias que tiene como máximo 2 hijos? R/ 2+4+21=27
f) ¿Cuántas familias tienen más de 1 hijo, pero como máximo 3? R/ 21+15=36
g) ¿Qué porcentaje de familias tiene más de 3 hijos?
R/ (0,12 + 0,02 + 0,02)*100 = 0,16*100 = 16%.

2. Un nuevo hotel va a abrir sus puertas en cierta ciudad. Antes de decidir el precio de sus habitaciones, el gerente investiga los precios por habitación de 40 hoteles de la misma categoría de esa ciudad. Los datos obtenidos en miles de dólares fueron:

3.9	4.7	3.7	5.6	4.3	4.9	5.0	6.1	5.1	4.5
5.3	3.9	4.3	5.0	6.0	4.7	5.1	4.2	4.4	5.8
3.3	4.3	4.1	5.8	4.4	4.8	6.1	4.3	5.3	4.5
4.0	5.4	3.9	4.7	3.3	4.5	4.7	4.2	4.5	4.8

Se pide:

a) ¿Cuál es la población objeto de estudio? R/ Los hoteles de una ciudad
b) ¿Qué variable estamos estudiando? R/Precio de alquiler de habitaciones
c) ¿Qué tipo de variable es? R/Cuantitativa continúa.
d) ¿Qué problema plantea la construcción de la tabla de frecuencias? R/ El problema que plantea es que existen muchos valores diferentes por tanto es bueno agrupar la serie en intervalos.

Clases $[L_{i-1},L_i)$	Marca de Clases C_i	Frecuencias Absolutas n_i	Frecuencias Absolutas Acumuladas N_i	Frecuencias Relativas $f_i = \frac{n_i}{n}$	Frecuencias Relativas Acumuladas $F_i = \frac{N_i}{n}$
[3.25 - 3.75)	3.25	3	3	0.075	0.075
[3.75 - 4.25)	4.00	8	11	0.200	0.275
[4.25 - 4.75)	4.50	14	25	0.350	0.625
[4.75 - 5.25)	5.00	6	31	0.150	0.775
[5.25 – 5.75)	5.50	4	35	0.100	0.875
[5.75 - 6.25)	6.00	5	40	0.125	1.00
		n=40		1.00	

e) ¿Cuántos hoteles tienen un precio entre 3,25 y 3,75? R/ 3 hoteles
f) ¿Cuántos hoteles tienen un precio superior a 4,75? R/ 15 hoteles
g) ¿Qué porcentaje de hoteles cuestan como mucho 4,25? R/ 27,5%

ACTIVIDADES

1. Escriba el tipo de variable a que pertenecen los enunciados siguientes:
   a) Número de músculos de los animales vertebrados. R/Cuantitativa Discreta
   b) La intención de voto de los ciudadanos salvadoreños. R/Cualitativa
   c) Talla de los pantalones de los alumnos de tu centro escolar. R/ Cuantitativa Discreta
   d) Tipos de refrescos que prefieren tus compañeros de aula. R/ Cualitativa

2. Realiza el siguiente experimento
   a) Lanza cuatro monedas al mismo tiempo (o una moneda cada cuatro veces sucesivas unas 20 veces. Cada moneda muestra “cara” o “cruz”.
   b) Anota cuántas veces aparece “cara”. Por ejemplo: 1 vez cara, 3 veces cara, etc; brevemente: 1, 3, 1,0,...
   c) Reúne los resultados en una tabla, indicando las frecuencias absolutas y relativas para los resultados.

Resultado	0 veces	1 vez	2 veces	3 veces	4 veces
Frecuencia Absoluta (Número de veces de la ocurrencia=X)					
Frecuencia Relativa(X entre el total de lanzamiento 20)					

3. Repaso de frecuencias, frecuencias relativas y redondeo

De 24 participantes, 20 han dado el examen de Matemática. ¿A qué porcentaje de participantes corresponde? Redondea el resultado a un porcentaje entero.

Frecuencia absoluta: 20
Frecuencia relativa: 83%

4. De 56 participantes, el 80% llega puntual a las reuniones de orientación. Fidel calcula que 44 son puntuales. Edita cree que son 45. ¿Quién tiene la razón?

   Frecuencia absoluta: 45
   Frecuencia relativa: 80%
   R/ Edita (45)

5. Juana practica en clase lanzando un dado unas 400 veces. Complete la tabla siguiente.

Puntos	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Frecuencia Absoluta	71	59	66	59	75	70	400
Frecuencia Relativa							100%

6. El SNET (Sistema Nacional de Estudios Territoriales), reporta la información referente a los sismos de mayor impacto en relación a pérdidas humanas que ocasionaron durante el final del siglo XX e inicios del siglo XXI. La cual se presenta a continuación.

Ranking	Fecha	Magnitud	Pérdidas Humanas	Lugar del Impacto
1	10/10/1986	5.4	1500	San Salvador
2	07/06/1917	6.7	1050	San Salvador
3	13/01/2001	7.6	944	Territorio Nacional
4	06/05.1951	6.2	400	Jucuapa-Chinameca
5	13/02/2001	6.6	315	Zona Paracentral(San Vicente, Cuscatlán, La paz, Usulután, Cabañas)
6	20/12/1936	6.1	100-200	San Vicente
7	03/05/1965	6.0	125	San Salvador
8	28/04/1919	5.9	100	Zona Central (San Salvador, La Paz, La libertad)
9	19/06/1982	7.0	8	Territorio Nacional

Fuente: SNET

Según la tabla presentada, realice lo siguiente:

1. Elabore una tabla de distribución de frecuencia no agrupada para la fecha en que se realizó el sismo y pérdidas humanas.
   a) ¿Cuál es el total de pérdidas humanas que han ocasionado los sismos en el salvador?
   b) ¿En qué fecha hubieron más pérdidas humanas?
   c) ¿Qué porcentaje de pérdidas humanas hubieron en el sismo registrado el 10/10/1986?
   d) ¿Cuántas pérdidas humanas hubieron en el sismo registrado el 06/05/1951?

e) ¿En qué fecha hubo menos pérdidas humanas?

2. Organice la información por décadas. Elabore una tabla de distribución de frecuencias no agrupadas de acuerdo a décadas y pérdidas humanas.
   a) ¿En que década hubieron menos pérdidas humanas? ¿Y qué porcentaje representa?
   b) ¿Cuántas pérdidas humanas hubo en la década de 1910?
   c) ¿Qué porcentaje de pérdidas humanas hubieron entre las décadas de 1980 y 2000?
   d) ¿Cuántas pérdidas humanas hubo en las décadas 1950-1980?
   e) ¿Qué porcentaje de pérdidas humanas hubieron en la década del 2000?
3. ¿Qué otras conclusiones importantes puede obtener de esta información?

GRÁFICOS ESTADÍSTICOS

Un gráfico es una representación más compacta de una serie de datos que permiten ser leídos de manera más fácil por las personas.

Los datos estadísticos se presentan comúnmente en forma de cuadro o de gráfico. La representación gráfica resulta eficaz para obtener rápidamente una impresión de conjunto de una serie de datos haciendo resaltar sus relaciones. Cuando el gráfico utilizado es el adecuado es una manera clara, simple y efectiva de presentar la información permitiendo un análisis visual. La representación gráfica puede ser una ayuda en la interpretación del contenido de un cuadro pero no un método sustitutivo. En la elaboración de un informe los gráficos deben ir acompañados de las correspondientes tablas, dado que los gráficos son útiles para proporcionar una idea de la situación general pero no de los detalles.

Al preparar un gráfico, la persona se manifiesta como un artista, que aporta su imaginación y su temperamento para comunicar un mensaje que su auditorio debe asimilar.

Antes de realizar cualquier análisis complejo, lo primero que se debe hacer es dar un análisis descriptivo de los datos que sea lo más sencillo y claro posible.

El objetivo de las representaciones gráficas es realizar una síntesis visual de la información contenida en una tabla de distribución de frecuencias y fundamentalmente se evidencian tres características de las distribuciones que son: su forma, acumulación o tendencia y dispersión o variabilidad.

Existen ciertas reglas generales que son comúnmente aceptadas en referencia a la construcción del gráfico. Las más importantes son las siguientes:

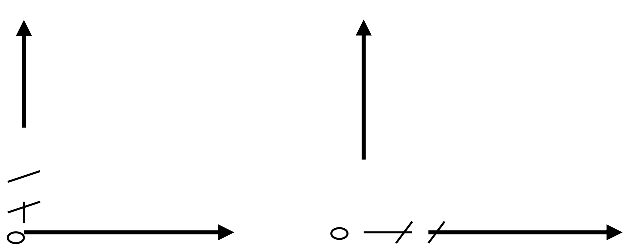
- El gráfico para alcanzar su objetivo debe ser sencillo por lo tanto no debería contener más líneas o símbolos que los que el ojo pueda seguir con comodidad.
- Una gráfica debe explicarse por sí misma por lo que debe contener título, origen, escalas, etc., necesarias para que el lector la interprete.
- El diagrama o gráfico progresa generalmente de izquierda a derecha y de abajo a arriba por lo que toda leyenda deber ser escrita para leerse hacia arriba o hacia la derecha.
- Las líneas que corresponden al gráfico deben ser más gruesas que los ejes.

Recomendaciones para la construcción de gráficos

- **Título:** Todo gráfico, al igual que las tablas, deberá tener un título en el que se manifieste de manera clara y concisa aquello que se quiere mostrar. El título se coloca habitualmente en la parte superior del gráfico pero puede aparecer en la parte inferior.

- **Texto:** Todos los Textos de un gráfico, incluyendo las escalas, los valores de la escala, símbolos y cualquier otra palabra, deben colocarse, si es posible, horizontalmente. A veces puede ser necesario colocar el nombre de la escala vertical en posición vertical. Deberán usarse letras minúsculas con la primera letra mayúscula para las escalas.

- **Cero en la Escala:** Deben incluirse siempre. Para los gráficos lineales y los histogramas se coloca siempre en la escala vertical y a veces en la horizontal. En algunas ocasiones puede ser necesario cortar la escala a fin de destacar mejor las diferencias entre las magnitudes representadas. En todos los casos debe indicarse con el 0 el origen de la escala. Cuando los datos disponibles se inician a partir de cifras muy distantes de cero se puede reducir el espacio requerido para el gráfico mediante algún corte como los siguientes:



Estos cortes deben ser claramente visibles.

Al marcarse las escalas debe recordarse que iguales distancias indicarán siempre magnitudes iguales. No se permite cambios de escala o de unidad de medida en el mismo eje del mismo gráfico estadístico.

- **Coordenadas:** A fin de asegurar la comprensión del gráfico deben marcarse claramente las dos escalas. No sólo deberá indicarse la naturaleza de la variable, sino también deberá expresarse la unidad de medida. A fin de resaltar la curva (o curvas) que representan los datos, éstas deben dibujarse con un trazado más grueso que las coordenadas.

Cuando se tracen varias curvas en un mismo gráfico, es esencial que cada curva se destaque con claridad; para ello pueden utilizarse distintos trazados o se puede recurrir al uso de colores, especialmente en los gráficos murales. **Ejemplo de distintos trazos.**

____
-.-
oooooooooooooooooooooooooooooooooooo

- **Escalas:** Las escalas del eje horizontal (también llamado de las X´s o abscisa) y del eje vertical (o de las Y´es u ordenada) siguen el sistema de coordenadas rectangulares, sólo que normalmente se utiliza el primer cuadrante para la mayoría de los gráficos estadísticos. Generalmente, el eje de las Y' es o la ordenada se utiliza para indicar la magnitud de los diagramas que representan a los datos: frecuencias absolutas, porcentajes, tasas, etc. y la escala del eje de las X´s o abscisa se usa para designar las categorías de la característica a la que se refiere el gráfico.

Es necesario indicar siempre cuál es la información dada por cada uno de esos ejes, pues sin este detalle resultaría imposible el análisis e interpretación de cualquier gráfico.

Siempre hay que anotar el nombre de cada escala e indicar la unidad de medida que corresponda y tomar el cero como punto de partida de las escalas.

- **Diagrama:** Los diagramas equivalen al cuerpo de un cuadro estadístico, se usan para presentar los datos del gráfico. Entre los tipos más comunes de diagramas usados en gráficos estadísticos están: líneas, barras, símbolos, mapas, etc.
- **Número:** Cuando se incluye en un trabajo más de un gráfico, éstos se numeran (en forma independiente de los cuadros) para facilitar la referencia a uno en particular.
- **Simbología:** Si para hacer comparaciones se utiliza sombreados, punteados o colores para identificar categorías de una misma serie u otras series estadísticas, es necesario agregar a la derecha del gráfico el significado de tal simbología.
- **Fuentes:** La fuente de los datos con los cuales se construyó el gráfico debe aparecer siempre en la parte inferior de éste. Si el gráfico fue construido tomando datos de uno o de varios cuadros incluidos en el trabajo, la fuente serían dichas tablas, indicándose cuáles por el número que corresponde a cada uno.
- **Notas Explicativas:** Como en las tablas, en un gráfico pueden aparecer notas explicativas, ya sea de tipo preliminar o al pié del mismo.
- **Otras Recomendaciones**: No es aconsejable tratar de presentar en un gráfico demasiada información o demasiado exacta. Para ello se usará una tabla. Un gráfico que contenga demasiadas líneas, curvas o números resulta confuso y no satisface el criterio primordial de toda buena presentación gráfica: selección acertada de los rasgos importantes del problema, claridad y comprensión rápida por parte del lector.

Existen una gran variedad de gráficos pero todo depende del problema y del investigador para escoger uno de ellos. De los más usuales se pueden mencionar:

- Diagrama de Barras
- Diagrama Circular
- Pictogramas
- Histogramas
- Polígono de Frecuencias
- Ojivas

Estas representaciones tienen la gran ventaja, de que no se necesita mucha formación para entender e interpretar la información que en ellos se presenta, por lo tanto llega con la misma facilidad a las personas comunes y corrientes, lo mismo que a los profesionales de cualquier rama del saber.

Un buen gráfico debería mostrar la información que se quiere transmitir y no desviar la información original bajo ningún motivo, y tampoco es recomendable saturarlo de información pues se perdería el interés en estudiarlo.

DIAGRAMA DE BARRAS

Este tipo de diagrama se utiliza para representar datos principalmente de variables de tipo cualitativas, pero también puede utilizarse para variables cuantitativas discretas, y en general para distribuciones de frecuencias no agrupadas. Un primer paso para su elaboración es crear la tabla de distribución de frecuencias no agrupada.

Consiste en dos ejes perpendiculares y una barra o rectángulo para cada categoría o valor de la variable. Por lo general, se suele colocar

en el eje horizontal las categorías o valores de la variable (aunque también se puede hacer en el vertical), y en el eje vertical las frecuencias absolutas o relativas, el eje se gradúa según los valores que toman las frecuencias. La representación gráfica consiste en dibujar una barra o un rectángulo para cada una de las categorías o valores de la variable cuya altura sea la del valor que alcanza la frecuencia en el eje vertical.

Todas las barras o rectángulos deben ir separadas y tanto el ancho como la distancia que las separa son arbitrarios, pero se debe tomar en cuenta que las barras no deben ser demasiado cortas y anchas o demasiado angostas y largas, dejando entre barra y barra un espacio que no sea menor que la mitad del ancho de una barra ni mayor que el ancho de la misma, una vez fijados deben mantenerse para todo el gráfico (igual ancho y estar igualmente espaciados); representándose tantas barras o rectángulos como categorías o valores tenga la variable.

Pasos para la construcción:

Paso 1: Representar las categorías o valores de la variable en el eje horizontal.

Paso 2: Usar una escala adecuada para representar la frecuencia en el eje vertical.

Paso 3: Dibuje una barra o rectángulo de igual ancho justo sobre cada categoría o valor del eje horizontal con altura igual a la frecuencia de cada categoría o valor.

La barra más alta corresponde a la categoría o valor de la variable que tiene mayor frecuencia, mientras que la más baja corresponde a la categoría o valor que tiene menor frecuencia.

A veces, se utilizan para hacer comparaciones múltiples, diferenciando con colores las barras de cada variable.

Ejemplo 4: Durante el empadronamiento llevado a cabo en el VI Censo de Población y V de Vivienda del año 2007, se contabilizó una población de 5,744,113 habitantes en El Salvador. La siguiente tabla presenta esta población distribuida por departamento.

Departamentos	Ahuachapán	Santa Ana	Sonsonate	Chalatenango	La Libertad
N$_0$. Habitantes	319,503	523,655	438,960	192,788	660,652

Departamentos	San Salvador	Cuscatlán	La Paz	Cabañas	San Vicente
N$_0$. Habitantes	$1_1$567,156	231,480	308,087	149,326	161,645

Departamentos	Usulután	San Miguel	Morazán	La Unión	Total
N$_0$. Habitantes	344,235	434,003	174,406	238,217	$5_1$744,113

Tabla 5: Densidad Población de El Salvador por Departamentos

Al elaborar el diagrama de barras, cada barra representa un departamento y la altura de la misma será el número de habitantes (frecuencia) para lo cual habrá que utilizar una escala adecuada.

El siguiente diagrama de barra presenta la densidad poblacional por departamento del censo de 2007 en términos de frecuencias absolutas.

Gráfico 1: Diagrama de Barras de la densidad Poblacional de El Salvador
Fuente: DIGESTYC

Interpretación:

- Se observa que la barra más alta corresponde al departamento de San Salvador, lo que significa que es el que tiene más número de habitantes.
- Las barras de los departamentos de Cabañas y San Vicente presentan una altura en sus barras muy similares lo que indican que tienen casi igual número de habitantes.

¿Qué otras conclusiones pueden obtener a partir de este diagrama de barras?

Características de los gráficos de columnas

- No muestran frecuencias acumuladas.
- Se prefiere para el tratamiento de datos cualitativos o cuantitativos discretos.
- La columna (o barra) con mayor altura representa la mayor frecuencia.
- Son fáciles de elaborar.
- Suelen utilizarse para representar tablas de distribución de frecuencia no agrupada.
- La sumatoria de las alturas de las columnas equivalen al 100% de los datos.

DIAGRAMA CIRCULAR

En el caso de variables de tipo cualitativas el diagrama circular se utiliza con mucha frecuencia; aunque también se utilizan para variables de tipo cuantitativo discretas sin agrupar, y es conocido también con el nombre de diagrama de sectores o tortas o pastel.

Este tipo de diagramas consideran una figura geométrica en que la distribución de frecuencias se reparte dentro de la figura como puede ser una dona, circulo o anillo, en el que cada porción dentro de la figura representa la información porcentual del total de datos, y consiste en un círculo en donde se representan los diferentes atributos o categorías de la variable, mediante partes o sectores circulares como frecuencias existan, que tienen una amplitud en grados proporcional a la frecuencia absoluta o relativa.

Para su elaboración se traza una circunferencia de radio arbitrario y se divide su círculo en tantas partes o sectores como categorías tenga la variable; de tal forma que a cada categoría le corresponda una parte o sector del círculo en grados proporcional a la frecuencia absoluta o relativa.

La circunferencia tiene en su interior 360°, los cuales se hacen corresponder al total (100%) de la información obtenida de la variable, para efectos de su elaboración y determinar el número en grados que corresponde a cada parte o sector de la categoría de la variable se divide la frecuencia absoluta correspondiente entre el total y esto se multiplicarlo por 360°; es decir:

$$\alpha_i = \left(\frac{n_i}{n}\right) * 360° \quad \text{o} \quad \alpha_i = f_i * 360°$$

Habrán tantos α_i como categorías tenga la variable y todo ellos deberán sumar 360°.

Para elaborar el dibujo se deberá utilizar compás, regla y transportador de ángulos. Y primero se traza la circunferencia y luego una línea del centro al radio en cualquiera posición que se utiliza como inicio de los sectores; luego se va midiendo para cada sector el ángulo(α_i) correspondiente con el transportador hasta completar todos los sectores en la primera línea trazada, y a cada sector de la circunferencia marcado se le coloca como rotulo o leyenda el nombre de la categoría de la variable y el porcentaje que le corresponde de la frecuencia relativa. De manera opcional cada sector se debe colorear de un color diferente, para diferenciarlos unos de otros.

Para su interpretación se observa la amplitud circular de cada uno de los sectores y compara con los demás sectores, tomando en cuenta la información de las leyendas de cada sector.

Ejemplo 5: Retomando la información del ejemplo 1, se hará un diagrama circular para la variable en estudio. Para elabora el diagrama circular se parte de la siguiente tabla:

Departamentos	Número de habitantes (n_i)	Frecuencia Relativa (f_i)	Frecuencia Relativa porcentual f_i*100	Grados α_i
Ahuachapán	319,503	0.06	5.56	20.02
Santa Ana	523,655	0.09	9.12	32.82
Sonsonate	438,960	0.08	7.64	27.51
Chalatenango	192,788	0.03	3.36	12.08
La Libertad	660,652	0.12	11.50	41.40
San Salvador	1,567,156	0.27	27.28	98.22
Cuscatlán	231,480	0.04	4.03	14.51
La Paz	308,087	0.05	5.36	19.31
Cabañas	149,326	0.03	2.60	9.36
San Vicente	161,645	0.03	2.81	10.13
Usulután	344,235	0.06	5.99	21.57
San Miguel	434,003	0.08	7.56	27.20
Morazán	174,406	0.03	3.04	10.93
La Unión	238,217	0.04	4.15	14.93
Total	5,744,113	1.00	100.00	360

Tabla 6: Densidad Población de El Salvador por Departamentos

Para elaborar el diagrama circular se hace una circunferencia de cualquier radio y como la variable departamentos tiene 14 categorías, por tanto el círculo estará dividido en 14 sectores y a cada sector se le hará corresponder el valor de los grados α_i de columna que se irán marcando a partir de una línea inicial con el transportador; por último cada sector se le pondrá la leyenda de la categoría y el valor porcentual de la frecuencia relativa. No olvidar escribir el título del diagrama.

A continuación se presenta el diagrama circular de la Densidad Poblacional de el Salvador del Censo 2007.

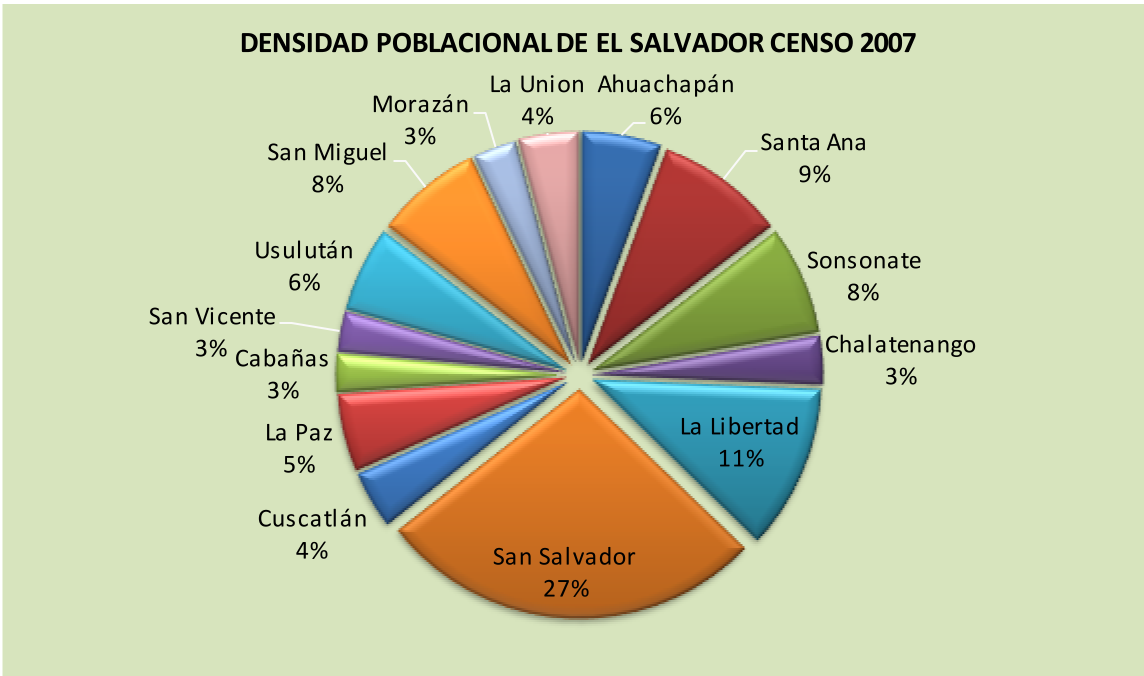


Gráfico 2: Diagrama circular de la densidad Poblacional de El Salvador
Fuente: DIGESTYC

Interpretación:

- Se observa que el sector más grande le corresponde al Departamento de San Salvador con un 27% del total de la población, lo que indica que más de la cuarta parte de la población del país reside en dicho departamento.
- Si a San Salvador se agrega la población del departamento de La Libertad (11%) y Santa Ana (9%) el porcentaje sube 47%, lo que indica que en estos tres departamentos reside casi la mitad de la población Salvadoreña.
- Cerca de la tercera parte de la población Salvadoreña (33%) reside en los departamentos de Sonsonate (8%), San Miguel (8%), Usulután (6%), Ahuachapán (6%) y La Paz (5%).

¿Qué otras conclusiones pueden obtener a partir de este diagrama circular?

PICTOGRAMAS

Son gráficos con dibujos alusivos a lo que se está estudiando y cuyo tamaño es proporcional a la frecuencia que representan. Se llaman también gráfica de figuras, estadística de figuras o lenguaje estadístico internacional. Se usan para hacer más llamativa la representación y de fácil comprensión por su sencillez; pero son diagramas poco precisos, y son especialmente útiles para fines publicitarios por ser atractivos y de fácil comprensión.

Consiste en un gráfico de barras horizontal o vertical en los que las barras se sustituyen por dibujos alusivos a la variable o lo que se quiere expresar. Cada dibujo representa un número determinado de unidades, por lo tanto, debe repetirse tantas veces como sea necesario para reflejar el valor de su frecuencia. Para interpretar estas gráficas basta conocer el valor a que equivale cada figura o signo.

Ejemplo 6: El censo Agropecuario llevado a cabo en los años 2007-2008 en El Salvador, reporta una superficie de frutales de 19,122 Mz. con una producción de 3,756,666 QQ. reporta que en El Salvador se cultivan 42 tipos de frutales y entre los cultivos con mayor producción y superficie se tiene: naranja, coco, limón y plátano.

A continuación se presenta un pictograma donde se observa la producción de frutales por departamento.

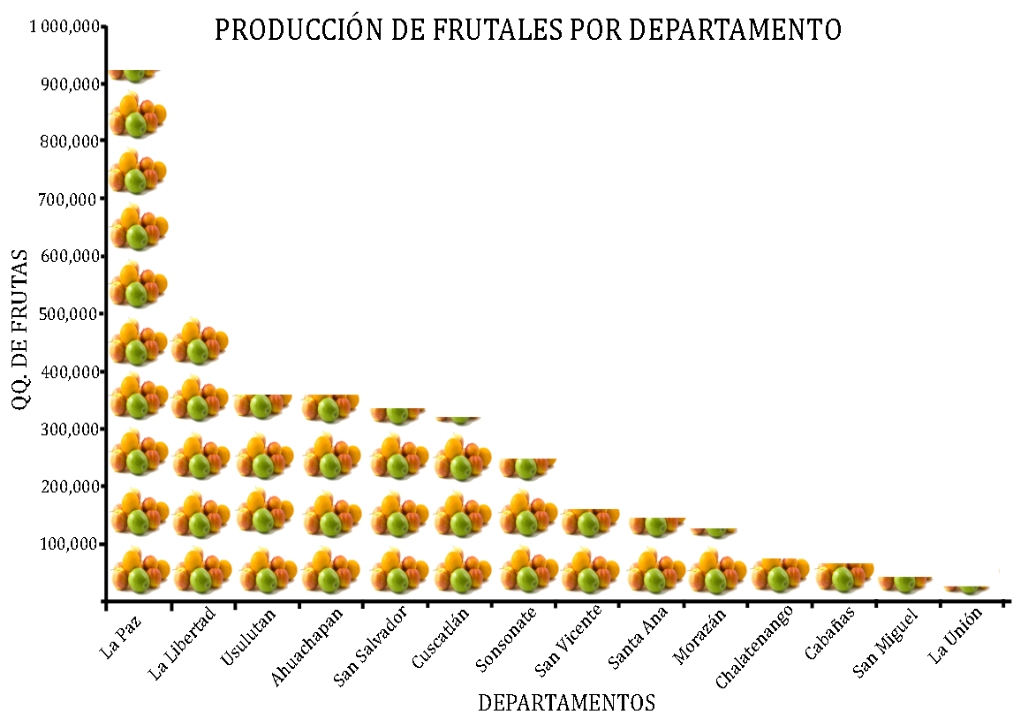


Gráfico 3: Pictograma de producción de frutales en El Salvador.
Fuente: DIGESTYC

Los departamentos con mayor producción de frutales son: La Paz, La Libertad y Usulután.

HISTOGRAMA

El histograma es una gráfica que se utiliza para representar variables de tipo cuantitativas continuas, y también para variables cuantitativas discretas que haya justificado su organización en distribuciones de frecuencias agrupadas en intervalos o clases; y es una de las gráficas más ampliamente utilizadas porque es más fácil de entender.

Se construyen de tal manera que en el eje horizontal se presentan los intervalos o clases de los valores de la variable y en el eje vertical sus respectivas frecuencias; y sobre cada intervalo de la variable se dibuja un rectángulo o barra de manera adyacente (sin huecos entre sí) de área proporcional a la frecuencia correspondiente a dicho intervalo.

En el eje vertical se puede representar no sólo el número de frecuencias, sino que también colocar la proporción y el porcentaje de observaciones para cada intervalo de clase, por eso se tienen varios tipos de nombres:

Sobre el eje vertical	Nombre
Número de observaciones	Histograma de frecuencias
Proporción de observaciones	Histograma de frecuencias relativas
Porcentaje de observaciones	Histograma porcentual

Si los intervalos son de amplitud constante, las alturas de los rectángulos serán iguales a las frecuencias absolutas respectivas, pues al ser las bases iguales las áreas son proporcionales a las alturas; pero si las amplitudes de los intervalos son diferentes, las alturas de los rectángulos deben calcularse dividiendo la frecuencia absoluta por la longitud del intervalo; ésta se puede representar por h_i y se obtiene como:

$$h_i = \frac{n_i}{a_i}$$

La altura h_i será la frecuencia correspondiente a cada unidad de medida de la variable en cada intervalo, y se le conoce a veces, con el nombre de densidad de frecuencia del intervalo.

Ya que:

Superficie = base x altura, por lo tanto, altura = Superficie/base, correspondiendo la superficie de los rectángulos a la frecuencia.

Y de esta forma, el área del rectángulo coincide con la frecuencia:

$$S_i = h_i * a_i = \frac{n_i}{a_i} * a_i = n_i$$

El primer paso para la construcción de un histograma es la creación de una tabla de distribución de frecuencias agrupada en intervalos de los datos de la variable.

En los histogramas en el eje vertical se pueden colocar, en lugar de las frecuencias absolutas, las frecuencias relativas o los porcentajes; y de acuerdo a ello se tienen los siguientes nombres.

- **Histograma de Frecuencias:** Si en eje vertical están representadas las frecuencias absolutas o número de observaciones.
- **Histograma de Frecuencias Relativas:** Si en eje vertical están representadas las frecuencias relativas o proporción de observaciones.
- **Histograma Porcentual:** Si en eje vertical están representados los porcentajes de las observaciones.

El objetivo del histograma es mostrar el tipo de distribución de la que se trata por lo tanto siempre resulta útil atender al efecto visual de este gráfico. Como los bloques representan el área de un rectángulo cuya base es la amplitud del intervalo y cuya altura es la frecuencia correspondiente a esta clase, a efectos de no distorsionar la impresión visual, se recomienda que los intervalos tengan la misma amplitud.

Ejemplo 7: En El Salvador en año 2010, se lleva a cabo la encuesta de Hogares de propósitos múltiples, de la cual se obtuvo la información referente al total de la población por rango de edades. A continuación se presenta en la siguiente tabla.

EDAD	[0-4]	[5-9]	[10-14]	[15-19]	[20-24]	[25-29]	[30-34]	[35-39]
POBLACIÓN(n_i)	523,447	618,241	753,284	705,337	566,569	449,024	441,549	403,067

[40-44]	[45-49]	[50-54]	[55-59]	[60-64]	[65-69]	[70 y Mas[	TOTAL
334,230	294,350	234,798	214,812	167,970	154,193	320,534	6,181,405

Tabla 7: Total de la población Salvadoreña por rango de edades.

Para elaborar el histograma en el eje horizontal se colocan los intervalos o clases de la variable EDADES y en eje vertical las frecuencias absolutas.

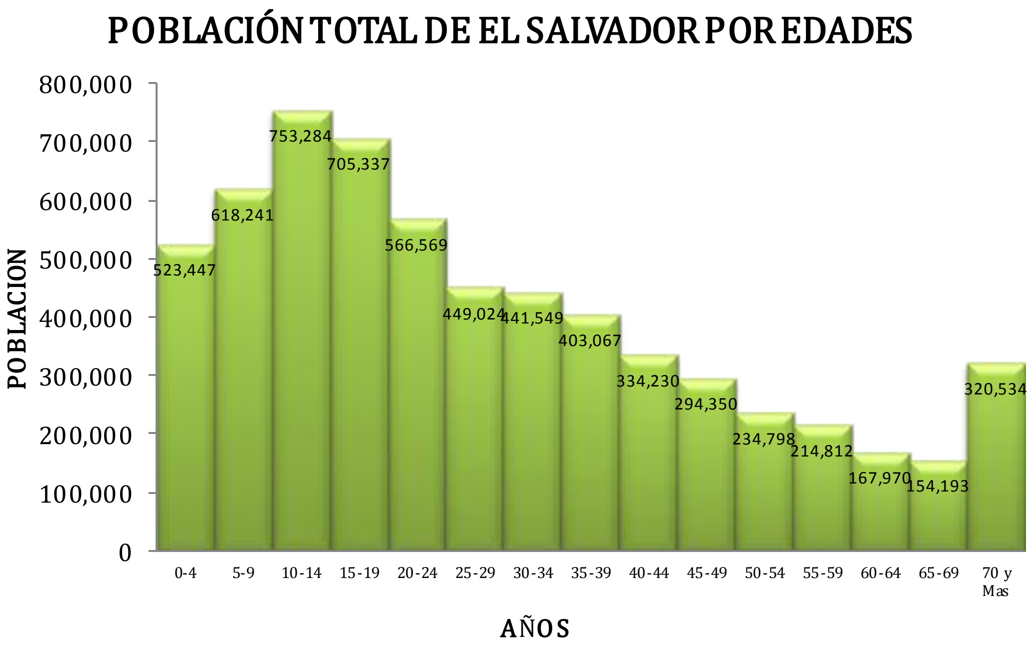


Gráfico 4: Histograma de la Población de El Salvador
Fuente: Ministerio de economía, dirección general de estadística y censos.
Encuesta de hogares de propósitos múltiples, 2010.

POLÍGONO DE FRECUENCIAS

Un polígono de frecuencias utiliza segmentos lineales conectados a puntos que se localizan directamente por encima de los valores de las marcas de clases o puntos medios del intervalo (c_i).

Las alturas de los puntos corresponden a las frecuencias de clase; en tanto que los segmentos lineales se extienden hacia la derecha y hacia la izquierda, de manera que la gráfica inicia y termina sobre el eje horizontal.

En el eje horizontal se colocan las marcas de clase de cada intervalo y para cada una de estas c_i se marcan las alturas en el eje vertical, las cuales vienen dadas por las frecuencias respectivas (absolutas o relativas). Luego, se marcan los puntos (c_i, n_i ó f_i) y se une con rectas en el para cada intervalo, y de acuerdo a ello se tienen los siguientes nombres.

- **Polígono de Frecuencias:** Si en el eje vertical están representadas las frecuencias absolutas o número de observaciones.
- **Polígono de Frecuencias Relativas:** Si en el eje vertical están representadas las frecuencias relativas o proporción de observaciones.
- **Polígono Porcentual:** Si en el eje vertical están representadas los porcentajes de las observaciones.

La diferencia con respecto al histograma es que el polígono de frecuencias sólo toma en consideración los puntos medios de las clases como representativo de cada clase o intervalo.

plano cartesiano. Para cerrar el polígono con el eje de las abscisas, se crean dos puntos ficticios, uno anterior al de la primera clase y otro posterior al de la última clase, cada uno con frecuencia igual a cero; al lado izquierdo se resta a la marca de clase inicial, la amplitud del intervalo (a_i) y al lado derecho, se suma a la marca de clase final, la amplitud del intervalo (a_i). De esta forma, el área que queda por debajo del polígono de frecuencias es igual al área contenida dentro del correspondiente histograma.

En el polígono de frecuencia como en el histograma, el valor de la variable aparece en el eje horizontal y en eje vertical la frecuencia absoluta; pero también se pueden representar la proporción y el porcentaje de observaciones

Otra alternativa de elaborar el polígono de frecuencia es a partir del histograma, marcando puntos a la mitad de cada barra en su parte superior, y uniendo estos puntos con segmentos de rectas; sin olvidar que en el eje horizontal antes del límite inferir del primer intervalo y después del último intervalo se corta el eje horizontal en los puntos ya mencionados. Se suele muchas veces representar el histograma y el polígono de frecuencia en una misma gráfica.

Ejemplo 8: El bibliotecario de un centro escolar está interesado en conocer el número de libros que sacaron en préstamos los 100 alumnos del tercer año de bachillerato a lo largo de su vida escolar en dicho centro. Los datos obtenidos se presentan a continuación.

Número de Libros prestados	Marca de clase (C_i)	Número de alumnos (n_i)	Frecuencias Relativas (fi)	Frecuencia porcentual
[35 – 39]	37	0	0	0%
[40 - 44]	42	2	0.02	2%
[45 - 49]	47	2	0.02	2%
[50 - 54]	52	8	0.08	8%
[55 - 59]	57	4	0.04	4%
[60 - 64]	62	8	0.08	8%
[65 - 69]	67	16	0.16	16%
[70 - 74]	72	16	0.16	16%
[75 - 79]	77	20	0.20	20%
[80 - 84]	82	12	0.12	12%
[85 - 89]	87	4	0.04	4%
[90 - 94]	92	6	0.06	6%
[95 - 99]	97	0	0	0%
[100 - 104]	102	2	0.02	2%
[105 - 109]	107	0	0	0%
Total		**100**	**1**	**100%**

Tabla 8: Número de Libros prestado por los alumnos.

POLÍGONO DE FRECUENCIA

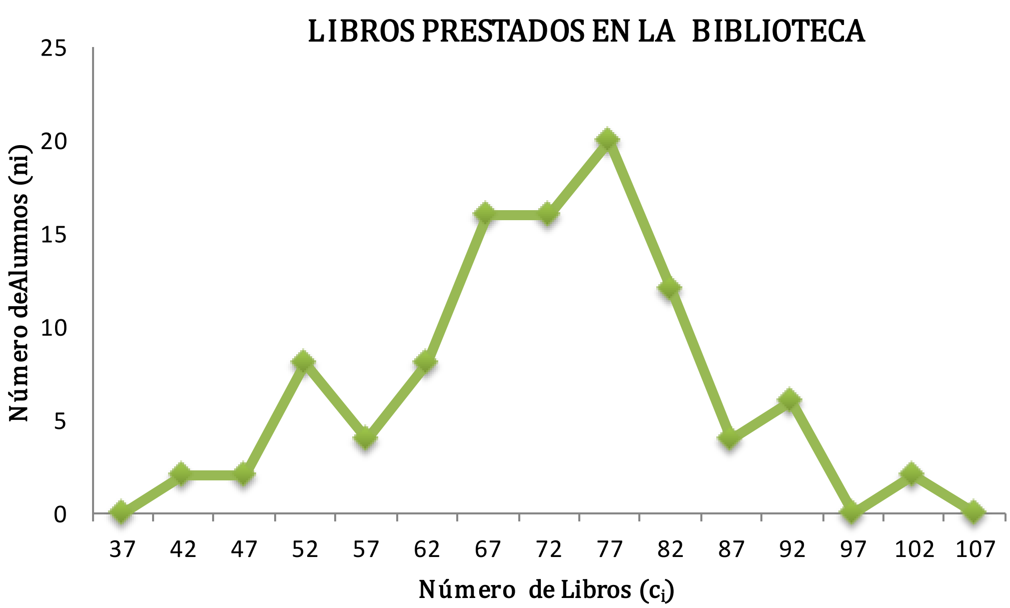


Gráfico 5: Polígono de Frecuencia de libros prestados

POLÍGONO DE FRECUENCIAS RELATIVAS

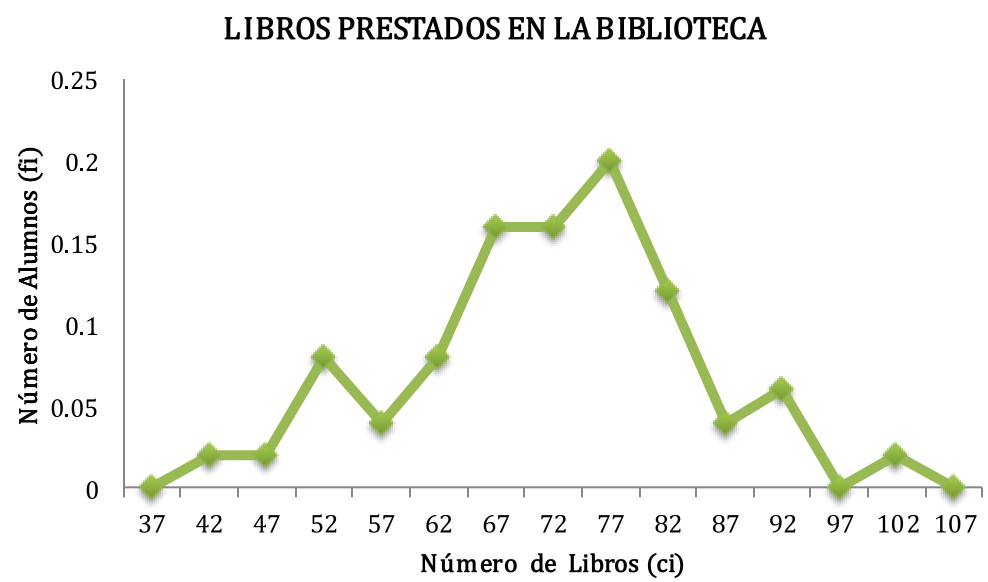


Gráfico 6: Polígono de Frecuencia Relativa de libros prestados

POLÍGONO PORCENTUAL

Gráfico 7: Polígono Porcentual de libros prestados

OJIVA

La ojiva es una gráfica similar al polígono de frecuencias y se conoce también como polígono de frecuencias acumuladas; ya que es una gráfica lineal que representa frecuencias acumuladas tanto de forma ascendente y descendente. Las frecuencias acumuladas permiten visualizar cuantas observaciones se encuentran por arriba o por debajo de ciertos valores, en lugar de limitarse a anotar los números de elementos dentro de los intervalos.

Se pueden construir ojivas "Menor que" y "Mayor que", la diferencia entre ambas gráficas es que la primera tiene pendiente positiva y crece, mientras que la segunda tiene pendiente negativa y decrece.

Una distribución de frecuencia acumulativa nos permite ver cuantas observaciones se hallan por arriba o por debajo de ciertos valores, en lugar de limitarnos a anotar los números de elementos dentro de los intervalos.

Construcción de Ojiva "Menor que"

En el eje horizontal se colocan sucesivamente los límites superiores de cada clase y en el vertical las frecuencias acumuladas ("Menor que"). Para cada límite superior de clase se marca con un punto su correspondiente frecuencia acumulada, partiendo desde el límite menor del primer intervalo que se le asigna una frecuencia igual a cero.

Las frecuencias "Menor que" se calculan a partir de la posición de la frecuencia del primer intervalo y se acumula hacia abajo hasta el último intervalo.

La frecuencia acumulada "menor que", nos muestra los valores que quedan después de un determinado dato.

Construcción de Ojiva "Mayor que"

En el eje horizontal se marca sucesivamente los límites inferiores de cada clase y en el vertical las frecuencias acumuladas ("Mayor que"). Para cada límite inferior de clase se marca con un punto su correspondiente frecuencia acumulada, culminando en el límite superior del último intervalo que se le asigna una frecuencia igual a cero.

Las frecuencias "Mayor que" se calculan a partir de la posición de la frecuencia del último intervalo y se acumula hacia arriba hasta el primer intervalo.

La frecuencia acumulada "mayor que", en cambio nos presenta los valores que se encuentran después de determinado dato.

En las Ojivas en el eje vertical se pueden colocar, en lugar de las frecuencias absolutas acumuladas, las frecuencias relativas acumuladas o los porcentajes acumulados; y de acuerdo a ello se tienen los siguientes nombres.

Ojiva: Si en eje vertical están representadas las frecuencias absolutas acumuladas o número de observaciones aculados.

- **Ojiva Relativa:** Si en eje vertical están representadas las frecuencias relativas acumuladas o proporción acumulada de las observaciones.
- **Ojiva Porcentual:** Si en eje vertical están representados los porcentajes acumulados de las observaciones.

Características de las Ojivas

- Muestran frecuencias acumuladas.
- Se prefiere para el tratamiento de datos cuantitativos.
- El punto de inicio equivale a una frecuencia de 0.

- Suelen utilizarse para representar tablas de distribuciones de datos agrupados.
- El punto final equivale al 100% de los datos.

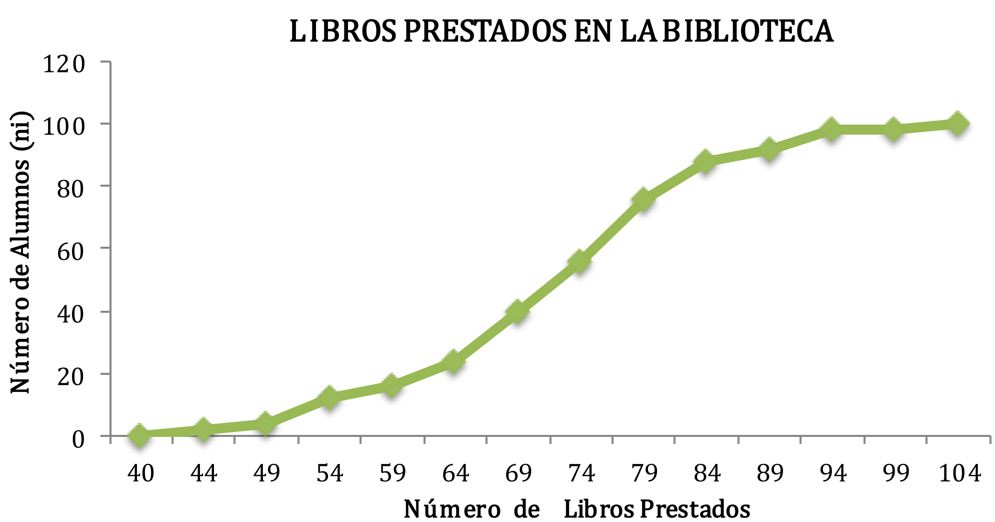
Diferencias fundamentales entre las ojivas y los polígonos de frecuencias.

- Un extremo de la ojiva no se "amarra" al eje horizontal, para la ojiva mayor que sucede con el extremo izquierdo; para la ojiva menor que, con el derecho.
- En el eje horizontal en lugar de colocar las marcas de clase se colocan las fronteras de clase. Para el caso de la ojiva mayor que es la frontera menor; para la ojiva menor que, la mayor.

Ejemplo 9: Retomando el ejemplo 8.

Número de Libros prestados	Número de alumnos n_i	Menor que n_i ↓	Mayor que n_i ↑
[40 - 44]	2	2	100
[45 - 49]	2	4	98
[50 - 54]	8	12	96
[55 - 59]	4	16	88
[60 - 64]	8	24	84
[65 - 69]	16	40	76
[70 - 74]	16	56	60
[75 - 79]	20	76	44
[80 - 84]	12	88	24
[85 - 89]	4	92	12
[90 - 94]	6	98	8
[95 - 99]	0	98	2
[100 - 104]	2	100	2
Total	**100**		

OJIVA "MENOR QUE"



Interpretación

Gráfico 8 : Ojiva "menor que" de libros prestados

El valor de 84 en el eje de las abscisas tiene un valor de 88, lo que indica que 88 alumnos prestaron una cantidad menor de 84 libros.

OJIVA "MAYOR QUE"

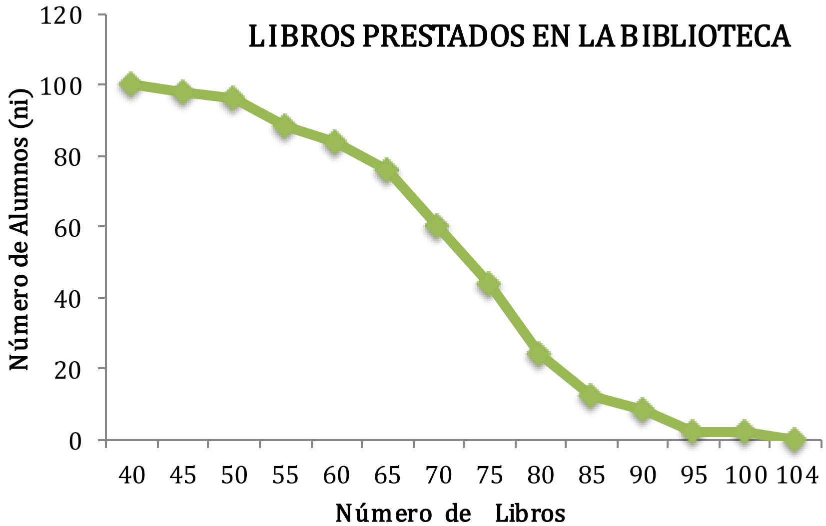


Gráfico 9: Ojiva "Mayor que" de libros prestados

ACTIVIDADES DE CONSOLIDACIÓN (GRUPALES)

1) En un jardín infantil se tomó una muestra de 14 niños para estudiar la relación entre la edad y el grado de adquisición del lenguaje. Como un primer paso para la investigación se desea conocer la distribución de las edades de la muestra (edades que se muestran en la tabla), para lo cual se elaboró el siguiente gráfico.

Niño	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Edad	3.6	1.2	2.1	1.2	1.5	3.6	3.6	1.2	2.1	3.6	3.2	1.5	2.0	1.2

Evalué críticamente el gráfico utilizado de acuerdo a lo que se quiere ilustrar en el estudio.

Respuesta:

La gráfica es errónea, ya que los valores graficados corresponden al número de niño, que sólo identifica a los participantes y no expresa ningún valor de un atributo o variable. Además, si la intención del investigador era graficar la distribución de edad, debió utilizar un gráfico de barras.

2) Aplica una encuesta con los 40 estudiantes de tu Instituto con respecto al medio de transporte utilizado con mayor frecuencia para trasladarse al instituto. Con los datos obtenidos elabora:
   a) Una distribución de frecuencias absolutas.
   b) Obtenga el gráfico circular,
   c) Gráfico de barras.
   d) Has tres conclusiones de tabla y gráficos obtenidos.
3) Recopila las estaturas (en metros) de todos los compañeros de tu sección de clase y reúne estos datos en una tabla de distribución de frecuencia. Haciendo su gráfico correspondiente y cuatros conclusiones de los resultados.

APLICANDO LO APRENDIDO

1. Los valores el pH sanguíneo en 80 pacientes del Hospital Rosales reflejaron los siguientes datos:

7.33 7.32 7.34 7.40 7.28 7.29 7.35 7.33 7.34 7.28 7.31 7.35 7.32 7.33 7.33 7.36

7.32 7.31 7.35 7.36 7.26 7.39 7.29 7.32 7.34 7.30 7.34 732 7. 39 7.30 7.33 7.33

7.35 7.34 7.33 7.36 7.33 7.35 7.31 7.33 7.37 7.38 7.38 7.33 7.35 7.30 7.31 7.33

7.35 7.33 7.27 7.33 7.32 7.31 7.34 7.32 7.32 7.32 7.31 7.36 7.30 7.37 7.33 7. 32

7.31 7.33 7.32 7.30 7.29 7.38 7.33 7.35 7.32 7.33 7.32 7.34 7.32 7.34 7.32 7.33

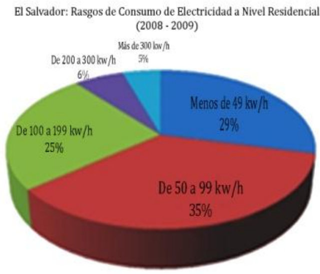
   a) Formar la tabla de frecuencias utilizando 15 intervalos de clase.
   b) Construir el histograma de frecuencias.
   c) Polígono de frecuencias.
   d) Construir el histograma de frecuencias acumuladas.
   e) Construir el polígono de frecuencias acumuladas.

2) En un Súper Selectos de San Salvador, se examinó un lote de 25 cajas de manzanas, cada una teniendo un contenido de 48 manzanas. El número de manzanas en mal estado que se encontraron en cada caja fue: 3 , 4 , 1 , 2 ,1 , 2 , 5 , 2 , 1 , 2 , 3 , 0 , 1 , 0 , 3 , 3 , 2 , 0 , 2 , 1 , 3 , 4 , 1 , 2 , 2.

Determine variable, tipo, población, muestra. Confecciona una Tabla de Frecuencias y en base a ella responder:

   a) ¿Cuántas cajas contienen menos de 3 manzanas en mal estado? (17 cajas)
   b) ¿Qué porcentaje de cajas contienen al menos 3 manzanas en mal estado? (32%)
   c) ¿Cuántas cajas contienen de 2 a 4 manzanas en mal estado? (15 cajas)
   d) ¿Qué porcentaje de cajas contienen a lo más 2 manzanas en mal estado? (68%)

3) En El Salvador en los años 2007 y 2008, del total de usuarios residenciales, el 64% consume menos de 100 kwh/mes. De acuerdo a datos de la SIGET, los datos se encuentran se ubicados en los siguientes rangos de consumo:



Fuente: Elaboración propia en base a los boletines estadísticos de la Superintendencia General de Electricidad y Telecomunicaciones (SIGET), 2007 y avance 2008

De acuerdo a la gráfica:

a) ¿Cuál es la población total de usuarios residenciales de El Salvador?
b) ¿Cuántos usuarios residenciales consumen menos de 49 kwh?
c) ¿Cuántos usuarios residenciales consumen más de 300 kwh?
d) Si el subsidio del gas se lo dan a los que consumen hasta 300 kwh. ¿A cuántos usuarios residenciales se les proporciona el subsidio del gas?

4) En la frontera de El Salvador y Guatemala se tomó el peso (en toneladas) de los Furgones que llegaron durante el mes de Octubre de 2009, obteniéndose los siguientes datos.

10.5, 12.0, 15.0, 12.3, 12.1, 14.3, 10.7, 13.0, 13.8, 13.5 , 11.2 , 11.8 , 11, 4 , 12.5 , 14.3 , 14.7 , 12.1 , 14.7 , 10.8 , 12.3 , 14.8 , 14.5 , 14.0 , 13.9 , 11.5 , 12.0 , 14.0 14.1 , 13.8 , 13,2 , 12,5 , 10.,8 , 12.9 , 14.0 , 10.2 , 12.5 , 10.6 , 11.2 , 14.6 , 13.0

a) Determine variable, tipo, población y muestra
b) Organizar los datos en una Tabla de Frecuencias con 4 intervalos.
c) Interpretar los valores que corresponden a N_1 , f_4 , N_2 y F_3 .

BIBLIOGRAFIA

Johnson R., Cuby P.(1999), *Estadística Elemental*. México: Internacional Thomson Editores, S.A de S.V.

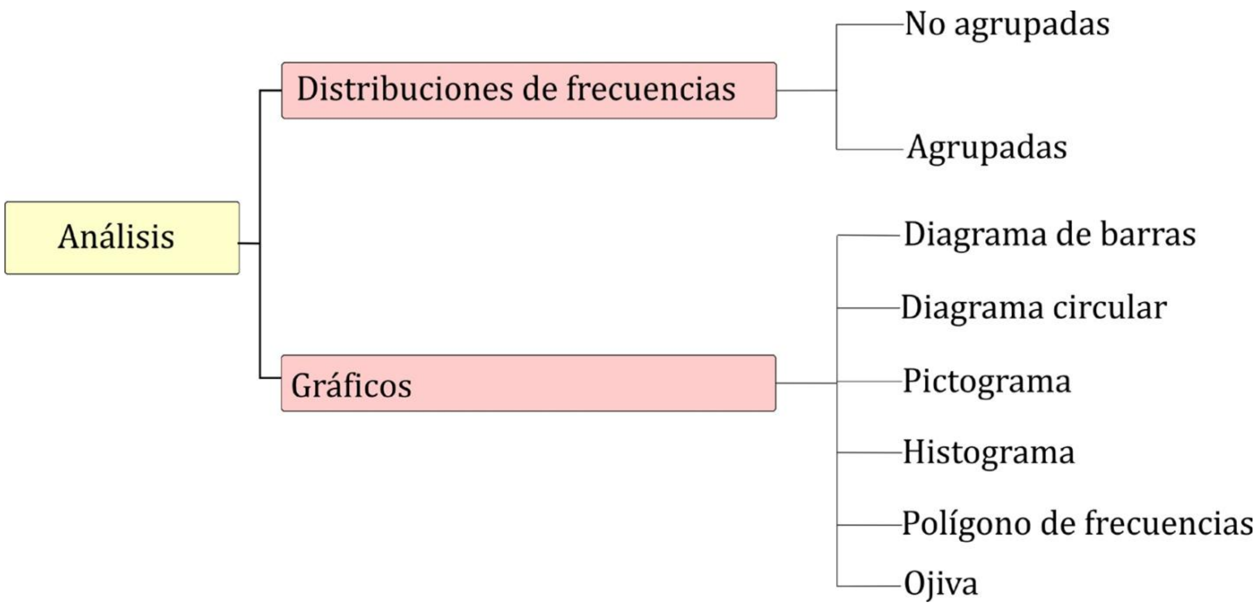
Pérez C., (2003).*Estadística. Problemas Resueltos y Aplicaciones*. Madrid: Pearson Educación, S.A.

Pérez-T. H.E ((2007), Estadística para las Ciencias Sociales, del Comportamiento y de la Salud. (3ª. Ed).México: Impreso Edamsa Impresiones, S.A. de C.V.

Sarabia J.M. (2000), *Curso Práctico de Estadística*.(2da ed.) . España: Impreso por Gráficas Rogar, S.A Navalcarnero (Madrid).

Triola, M., (2009). *Estadística.* (10a ed.). México: Pearson Educación.

DIAGRAMA DE CONTENIDOS



Lección 3 **Primer año de Bachillerato** **Unidad V** **Tiempo: 8 horas clase**

Medidas de tendencia central

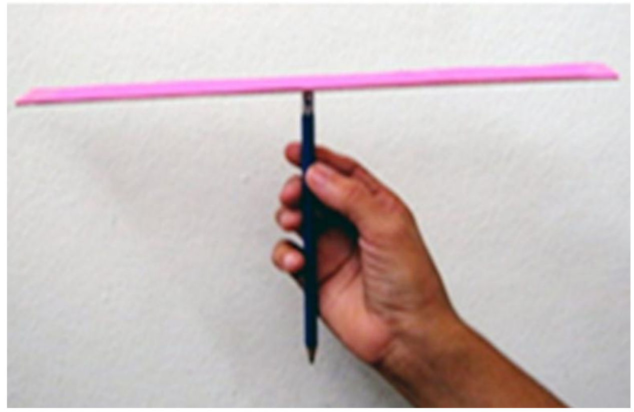


Figura 1. Punto de equilibrio de una regla.

Introducción del Tema

Al tener una colección de datos del tipo numérico; sean estos de una población o de una muestra, en lugar de manejar todos los datos representados o no en una distribución de frecuencias se pueden caracterizar mediante las medidas de tendencia central también llamadas medidas de posición las cuales se utilizan para representar con un solo número todo un conjunto de datos; es decir son valores numéricos que localizan, de alguna manera, el centro de un conjunto de datos.

Por tendencia central se entiende un valor que representa al conjunto de valores de la distribución de una variable. En el caso extremo de una distribución en la que todos los sujetos tuvieran el mismo valor, este dato daría cuenta de todos ellos. Pero, como su propio nombre indica, las variables se caracterizan por no presentar valores únicos. Por ello, hay varios procedimientos para obtener una medida de tendencia central. Los más empleados son la media, la moda y la mediana.

Descripción

Se estudian las principales medidas de tendencia central como son: Media, Mediana y Moda. Las fórmulas de los cálculos se especifican dependiendo como se tengan organizados los datos, los cuales pueden estar de manera simple o agrupada en tablas de distribuciones de frecuencias en intervalos o no.

Objetivos

- Calcular e interpretar las medidas estadísticas de centralización más importantes.
- Identificar las propiedades de las medidas de tendencia central.
- Calcular las medidas de tendencia central de datos no agrupados y agrupados de problemas propuestos utilizando las herramientas tecnológicas digitales como recursos didácticos.

Importancia

En la vida cotidiana la herramienta estadística más utilizada son las medidas de tendencia central, ¿quién no ha hablado sobre promedio de notas, gastos diarios, gasto de transporte, alimentación, educación, de producción por hectárea, humedad, temperatura, goles, nacimientos, muertes, delitos, heridos, accidentes?; y sobre estos resultados se toman decisiones ya sean de control, distribución e inversión, según el caso, demostrándose que la estadística es de gran importancia para la toma de decisiones.

Las medidas de tendencia central (media, mediana y moda) sirven como puntos de referencia para interpretar las calificaciones que se obtienen en una prueba. Supongamos por ejemplo que la calificación promedio de la prueba que hizo Pedro fue de 20 puntos. De ser así podemos decir que la calificación de Pedro se ubica notablemente sobre el promedio. Pero si la calificación promedio fue de 60 puntos, entonces la conclusión sería muy diferente, dado que se ubicaría muy por debajo del promedio de la clase.

Estas medidas de tendencia central se definen como un indicador de localización central empleado en la descripción de las distribuciones de frecuencia. Una distribución de frecuencia representa una organización de datos pero no nos permite por si misma establecer proposiciones cuantitativas, ya sea describiendo la distribución o comparando dos o más distribuciones.

Las medidas de tendencia central son como el centro de gravedad de los cuerpos. En Dinámica, describir el movimiento del centro de gravedad, equivale a describir el movimiento total del cuerpo. Si la línea de acción del centro de gravedad, pasa por el punto de apoyo, el cuerpo se encuentra en equilibrio.

Competencias a reforzar.

Calcula, analiza e interpreta las medidas de tendencia central, para tomar decisiones acertadas ante una situación real.

Presaberes

Diferencia entre variable cualitativa y cuantitativa.
Conocimiento de las operaciones básicas.

MEDIA ARITMÉTICA

La media aritmética es el valor medio de todos los valores que toma la variable estadística de una serie de datos. Por lo tanto, se considera como la medida posicional más utilizada en los estudios estadísticos; por su fácil cálculo e interpretación, es la medida de posición más conocida. La media es el valor más representativo de la serie de valores, es el punto de equilibrio, es el centro de gravedad de la serie de datos.

Si se obtienen las calificaciones de 14 alumnos de primer año de bachillerato de la asignatura de Matemática: 0, 1.75, 3, 4.25, 5, 5, 5, 6, 6, 6, 6, 7, 8.5, 10.

La media aritmética no es otra cosa que el centro de gravedad o punto de equilibrio de la distribución.

En la siguiente figura, cada nota está representada por una bola, donde el tamaño es proporcional al número de veces que se repite cada dato. Por ejemplo, el número 6 se repite más veces (cuatro) por lo tanto, estará representado por la bola más grande.

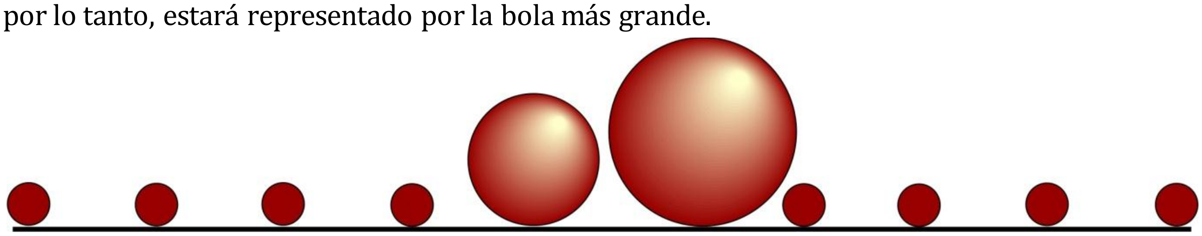


Figura 2. Representación de los datos en una línea recta

Si el grosor de cada bola es representado en términos de peso, el centro de gravedad o punto de equilibrio de la distribución de los datos se ubica aproximadamente donde hemos trazado la línea vertical.

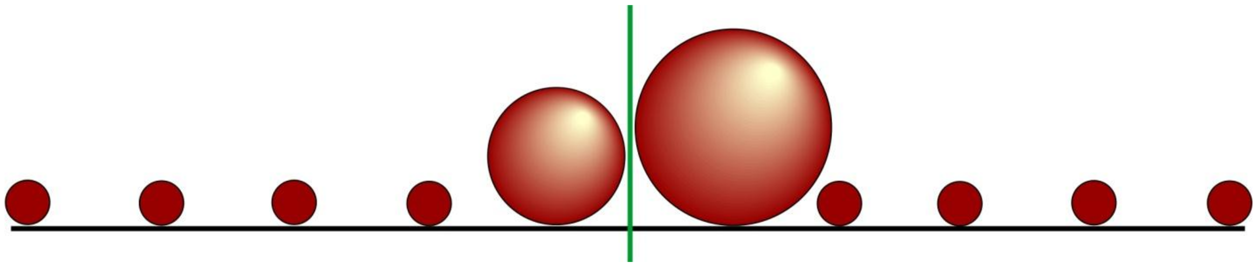


Figura 3. Posición del centro de gravedad de los datos.

Con lo anterior, el centro de gravedad de una distribución de datos, debe ser aquel valor de la variable que equilibre la distribución, entendiendo este equilibrio en el sentido de que las desviaciones positivas y negativas con respecto a ella deben sumar cero.

Ahora, si nos imaginamos el diagrama de barras o el histograma de frecuencias de la distribución de los datos apoyado en un punto del eje horizontal de tal forma que quedase en equilibrio, el valor de este punto en dicho eje sería el valor de la media aritmética, que es el centro de gravedad de la distribución estadística.

De acuerdo a lo planteado anteriormente, la forma del diagrama de barras o del histograma nos permite calcular "a ojo", con bastante aproximación, el valor de la media aritmética de los datos representados. La media aritmética coincide con el "punto de equilibrio" del gráfico o, dicho de otro modo, está en la vertical que pasa por su centro de gravedad.

Imagina por un momento que el gráfico es un objeto con masa y quisiéramos colocarlo, en equilibrio, sobre un punto del eje horizontal: el punto de apoyo ha de estar situado en la media aritmética de los datos representados.

Si ese punto de apoyo estuviera situado a la izquierda o a la derecha de la media, el gráfico se desequilibraría, hacia un lado o hacia el otro.

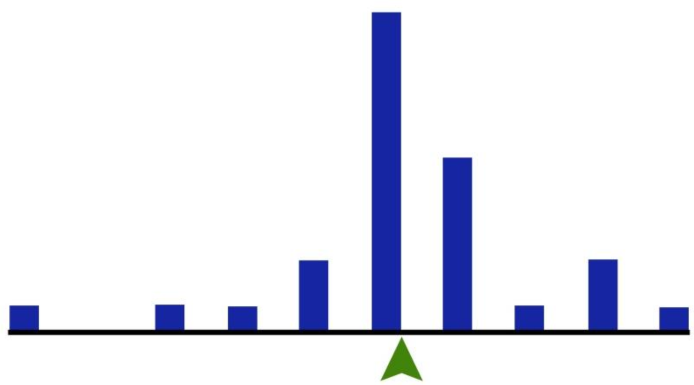


Figura 4. Centro de gravedad o punto de equilibrio de un diagrama de barras.

Notación y Cálculo

La media aritmética de n observaciones de la variable X se denotará por el símbolo: $\overline{X}$ que se lee como:

"x barra" o **"media de la muestra"**.

La fórmula para su cálculo depende cómo se tengan organizados los datos; ya que pueden estar sin agrupar o agrupados en una tabla de distribución de frecuencias, que pueden ser en intervalos o no.

Caso 1: Datos sin agrupar

La media aritmética se define como la suma de todos los valores de la distribución (variable X) dividida por el número total de datos, n. Lo anterior se expresa con una fórmula como:

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + x_3 + \ldots + x_{n-1} + x_n}{n}$$ **Ecuación 1**

Donde

x_i : i-ésimo elemento de la muestra.
n : Número total de observaciones.
$\overline{X}$: Media de la muestra.

Caso 2: Datos agrupados en tablas de frecuencias

Si el valor X_i de la variable X se repite n_i veces, la expresión de la media aritmética es de la forma:

$$\overline{X} = \sum_{i=1}^{k} \frac{x_i n_i}{n}$$ **Ecuación 2**

Donde:

x_i : i-ésimo elemento de la muestra.
n : Número total de observaciones.
n_i : Frecuencia absoluta
k : Cantidad de valores que toma x_i
$\overline{X}$: Media de la muestra.

Como $f_i = \frac{n_i}{n}$, una expresión equivalente para el cálculo de media será $\overline{X} = \sum_{i=1}^{k} x_i f_i$; donde f_i es la frecuencia relativa de los valores de la variable.

Caso 3: Datos agrupados en intervalos.

En el caso que se tuviera una distribución con datos agrupados en intervalos, los valores individuales de la variable serían desconocidos y, por tanto, no se podrían utilizar las fórmulas anteriores. En este supuesto los datos estarán agrupados en clases, y se postula la hipótesis de que el punto medio del intervalo de clase (c_i marca de clase) representa adecuadamente el valor medio de dicha clase, y aplicaríamos la fórmula siguiente:

$$\overline{X} = \sum_{i=1}^{k} \frac{c_i n_i}{n}$$ **Ecuación 3**

Igual que el caso anterior como $f_i = \frac{n_i}{n}$ entonces $\overline{X} = \sum_{i=1}^{k} c_i f_i$ con $c_i = \frac{\mathbf{L_{i-1}} - \mathbf{L_i}}{2}$

Donde

c_i : Marca de clase i-ésima.
L_{i-1}: Límite inferior de clase
L_i : Límite superior de clase
n : Número total de observaciones.
n_i : Frecuencia absoluta de la clase i-ésima.
k : Número de clases
f_i : Frecuencia relativa de la clase i-ésima.
$\overline{X}$: Media de la muestra.

Ejemplos de aplicación

Ejemplo 1: Supongamos que en un almacén tienen empleados a 12 vendedores, y sus ingresos mensuales son: \$ 585, \$ 521, \$ 656, \$ 465, \$ 536, \$ 487, \$ 564, \$ 490, \$ 563, \$ 1234, \$ 469 y \$ 547. **Se pide:**

a) Determinar la media de los ingresos de los 12 vendedores.

b) ¿Cuántos vendedores ganan más que el promedio?

Solución

a) Cálculo de la media

Por la forma en que se presentan los datos se utilizará la **Ecuación 1**:

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + x_3 + \ldots + x_{n-1} + x_n}{n}$$

La cantidad de observaciones son 12 (n=12), por lo tanto: X_1= 585, X_2= 521, X_3= 656, X_4= 465, X_5= 536, X_6=487, X_7= 564, X_8= 490 X_9= 563, X_{10}= 1234, X_{11}= 469, X_{12}= 547

Al sustituir queda:

$$\overline{X} = \frac{\sum_{i=1}^{12} x_i}{12} = \frac{x_1 + x_2 + x_3 + \ldots + x_{11} + x_{12}}{12}$$

$$= \frac{585 + 521 + 656 + \ldots + 469 + 547}{12} = \frac{7117}{12}$$

$$\overline{X} = \$\ 593.08$$

Interpretación del resultado: Los ingresos promedios de los 12 vendedores del almacén son $ 593.08.

b) Para determinar el número de vendedores que ganan más que el promedio obtenido en a) se debe ordenar los datos de menor a mayor y contar cuantos valores de sueldos son mayores que $ 593.08.

$ 465, $ 469 , $ 487 , $ 490 , $ 521 , $ 536 , $ 547 , $ 563, $ 564 , $ 585, **$ 656 , $ 1234**

Como puede observarse solo son dos sueldos que superan al promedio, los cuales aparecen en negrito.

Ejemplo 2: Los resultados obtenidos de las calificaciones de 50 alumnos de la asignatura de ciencias se tienen organizados en una tabla de distribución de frecuencia como se muestra a continuación.

x_i	n_i	N_i	f_i	F_i
0	1	1	0.02	0.02
1	1	2	0.02	0.04
2	2	4	0.04	0.08
3	3	7	0.06	0.14
4	6	13	0.12	0.26
5	11	24	0.22	0.48
6	12	36	0.24	0.72
7	7	43	0.14	0.86
8	4	47	0.08	0.94
9	2	49	0.04	0.98
10	1	50	0.02	1.00
n	50		1.00	

a) Encontrar la calificación media de este grupo de estudiantes.

b) Determinar la cantidad aproximada de alumnos que tienen una calificación en ciencias menor que la media.

Solución.

a) Cálculo de la media

Por la forma en que se presentan los datos se utilizará la **Ecuación 2**:

$$\overline{X} = \sum_{i=1}^{n} \frac{x_i n_i}{n}$$

La cantidad de observaciones son 50 (n=50), por lo tanto al sustituir los valores queda:

$$\overline{X} = \frac{\sum_{i=1}^{11} x_i * n_i}{50} = \frac{x_1 * n_1 + x_2 * n_2 + x_3 * n_3 + \ldots x_{10} * n_{10} + x_{11} * n_{11}}{50}$$

$$= \frac{0*1 + 1*1 + 2*2 + \ldots + 9*2 + 10*1}{50} = \frac{275}{50}$$

$$\overline{X} = 5.5$$

Interpretación del resultado: La calificación media de la asignatura de ciencias del grupo de 50 alumnos es 5.5

b) Para determinar la cantidad aproximada de alumnos que tienen una calificación menor que el promedio, se observa en la columna N_i de la tabla que corresponde a las frecuencias absolutas acumuladas, y el valor que corresponde a un x_i menor que 5.5; en este caso es N_6 =24. Por lo tanto, hay 24 alumnos que obtuvieron una calificación menor que el promedio.

Ejemplo 3: Calcular la media aritmética (vida media) de la siguiente distribución de frecuencia del número de meses de duración de una muestra de 40 baterías para vehículo.

$[L_{i-1}-L_i]$	c_i	n_i	N_i	f_i	F_i
[15-19]	17	2	2	0.05	0.05
[20-24]	22	1	3	0.025	0.075
[25-29]	27	4	7	0.1	0.175
[30-34]	32	15	22	0.375	0.55
[35-39]	37	10	32	0.25	0.8
[40-44]	42	5	37	0.125	0.925
[45-49]	47	3	40	0.075	1.00
n		40		1.00	

Solución

La variable X en estudio es la duración en meses de las baterías de vehículo.

Por la forma en que se presentan los datos se utilizará la **Ecuación 3**:

$$\overline{X} = \sum_{i=1}^{k} \frac{c_i n_i}{n}$$

Se tiene que la cantidad de observaciones son 40 (n=40), por lo tanto:

$$\overline{X} = \sum_{i=1}^{7} \frac{c_i n_i}{40} = \frac{c_1 * n_1 + c_2 * n_2 + c_3 * n_3 + ... + c_6 * n_6 + c_7 * n_7}{40}$$

$$= \frac{17*2 + 22*1 + 27*4 + ... + 42*5 + 47*3}{40} = \frac{1365}{40}$$

$\overline{X} = 34.12$ meses

Interpretación del resultado: La vida media de las 40 baterías es de 34.12 meses; es decir que las 40 baterías duran en promedio 34.12 meses.

Características de la Media Aritmética

1. En su cálculo se toman en cuenta todos los valores de la variable.
2. La media aritmética es el parámetro de centralización más utilizado.
3. Es una medida totalmente numérica o sea sólo puede calcularse en datos de características cuantitativas.
4. No puede ser calculada en distribuciones de frecuencia que tengan clases abiertas.
5. Es única, o sea, un conjunto de datos numéricos tiene una y solo una media aritmética.
6. La media aritmética viene expresada en las mismas unidades que la variable.
7. Es el centro de gravedad de toda la distribución, representando a todos los valores observados.
8. Su principal inconveniente es que se ve afectada por los valores extremadamente grandes o pequeños de la distribución y que, por consiguiente, puede estar muy lejos de ser una representación de la muestra, por lo que no es recomendable usarla en distribuciones muy asimétricas.

Propiedades de la media aritmética

Dada la importancia de la media y su uso frecuente, conviene considerar algunas de sus principales propiedades matemáticas:

Propiedad 1: La suma de las desviaciones de los valores de la variable respecto a la media vale cero, simbólicamente se expresa de la manera siguiente:

Para datos no agrupados: $\sum_{i=1}^{n}(x_i - \overline{x}) = 0$

Para datos agrupados: $\sum_{i=1}^{n}(x_i - \overline{x})n_i = 0$

La demostración de esta propiedad es como sigue:

$$\sum_{i=1}^{n}(x_i - \overline{x})n_i = \sum_{i=1}^{n} x_i n_i - \overline{x}\sum_{i=1}^{n} n_i$$

$$= \sum_{i=1}^{n} x_i n_i - N\overline{x} = \sum_{i=1}^{n} x_i n_i - \sum_{i=1}^{n} x_i n_i = 0$$

Ejemplo 4: Los pesos de seis amigos son: 84, 91, 72, 68, 87 y 78 kg. Hallar el peso medio y verificar que las desviaciones respecto a la media es igual a cero.

Solución

Media aritmética

$$\overline{X} = \frac{\sum_{i=1}^{6} x_i}{6} = \frac{84+91+72+68+87+78}{6} = \frac{480}{6}\ \overline{X} = 80$$

Kg.

La suma de las desviaciones es:

$$\sum_{i=1}^{6}(x_i - \overline{x}) = (84-80)+(91-80)+(72-80)$$
$$+(68-80)+(87-80)+(78-80)$$
$$= 4 + 11 - 8 - 12 + 7 - 2 = 22 - 22 = 0$$

Por lo tanto se cumple que $\sum_{i=1}^{n}(x_i - \overline{x}) = 0$

Propiedad 2: La media aritmética de un valor constante, es igual a la misma constante.

Demostración

Si cada uno de los valores observados X_1, X_2, X_3……X_n es igual a una constante c, entonces

$$\overline{X} = \frac{x_1 + x_2 + x_3 + \ldots + x_{n-1} + x_n}{n} = \frac{c + c + \ldots + c}{n} = \frac{nc}{n} = c$$

Ejemplo 5: A un restaurante asisten cinco familias con el objetivo de celebrar los quince años de sus hijos mayores. Cuál es la edad media de dichos jóvenes.

Solución

Como son cinco familias y cada una de ellas tiene un hijo mayor, con quince años de edad entonces se tienen cinco valores iguales que son: 15, 15, 15, 15, 15. Valor constante para cada uno de ellos.

Entonces la edad media es:

$$\overline{X} = \frac{15+15+15+15+15}{5} = \frac{75}{5} = 15 \text{ años.}$$

Propiedad 3: Relación de la media aritmética con las operaciones básicas.

a) Si le sumamos a todas las observaciones un mismo número, la media aumentará en dicha número.

Demostración

Si los valores observados son: $x_1, x_2, \ldots, x_{n-1}, x_n$

Al sumarle la constante C, a cada uno de ellos, resulta: $x_1+c, x_2+c, \ldots, x_{n-1}+c, x_n+c$

La media aritmética de estos nuevos datos es:

$$\overline{X+c} = \frac{(x_1+c)+(x_2+c)+(x_3+c)+\ldots+(x_{n-1}+c)+(x_n+c)}{n}$$
$$= \frac{x_1 + x_2 + x_3 + \ldots + x_{n-1} + x_n + c + c + \ldots + c}{n}$$
$$= \frac{x_1 + x_2 + x_3 + \ldots + x_{n-1} + x_n}{n} + \frac{c + c + \ldots + c}{n}$$
$$= \overline{X} + \frac{nc}{n} \quad = \overline{X} + c$$

Ejemplo 6: Retomando el ejemplo 5 de los pesos de los amigos. La media aritmética obtenida fue 80 kg.

Si a estos valores se les suma la constante c=2; los nuevos datos serán: 86, 93, 74, 70, 89 y 80.

La nueva media aritmética según propiedad será: 80+2 = 82 kg.

Comprobando:

$$\overline{X} = \frac{\sum_{i=1}^{6} x_i}{6} = \frac{86+93+74+70+89+80}{6} = \frac{492}{6}$$

$$\overline{X} = 82\ Kg$$

b) Si le restamos a todas las observaciones un mismo número, la media aritmética queda disminuida en dicha cantidad.

Demostración

Si los valores observados son: $x_1, x_2, \ldots, x_{n-1}, x_n$

Al restar la constante *c* a cada uno de ellos, resulta: $x_1-c, x_2-c, \ldots, x_{n-1}-c, x_n+c$

La media aritmética de estos nuevos datos es:

$$\overline{X-c} = \frac{(x_1-c)+(x_2-c)+(x_3-c)+\ldots+(x_{n-1}-c)+(x_n-c)}{n}$$

$$= \frac{x_1+x_2+x_3+\ldots+x_{n-1}+x_n-(c+c+\ldots+c)}{n}$$

$$= \frac{x_1+x_2+x_3+\ldots+x_{n-1}+x_n}{n} - \frac{c+c+\ldots+c}{n}$$

$$= \overline{X} - \frac{nc}{n} \quad = \overline{X} - c$$

Ejemplo 7: Retomando el ejemplo 5 de los pesos de los amigos.

La media aritmética obtenida fue 80 kg.

Si a cada valor se le resta la constante c=3; los nuevos datos serán: 81, 88, 69, 65, 84 y 75 kg.

La nueva media aritmética según propiedad será: 80-3 = 77 kg.

Comprobando:

$$\overline{X} = \frac{\sum_{i=1}^{6} x_i}{6} = \frac{81+88+69+65+84+75}{6} = \frac{462}{6}$$

$$\overline{X} = 77\ Kg$$

b) Si multiplicamos todas las observaciones por un número constante, la media queda multiplicada por dicho número.

Demostración

Si los valores observados son: $x_1, x_2, \ldots, x_{n-1}, x_n$

Al multiplicar la constante *c* a cada uno de ellos, resulta: $x_1*c, x_2*c, \ldots, x_{n-1}*c, x_n*c$

La media aritmética de estos nuevos datos es:

$$\overline{X*c} = \frac{x_1*c+x_2*c+x_3*c+\ldots+x_{n-1}*c+x_n*c}{n}$$

$$= \frac{(x_1+x_2+x_3+\ldots+x_{n-1}+x_n)}{n} * c = \overline{X} * c$$

Ejemplo 8: Retomando el ejemplo 5 de los pesos de los amigos.

Si cada valor se multiplica por una constante c=2 se tienen los nuevos valores: 168, 182, 144, 136, 174 y 156 kg.

La media aritmética obtenida es: 80 kg.

La nueva media aritmética según propiedad será: 2*80 = 160 kg.

Comprobando:

$$\overline{X} = \frac{\sum_{i=1}^{6} 2x_i}{6} = \frac{168+182+144+136+174+156}{6} = \frac{960}{6}$$

$$\overline{X} = 160\ Kg.$$

d) Si dividimos todas las observaciones por un número constante, la media queda dividido por dicho número.

Si los valores observados son: $x_1, x_2, \ldots, x_{n-1}, x_n$

Al dividir la constante *c* a cada uno de ellos, resulta: $\frac{x_1}{c}, \frac{x_2}{c}\ \frac{x_3}{c}, \ldots, \frac{x_{n-1}}{c}, \frac{x_n}{c}$

La media aritmética de estos nuevos datos es:

$$\overline{\left(\frac{X}{c}\right)} = \frac{\frac{x_1}{c}+\frac{x_2}{c}+\frac{x_3}{c}+\ldots+\frac{x_{n-1}}{c}+\frac{x_n}{c}}{n}$$

$$= \frac{x_1}{cn}+\frac{x_2}{cn}+\frac{x_1}{cn}+\ldots+\frac{x_{n-1}}{cn}+\frac{x_n}{cn}$$

$$=\frac{1}{c}\left(\frac{x_1+x_2+x_2+....+x_{n-1}+x_n}{n}\right)$$

$$=\frac{1}{c}\overline{X} \quad =\frac{\overline{X}}{c}$$

Ejemplo 9: Retomando el ejemplo 5 de los pesos de los amigos.

Si cada valor se divide por una constante c igual a 2 se tienen los nuevos valores: 42, 45.5, 36, 34, 43.5, 39 kg.

La media aritmética obtenida fue 80 kg.

La nueva media aritmética será: $\overline{X}=\frac{80}{2}=40$ kg.

Comprobando:

$$\overline{X}=\frac{\sum_{i=1}^{6}\frac{x_i}{2}}{6}=\frac{42+45.5+36+34+43.5+39}{6}=\frac{240}{6}$$

$\overline{X}=40\ Kg.$

Propiedad 4: La media aritmética de una muestra dividida en submuestras, es igual, a la media ponderada de las submuestras, tomando como ponderación los tamaños de las submuestras. Esto es, $\overline{x}=\frac{\sum_{i=1}^{m}\overline{x}_i n_i}{n}$,

Donde n= $n=n_1+n_2+\ldots+n_m$, $\overline{x}_i$ la media de cada submuestra *y* n_i la cantidad de elementos de cada submuestra.

Demostración.

Sea la distribución $x_1, x_2, x_3, x_4,\ldots\ldots x_n, x_{n+1}, x_{n+2}\ldots\ldots\ldots x_k$, observando que habrían como dos submuestras de n y k-n elementos cada uno.

Si consideramos la media aritmética de la distribución: $\overline{X}=\frac{\sum\overline{x}_i n_i}{n}$ y calculamos los sumatorios para los dos subconjuntos, la expresión de la media quedaría:

$$\overline{X}=\frac{\sum_{j=1}^{n}x_j n_j+\sum_{r=n+1}^{k}x_r n_r}{n}=\frac{\sum_{j=1}^{n}x_j n_j}{n}+\frac{\sum_{r=n+1}^{k}x_r n_r}{n}$$

Si se multiplica el numerador y denominador de cada una de las fracciones por una misma cantidad el resultado no varía, por tanto, multiplicaremos la primera por n_1 que es su número de elementos del primer subconjunto y la segunda por n_2 que es el correspondiente, la expresión quedará:

$$\overline{X}=\frac{n_1\sum_{j=1}^{n}x_j n_j}{n_1 n}+\frac{n_2\sum_{r=n+1}^{k}x_r n_r}{n_2 n}=\frac{n_1\left(\frac{\sum_{j=1}^{n}x_j n_j}{n_1}\right)}{n}+\frac{n_2\left(\frac{\sum_{j=1}^{n}x_j n_j}{n_2}\right)}{n}$$

Como $\frac{\sum_{j=1}^{n}x_j n_j}{n_1}=\overline{x}_1$ y $\frac{\sum_{r=n+1}^{kn}x_{rj}n_{jr}}{n_2}=\overline{x}_2$ son la media de la primera y segunda submuestra respectivamente, la expresión la podemos expresar de la siguiente manera:

$$\overline{X}=\overline{X}_1\frac{n_1}{n}+\overline{X}_2\frac{n_2}{n}=\frac{\overline{X}_1 n_1+\overline{X}_2 n_2}{n}$$

que es lo que queríamos demostrar; ya que si las frecuencias se multiplican o dividen por un mismo número, la media no varía

Ejemplo 10: Para demostrar estos datos utilizaremos 3 conjuntos de datos:

Primero Conjunto: 5, 6, 8, 5, 4, 3, 9, 7

Calculamos su media aritmética

$$\overline{X}_1=\frac{5+6+8+5+4+3+9+7}{8}=\frac{47}{8}=5.875$$

Segundo Conjunto: 12, 10, 9, 15, 8

Calculamos su media aritmética

$$\overline{X_2} = \frac{12+10+9+15+8}{5} = \frac{54}{5} = 10.8$$

Tercer Conjunto: 54, 60, 50, 52, 58, 65, 57

Calculamos su media aritmética

$$\overline{X_3} = \frac{54+60+50+52+58+65+57}{7} = \frac{396}{7} = 56.571$$

Si calculamos la media aritmética de la forma tradicional tendríamos que hacer los cálculos de la siguiente manera:

$$\overline{X} = \frac{5+6+8+5+4+3+9+7+12+10+9+15+8+54+60+50+52+58+65+57}{20}$$

$$\overline{X} = \frac{497}{20} = 24.85$$

Ahora bien, con la propiedad anterior podemos calcularlo directamente con la fórmula:

$$\overline{X} = \frac{\overline{X_1}n_1 + \overline{X_2}n_2 + \overline{X_3}n_3}{n_1 + n_2 + n_3}$$

Sustituyendo

$$\overline{X} = \frac{5.875*8 + 10.8*5 + 56.75*7}{8+5+7}$$

$$\overline{X} = \frac{496997}{20} = 24.85$$

Observamos que ambas medias aritméticas dieron como resultado 24.85.

MEDIANA

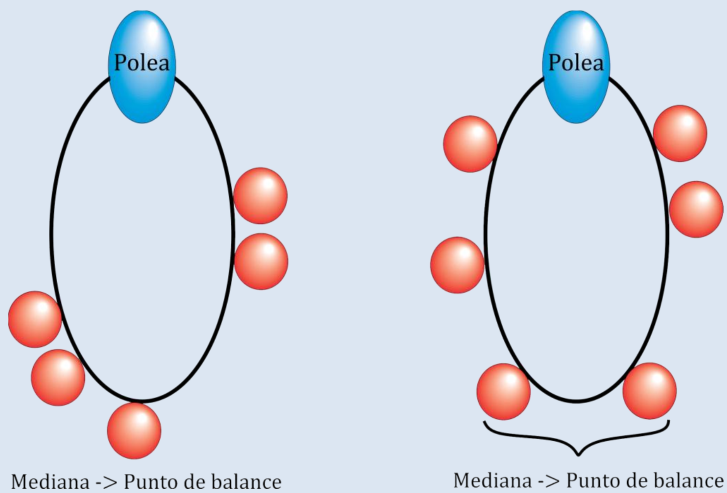


Figura 5: Representación de la Mediana en una polea

La mediana es un término que se utiliza muy frecuentemente en la vida cotidiana; a los niños se les enseña desde que aprenden a distinguir tamaños de objetos; por ejemplo, desarrollan actividades como: Ordenar sus juguetes por tamaño, o colorear la figura mediana, entre otras.

La mediana se define como el valor que está en el centro de todos los datos; no es ni tan grande, ni tan pequeño; así nos podemos encontrar con una gran cantidad de situaciones en donde esté presente la mediana como por ejemplo:

Mesa mediana

Figura 6: Mediana en la Vida Cotidiana

Específicamente, la mediana es el valor del término medio que divide una serie de datos ordenados en dos mitades con una mitad de las observaciones mayores que ésta y la otra mitad menores a la mediana; es decir, el 50% de los datos se ubican sobre la mediana o hacia los puntajes altos y el 50% restante hacia los puntajes bajos.

Es importante tomar en cuenta que si hay datos repetidos deben ser incluidos en el ordenamiento.

En términos de frecuencia, la Mediana es el valor de la distribución que, una vez ordenados los valores de la variable de menor a mayor, deja igual número de frecuencias a su izquierda que a su derecha, es decir, el valor que ocupa el lugar central. Puede entenderse también como aquel valor cuya frecuencia absoluta acumulada es $\frac{n}{2}$.

Notación y Cálculo

La mediana de n observaciones de la variable X se denotará por el símbolo: $\bar{X}$ o M_d y en algunas ocasiones como M_e.

La fórmula para su cálculo depende cómo se tengan organizados los datos; ya que pueden estar sin agrupar o agrupados en una tabla de distribución de frecuencias, que pueden ser en intervalos o no.

Caso 1: Datos sin agrupar

Para realizar este cálculo hemos de ordenar las puntuaciones en orden creciente o decreciente y fijarnos en el puesto mediano, que será el que deje por encima y por debajo el mismo número de datos de la serie.

Se distinguen dos situaciones:

a) Si el número n de datos es impar, la mediana es el dato que se encuentra exactamente en el centro de la lista. Para calcular su posición se aplica la siguiente ecuación:

$$\tilde{x} = x_{\frac{n+1}{2}}$$

b) Si el número n de datos es par, la mediana es la media aritmética de los dos datos que se encuentran a la mitad de la lista.

Para calcular su posición se aplica la siguiente ecuación:

$$\tilde{x} = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$$

Caso 2: Datos agrupados en tablas de frecuencias

En este caso hay que acudir al concepto de frecuencias acumuladas para determinar la mediana. Se calcula $\frac{n}{2}$ y se construye la columna de las frecuencias absolutas acumuladas. Se observa cuál es la primera frecuencia absoluta acumulada que supera o iguala $\frac{n}{2}$, **distinguiéndose dos casos:**

a) Si la frecuencia absoluta acumulada es mayor que $\frac{n}{2}$, la mediana es la x_i que corresponde a dicha frecuencia.

b) Si la frecuencia absoluta cumulada es igual a $\frac{n}{2}$, la mediana es la media aritmética de x_i y del siguiente x_{i+1}. Si este resultado no fuera admisible, la mediana sería los dos valores conjuntamente.

Caso 3: Datos agrupados en intervalos.

En el caso de estar la distribución agrupada en intervalos (sean o no de la misma amplitud) al buscar el valor que ocupa el lugar $\frac{n}{2}$ nos encontramos con un intervalo $\mathbf{I_i = [L_{i-1} - L_i]}$ y no con un dato; cuyo intervalo se denomina intervalo mediano. Para determinar un único representante de dicho intervalo como mediana, se observa la columna de las frecuencias absolutas acumuladas y se busca el primer intervalo cuya N_i sea mayor o igual que $\frac{n}{2}$, que será el intervalo que contiene a la mediana, **distinguiéndose dos casos:**

a) Si la frecuencia absoluta acumulada es igual a $\frac{n}{2}$, la mediana es el límite superior del intervalo mediano; es decir $\tilde{X} = L_i$.

b) Si la frecuencia absoluta acumulada es mayor que $\frac{n}{2}$, entonces el intervalo mediano es $[L_{i-1} - L_i]$ y la mediana es:

$$\tilde{X} = L_{i-1} + \frac{\frac{n}{2} - N_{i-1}}{n_i} * a_i \quad \textbf{\textit{Ecuación 4}}$$

Donde:

n : Número total de datos

L_{i-1}: Limite inferior del intervalo mediano

N_{i-1}: Frecuencia absoluta acumulada anterior a la correspondiente a dicho intervalo.

n_i : Frecuencia absoluta del intervalo mediano

a_i : Amplitud del intervalo y $a_i = \mathbf{L_i - L_{i-1}}$

Ejemplos de aplicación

Ejemplo 11: Calcular la mediana de las siguientes calificaciones del curso de Matemática evaluadas sobre diez: 10, 8, 9, 6, 4, 8, 9, 7, 10 y 9.

Solución

Como los datos se encuentran sin agrupar (**caso 1**) y el número de datos (n=10) es par.

Primero: se ordena los valores de menor a mayor.

4	6	7	8	8	9	9	9	10	10

Segundo: se identifican las posiciones que ocupa cada valor.

4	6	7	8	8	9	9	9	10	10
X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}

Tercero: aplica la fórmula $\tilde{x} = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$, para encontrar la posición de los dos valores centrales de los datos y la mediana.

$$\tilde{x} = \frac{x_{\frac{10}{2}} + x_{\frac{10}{2}+1}}{2} = \frac{x_5 + x_6}{2} = \frac{8+9}{2} = \frac{17}{2}$$

$$\tilde{x} = 8.5$$

Interpretación del resultado: La mitad de los alumnos tienen una nota menor que 8.5 y la otra mitad una nota mayor que 8.5.

Ejemplo 11: El Centro de Salud del municipio Ilopango realiza una encuesta en la colonia San José para estimar el número de mascotas que tienen los vecinos; obteniéndose los siguientes datos:

No. de mascotas	0	1	2	3	5	n
n_i	15	11	9	5	2	42
N_i	15	26	35	40	42	

Encuentre la mediana para el número de mascotas de los vecinos.

Solución

La variable aleatoria a estudiar X es el número de mascotas de las personas que viven en la colonia san José.

Por la forma en se presentan los datos se aplica el **caso 2**, y nos piden calcular la mediana.

Primero se calcula $\frac{42}{2} = 21$ y luego se observa en la columna de frecuencias absoluta acumulada qué valor de ellas supera o iguala a 21, en este caso es $N_2 = 26$, entonces $X = 1$.

Interpretación del resultado: La mitad de las personas encuestadas tienen menos de una mascota y la otra mitad tiene más de una mascota.

Ejemplo 12: Dada la siguiente distribución de frecuencia que corresponde a las horas extras laboradas por un grupo de obreros. Realice los cálculos respectivos para completar el siguiente cuadro.

N° de Horas extras $[L_{i-1} - L_i]$	Número de obreros (ni)	N_i
[55 – 59]	6	6
[60 – 64]	20	26
[65 – 69]	18	44
[70 – 74]	**50**	**94**
[75 – 79]	17	111
[80 – 84]	16	127
[85 – 89]	5	132
n	132	

Calcule la mediana de las horas extras.

Solución:

Por la forma en que se presentan los datos se aplica el **caso 3**.

Primero se calcula $\frac{132}{2} = 66$ y luego se observa que valor de N_i supera o iguala a 66; para el caso es $N_4 = 94$, entonces el intervalo mediano es $[L_{i-1} - L_i] = [70-74]$ que es donde se encuentra la mediana. Ya encontrado el intervalo mediano utilizar la siguiente fórmula:

$$\tilde{x} = L_{i-1} + \frac{\frac{n}{2} - N_{i-1}}{n_i} * a_i$$

Antes de aplicar la fórmula se identifican los términos que la forman.

El intervalo modal es: $[L_{i-1} - L_i] = [70-74]$ que corresponde al intervalo número 4 (i=4), de donde $L_{i-1} = 70$, $n = 132$, $\frac{n}{2} = 66$, $N_{i-1} = 44$ $n_4 = 50$, $a_4 = 5$; ya que los dos valores de los extremos del intervalo están incluidos en él.

Sustituyendo se tiene:

$$\tilde{x} = 70 + \frac{66-44}{50} * 5 = 70 + \frac{22}{50} * 5 = 70 + \frac{110}{50}$$

$$= 70 + 2.2$$

$\tilde{x} = 72.2$ horas

Interpretación del resultado: Significa que la mitad de los obreros, trabajan horas extras por debajo de 72.2 horas y la otra mitad trabaja horas extras por encima de 72.2 horas.

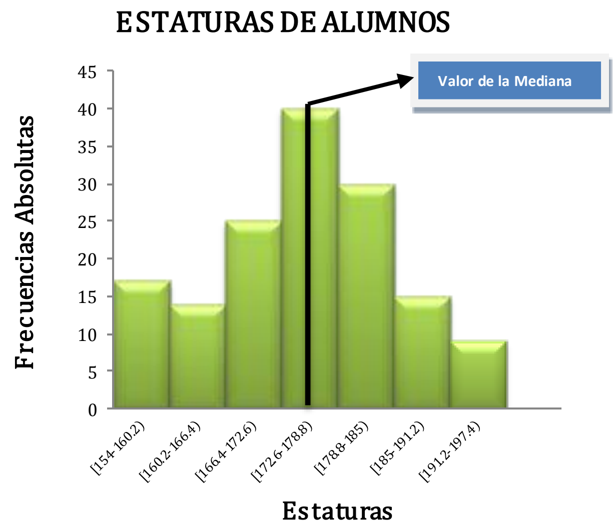
Propiedades de la Mediana

a) Puede ser usada no sólo para datos numéricos sino además para datos en escala ordinal (datos que pueden ser ordenados), ya que para calcularla sólo es necesario establecer un orden en los datos; siendo la medida más representativa de estos por describir la tendencia central de los mismos.

b) Su cálculo es sencillo e interesa que los valores estén ordenados de menor a mayor o viceversa.

c) En ella solo influyen los valores centrales de la distribución y es insensible a los valores extremos lo cual es útil cuando existen muchos valores extremos que invaliden otras medidas de posición central.

d) La mediana es el valor cuya vertical divide al histograma en dos partes de igual superficie.



MODA

Cuando en el medio cotidiano se escucha la palabra moda siempre lo relaciona con vestidos, trajes, corbatas, faldas, pantalones, zapatos, etc. Y precisamente el término de moda está presente en la vida cotidiana; y podría surgir la ¿Por qué sabemos que algún producto está de moda? Seguramente responderás… "Por qué lo usan muchas personas, o porque lo vemos frecuentemente en la calle", y efectivamente eso es la moda, aquello que se impone, la generalidad de las personas lo lleva.

El concepto de moda en estadística es exactamente lo mismo; considerándose en una serie de números el valor que se presenta con mayor frecuencia; es decir, el que se repite un mayor número de veces. Es por tanto, el valor más común.

Una lista de valores puede tener más de una moda; esto es posible cuando se encuentran dos o más valores que se repiten el mismo número de veces, e incluso puede no existir. Por lo tanto se puede tener las siguientes situaciones:

- Si no se repite ningún valor en la serie de valores o si dos o más valores están empatados en cuanto a mayor frecuencia (número de ocurrencias), se dice que no hay moda y la serie es amodal;

- Si hay dos valores que se repiten el mismo número de veces, se dice que la serie es Bimodal;
- Si hay tres valores que se repiten el mismo número de veces, se dice que la serie es trimodal;
- Y en general se pueden encontrar una serie que sea multimodal.

De manera gráfica, la moda equivale al valor que alcanza la frecuencia máxima o pico en el polígono de frecuencias; y para en un histograma, una moda es un máximo relativo ("un salto"); es decir la barra que tiene mayor tamaño.

Si se trabaja con datos sin agrupar se recomienda ordenar los datos de menor a mayor para tener una mejor visión de los datos que están repetidos y si hay datos repetidos deben ser incluidos en el ordenamiento.

Notación y Cálculo

La moda de n observaciones de la variable X se denotará por el símbolo: M_o.

Al igual que las medidas anteriores para su cálculo depende cómo se tengan organizados los datos; ya que pueden estar sin agrupar o agrupados en una tabla de distribución de frecuencias, que pueden ser en intervalos o no.

Caso 1: Datos sin agrupar

Si se tienen datos sin agrupar, se encuentra fácilmente simplemente observando cual es el valor que más se repite, y se recomienda ordenar los valores en orden creciente o decreciente para tener una mejor visibilidad de las veces que se repite cada valor.

Caso 2: Datos agrupados en tablas de frecuencias

Una vez construida la tabla de frecuencias absolutas, se localiza la mayor frecuencia absoluta, y la moda es su correspondiente valor de la variable; es decir $M_o = x_i$.

Caso 3: Datos agrupados en intervalos.

Cuando la variable viene agrupada en intervalos de clases, el primer paso será calcular el intervalo (o intervalos) modal o modales; será el que mayor frecuencia absoluta tenga.

Visto gráficamente en el histograma el intervalo modal será aquel intervalo tal que su histograma le corresponda el rectángulo de mayor área por unidad de base.

Luego de haber identificado el intervalo modal se supone que dicho intervalo tiene de extremos $I_i=[L_{i-1}- L_i]$ y que todos los intervalos son de igual amplitud (a_i). En este caso, la moda es un valor situado dentro de este intervalo, y se calcula utilizando la siguiente fórmula:

$$M_o = L_{i-1} + \frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})} * a_i$$ **Ecuación 5**

Dónde:

n: Número total de datos

n_i : Frecuencia absoluta del intervalo modal.

L_{i-1}: Límite inferior del intervalo modal.

n_{i-1}:Frecuencia absoluta anterior del intervalo modal.

n_{i+1}:Frecuencia absoluta posterior del intervalo modal.

a_i : Amplitud del intervalo.

$$a_i = L_i - L_{i-1}$$

Ejemplos de aplicación

Ejemplo 13: Al preguntar a diez conductores cuántos litros de gasolina consume su vehículo en la carretera por cada 100 km, éstas fueron sus respuestas: 8, 9, 10, 8, 6, 6, 5, 7, 7, 7.

Determine la moda del consumo.

Solución.

Por la forma en que se presentan los datos se aplica el **caso 1**.

Primero se ordena los valores de menor a mayor: 5, 6, 6, 7, 7, 7, 8, 8, 9, 10. Y se observa el número de veces que aparece cada valor.

El 5 aparece 1 vez, el **6** aparece 2 veces, el **7** aparece 3 veces, el **8:** aparece 2 veces, el **9** aparece 1 vez, el valor **6** aparece 1 vez.

Luego, se toma como moda el que aparece el mayor número de veces; entonces M_o =7 litros

Interpretación del resultado: 7 litros de gasolina es la cantidad que la mayoría de conductores consumieron por cada 100 km de recorrido.

Ejemplo 14: Se les pregunto a 60 alumnos de primer año de bachillerato general, el número de asignaturas reprobadas en el primer trimestre de este año; y obtuvo la siguiente tabla de distribución de frecuencias.

x_i	0	1	2	3	4	5	n
n_i	8	11	13	**15**	10	3	60

Determinar el número de asignatura que más reprobaron los alumnos.

Solución.

La variable X a estudiar es el número de asignaturas reprobadas por los alumnos en el primer trimestre.

Por la forma en que se presentan los datos se aplica el **caso 2.**

Para encontrarla simplemente se observa la columna de las frecuencias absolutas (n_i) y la moda será el valor de la variable (x_i) que tenga mayor valor de la frecuencia absoluta.

En este caso el mayor valor de frecuencia absoluta es $n_4 = 15$; por lo tanto la moda $M_o = 3$.

Interpretación del resultado: La cantidad de asignaturas que reprobaron, la mayoría de los alumnos del primer año de bachillerato en el primer trimestre fue 3.

Ejemplo 15: Dada la siguiente distribución de frecuencia correspondiente al peso en Kg. de un grupo de trabajadores de una empresa, calcule la moda.

Peso (Kg) $[L_{i-1}- L_i]$	**Número de trabajadores (ni)**
[30 – 39]	2
[40 – 49]	2
[50 – 59]	7
[60 – 69]	11
[70 – 79]	12
[80 – 89]	16
[90 – 99]	2
n	52

Solución.

Por la forma en que se presentan los datos se aplica el **caso 3**

Primero se debe identificar el intervalo modal y es que tiene mayor valor de la frecuencia absoluta, para el caso el intervalo modal es: $[L_{i-1}- L_i]$ = [80-89]; que es donde se encuentra la moda .

Y luego se utiliza la siguiente fórmula:

$$M_o = L_{i-1} + \frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})} * a_i$$

Antes de aplicar la formula identificar los términos que tiene.

El intervalo modal es: $[L_{i-1} - L_i] = [80\text{-}89]$ que corresponde al intervalo número 6 (i=6), de donde $L_{i-1} = 80$, $n_{i+1} = n_7 = 2$, $n_{i-1} = n_5 = 12$ $n_6 = 16$, $a_6 = 10$ ya que los dos valores de los extremos del intervalo están incluidos en él.

Sustituyendo se tiene:

$$M_o = 80 + \frac{16-12}{(16-12)+(16-2)} * 10 = 80 + \frac{4}{4+14} * 10$$

$$= 80 + \frac{4}{18} * 10 = 80 + \frac{40}{18}$$

$$= 80 + 2.2222$$

$$M_o = 82.22\,Kg.$$

Interpretación del resultado: 82.2 kg es el peso que la mayoría de los trabajadores tienen.

Características de la Moda

a) No es única, o sea, un conjunto de datos numéricos puede tener más de una moda.

b) Puede ser afectada grandemente por el método de elaboración de los intervalos de clases.

c) No es afectada por la magnitud de los valores extremos de una serie de valores, como sucede en la media aritmética.

d) La moda se puede obtener en una forma aproximada muy fácilmente, puesto que la obtención exacta es algo complicado.

e) Tiene poca utilidad en una distribución de frecuencia que no posea suficientes datos y que no ofrezcan una marcada tendencia central.

f) No es susceptible de operaciones algebraicas posteriores.

g) Se utiliza cuando se trabaja con escalas nominales aunque se puede utilizar con las otras escalas.

h) Es útil cuando se está interesado en tener una idea aproximada de la mayor concentración de una serie de datos.

ACTIVIDADES DE ALUMNAS Y ALUMNOS

Cálculo Media, Moda y Mediana

Actividad 1: Recoger datos

Solicitar a cada estudiante, al menos un día antes de desarrollar esta actividad, que lleve sus datos: edad, estatura y peso. Con la información de toda la clase, completar una tabla similar a la siguiente.

Organizar grupos de trabajo de tres o cuatro alumnos(as). Indíqueles que calculen las Medidas de Tendencia Central: La Media Aritmética, la Mediana y la Moda, de cada una de las tres categorías (Edad, estatura y peso).

Nombre de alumnos(as)	Edad	Estatura	Peso

Actividad 2: Nombres de Países de Centro América.

La siguiente actividad se puede hacer en grupos de 3 o 4 alumnos.

Material: Papel cuadriculado, tijera, lapicero.

Desarrollo de Actividad

1) Cortar 5 tiras de papel y en cada una escribir los nombres de los países de Centro América.

G	*U*	*A*	*T*	*E*	*M*	*A*	*L*	*A*

H	*O*	*N*	*D*	*U*	*R*	*A*	*S*

E	L	S	A	L	V	A	D	O	R

N	*I*	*C*	*A*	*R*	*A*	*G*	*U*	*A*

C	*O*	*S*	*T*	*A*	*R*	*I*	*C*	*A*

2) Ordenar los nombres de los países, del más corto al más largo.

3) En otra tira de papel colocar la cantidad de letras de cada nombre de los países, un número por cada país. En nuestro ejemplo tendríamos:

8	*9*	*9*	*9*	*10*

4) ¿Cuál es el número de letras que más se repite? _________

¿Cómo se le llama a este valor?_________

5) Doblar la tira a la mitad, por donde marca la línea vertical. ¿Qué medida de tendencia central se obtiene? ________________ , ¿Qué valor le corresponde? _______

8	*9*	*9*	*9*	*10*

6) Calcular la media aritmética.

7) Escribe los valores de cada medida de tendencia central:

Media Aritmética: ________ Mediana: _______ Moda:________

Actividad 3: ¿Quiénes son más altos? ***Duración: 2 horas***

El equipo de Basquetbol de Primer año Comercial se enfrentará al de Primer año General. Siempre se ha corrido el rumor que el equipo de Primer año General tiene jugadores de mayor estatura, por lo que tienen más ventaja.

Los encargados de dirigir el partido tienen el registro de las estaturas de los jugadores que participarán en este encuentro deportivo y los publica para ambos equipos.

Equipo de Primer año General : 1.73, 1.75, 1.83, 1.75, 1.89, 1.95, 1.85, 1.76, 1.75, 1.82, 1.90 en metros.

Equipo de Primer año Comercial: 1.85, 1.69, 1.92, 1.89, 1.78, 1.78, 1.89, 1.88, 1.69, 1.95, 1.89 en metros.

Con base en esta información, ¿Qué equipo tiene los jugadores más altos? ¿Cómo puedes comparar las estaturas de ambos equipos para que nos ayude a saber quién tiene mayor ventaja por su estatura?

Reúnase con su grupo de trabajo, analicen la información y resuelvan lo siguiente:

1. Para cada equipo, realiza la suma de todos los datos y divídelo entre el total de ellos. ¿Cómo se llama este resultado.
2. ¿Cuál es la diferencia entre las medias de cada equipo?
3. ¿Cuál de los equipos se puede decir que supera al otro en la estatura de sus jugadores?____________
4. Ordena de menor a mayor cada una de las estaturas, para cada equipo. ¿Qué estatura es la que se encuentra a la mitad de cada lista? _______ y ______ ¿Con qué nombre se conoce este valor?:____________
5. ¿Es la mediana muy diferente a la media aritmética? ______ En cuanto Difieren?:_______________
6. ¿Consideras que ambas pueden ayudarte a realizar la comparación de ambos equipos?_______
7. ¿Cuál prefieres?______________
8. De cada lista de jugadores, ¿cuál estatura es la que más se repite?___________
9. ¿Encuentras alguna similitud de este valor con la media y la mediana?_________
   Como se llama este valor? _______________
10. ¿En qué situaciones que conozcas puedes utilizar el concepto de moda? Escribe tres ejemplos.
11. Un día antes del encuentro, decidieron los encargados de dirigir el partido aumentar a la lista de jugadores tres personas más por equipo. **Primer año General** llevará a José, Arturo y Pedro, de 1.75, 1.84 y 1.68 m de estatura, respectivamente. Mientras que **Primer año Comercial** llevará a Luis, Jorge y Santiago de 1.78, 1.69 y 1.78 m.

    Determinen para cada equipo la media aritmética, la mediana y la moda con estos nuevos datos.

    ¿Qué equipo tiene más ventaja por su estatura?

    Una vez terminada la actividad entrégala a tu profesor para su revisión.

APLICANDO LO APRENDIDO

Indicación: Resolver de manera clara y ordenada las siguientes situaciones problemáticas.

1. Al tomar una muestra de 40 personas y observar el número de caries que presentan, se ha registrado los siguientes datos.

Número de Caries	1 2 3 4 5 6 7 8
Número de Persona	2 6 10 5 10 3 2 2

a) Encontrar la cantidad media de caries encontradas en este grupo de personas **R/** 4.05
b) Cuál es la cantidad de caries que más poseen estas persona. Es única? **R/** 3 y 6
c) Cuál es la cantidad de caries que tienen como máximo la mitad de las personas. **R/** 4

2. Las puntuaciones de Cristina en cuatro pruebas son 7, 6,5 y 5. ¿Cuánto debe sacar en la quinta prueba para que la media aritmética de las puntaciones de las cinco pruebas sea 6? **R/** 7
3. El promedio de 3 estudiantes es 5,4 y el promedio de otros 4 estudiantes es 6,7, ¿Cuál es el promedio de los 7 estudiantes?
4. El número de emergencia que se han atendido en el Hospital Rosales en 30 noches se presentan en la siguiente tabla.

No. de Emergencias	**0**	**1**	**2**	**3**	**4**	**5**	**6**
No. de días	7	8	5	4	3	1	2

a) Determinar el número promedio de Emergencias atendidas **R/**1.97
b) Encontrar la cantidad de emergencias que más se atendieron. **R/** 1
c) Determinar la cantidad de emergencias que tiene como máximo la mitad de los días. **R/** 1.5

5. La estatura, en centímetros de un grupo que pertenecen a los alumnos y alumnas del primer año de bachillerato es:

   150 , 169 , 171 , 172 , 175 , 181 , 182 , 183 ,177 , 179 , 176 , 184 ,158

   Calcule las medidas de tendencia central y de una interpretación razonable de ellas.

   R/ Media: 173.6 Mediana: 176 Moda: No hay

6. Se pidió a 15 estudiantes que dijeran el número de horas que habían dormido la noche anterior. Los datos fueron: 5, 6, 6, 8, 7, 7, 9, 5, 4, 8, 11, 6, 7, 8, 7.Calcula el promedio de horas que durmieron los encuestados. **R/** 6.9
7. La media aritmética de estas dos series de datos es 5:

   Serie 1: 5, 5, 5, 5, 5 Serie 2: 20, 0, 0, 4, 1

   ¿A cuál de las dos series representa mejor la media aritmética? **R/** Representa mejor a la serie 1. ¿Por qué?: los valores de la serie 1 están alrededor de 5 y la serie 2 los datos están muy alejados de 5.

8. Un comerciante realizó una encuesta para saber cuáles eran las tallas de cinturón para caballero que debía tener en la bodega. Los resultados aparecen en la tabla:

Talla	30	31	32	33	34	35	36	37	38	39	40	Total

Frecuencia Absoluta	2	2	3	5	6	8	14	11	8	3	1	63

a) ¿A qué tallas corresponde la mayor frecuencia? **R/** Talla 36 y Talla 37
b) ¿A qué talla le corresponde la menor frecuencia? **R/** Talla 40
c) ¿Qué medidas corresponden a la frecuencia 3? **R/** Talla 32 y Talla 39
d) ¿Cuál es el promedio? **R/** 35.54 o Talla 36
e) ¿El promedio indica que los encuestados son gordos o delgados?.**R/** La talla de cinturón promedio es 36 (como valor entero), no indica si son gordos o delgados, aunque debe haber más tallas grandes que pequeñas.

PRUEBA OBJETIVA

PROBLEMA 1: Las calificaciones de Matemáticas en el último examen fueron: 6, 3, 5, 4, 5, 6, 10, 5, 5, 4, 6, 6, 7, 10, 5, 6, 6, 7, 2, 4, 8, 9, 5, 7, 3, 8, 6

a) ¿Cuántos alumnos presentaron el examen? **R/** 27 alumnos
b) Ordena los datos y haz una tabla con las frecuencias absolutas.
c) ¿Cuál es la nota más baja? ¿Y la más alta? **R/** La nota más baja es 2 y la más alta es 10
d) ¿Cuántos alumnos obtuvieron la peor calificación y cuántos la mejor?, **R/** uno la peor nota y dos la mejor.
e) ¿Cuántos reprobaron? **R/** 12 alumnos reprobaron
f) Encuentra la moda. **R/** la moda es 6
g) Encuentra el promedio de calificaciones e interprétalo. **R/** 5.85

PROBLEMA 2: A 150 personas se les ha realizado un test de 50 preguntas sobre seguridad vial y se han obtenido los siguientes resultados agrupados en intervalos.

Intervalo	[0-10[	[10-20[	[20-30[	[30-40[	[40-50]
Frecuencia	24	32	48	26	20

Obtener las medidas de tendencia central e interpretarlas.

R/ Media: 24.1 Moda: 24.2 Mediana: 24.0

PROBLEMA 3: Se ha estudiado el coeficiente intelectual de los 210 alumnos de Bachillerato de un Centro Escolar, obteniéndose los siguientes resultados.

X_i	[82,90[	[90,98[	[98,106[	[106,114[	[114.122[	[122,130[	[130,138[	[138,146]
n_i	12	32	49	54	30	17	11	5

Donde X_i: Coeficiente Intelectual n_i: Número de Alumnos

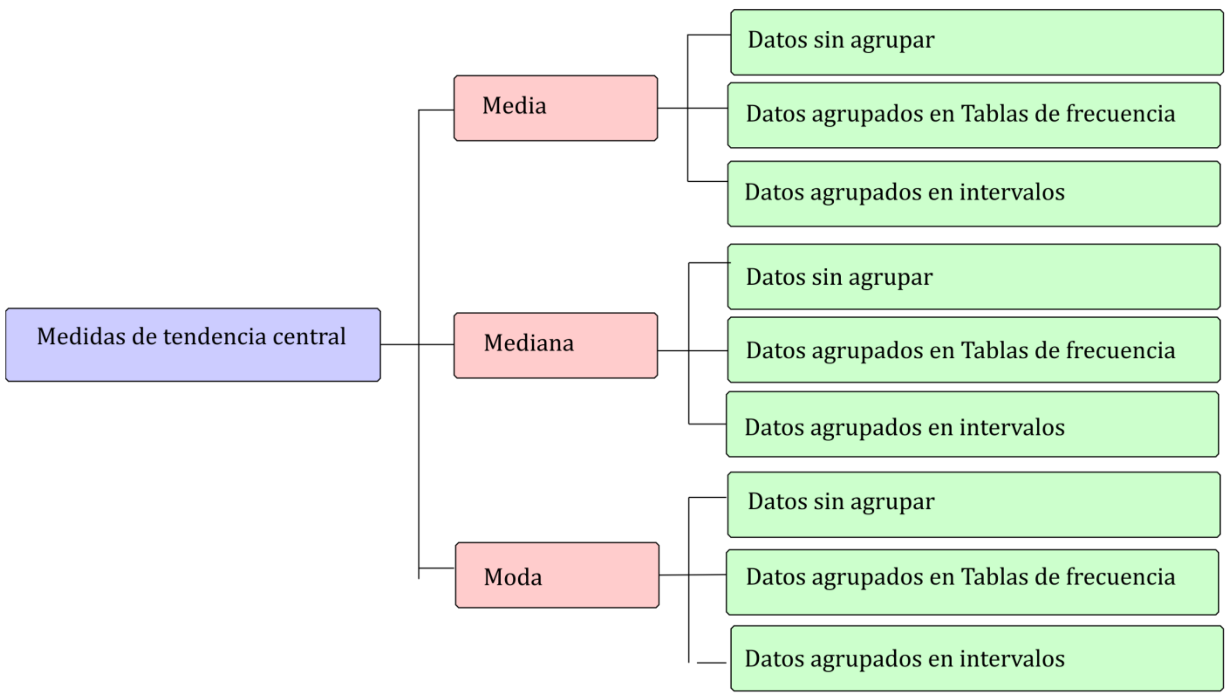
Obtener las medidas de tendencia central e interpretarlas.

R/ Media: 108.8 Moda: 107.4 Mediana: 107.8

BIBLIOGRAFIA

1. Johnson R., Cuby P.(1999), *Estadística Elemental*. México:Internacional Thomson Editores, S.A de S.V .
2. Pérez C., (2003).*Estadística. Problemas Resueltos y Aplicaciones*. Madrid: Pearson Educación, S.A.
3. Sarabia J.M. (2000), *Curso Práctico de Estadística*.(2da ed.) . España: Impreso por Gráficas Rogar, S.A Navalcarnero (Madrid).
4. Triola, M.,(2009). *Estadística*. (10ª ed.). México: Pearson Educación.

DIAGRAMA DE CONTENIDOS



Lección 4 **Primer año de Bachillerato** **Unidad VI** **Tiempo: 8 horas clase**

Medidas de posición

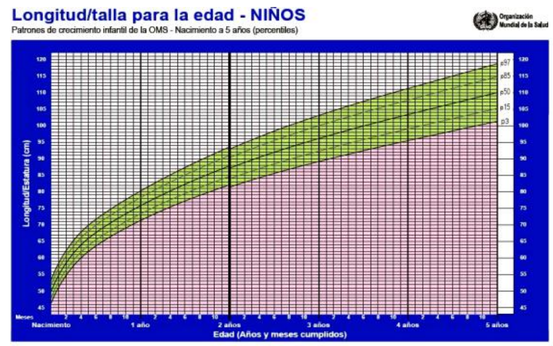


Figura 1. **Curva de crecimiento de talla y peso de bebes.**

Introducción del Tema

Un conjunto de mediciones puede dividirse en cierto número de partes iguales mediante la selección de valores que correspondan a una posición determinada en dicho conjunto. Por ejemplo, la mediana divide un conjunto de valores dados en dos partes iguales, y su posición es, en consecuencia, a la mitad del mismo, de manera que el 50% de los valores quedan a uno y otro lado de dicho valor estadístico. Las medidas de posición no centrales permiten conocer otros puntos característicos de la distribución que no son los valores centrales. En general se llaman cuantiles a estos valores con esa posición divisoria determinada; que trata de resumir lo que ocurre en determinados tramos o intervalos del conjunto de datos.

Entre otros indicadores, se suelen utilizar una serie de valores que dividen la muestra en tramos iguales.

Las medidas de posición no central que más se utilizan son: los cuartiles, deciles y percentiles y se asemejan a la mediana por que dividen el conjunto de datos en partes iguales, la mediana lo hace en dos los que están por encima y por debajo de ella, mientras que los cuartiles dividen el conjunto de valores en cuatro partes iguales, los deciles en diez y los percentiles en cien.

Objetivos

- Calcular las medidas de posición no centrales más importantes.
- Diferenciar entre las tres medidas de posición (cuartil, decil y percentil), y calcularlas estando agrupadas por intervalos.
- Interpretar las medidas de posición no central.

Importancia

Estas medidas se utilizan para saber si el desarrollo de los bebes, niños y niñas es el adecuado en talla y peso; los médicos utilizan las curvas de crecimiento que son gráficas donde están representados los percentiles. Hay una para la talla o longitud, otra para el peso y otra para el perímetro craneal de los bebes, están diferenciadas para los niños y niñas, y por edades de 0 a 2 años y de 2 a 14 años pues su desarrollo es distinto.

Cada línea corresponde a un percentil, para que el desarrollo de la niña o el niño sea el adecuado debe mantenerse siempre más o menos en el mismo percentil, no sería normal que pasara en poco tiempo de un percentil 90 a un percentil 50, seria signo de algún problema de salud.

Si una niña, un niño o un bebe está situado en el percentil 10 de talla y peso, quiere decir que el 10% de la población está por debajo de esos valores y el 90% de la población por encima. En cambio si se sitúa en el percentil 90 implica que es de los de mayor estatura de la población, ya que solo el 10% lo supera.

La determinación de los cuartiles demuestra su utilidad con mucha frecuencia. Por ejemplo la mayoría de las universidades solo reciben a los estudiantes cuyas pruebas de estado los ubique en el tercer cuartil, es decir, que estén ubicados en el 25% superior de los aspirantes. También es muy frecuente encontrar que los dirigentes de las universidades se muestran muy interesados en encontrar las causas que crean problemas entre los estudiantes universitarios cuyo desempeño está ubicado en el primer cuartil, es decir, en el 25% inferior de toda la comunidad universitaria.

Cabe mencionar que las grandes multinacionales solo están interesadas en contratar profesionales que estén ubicados en el 10% superior de todos los aspirantes, es decir, que este ubicado en decil nueve o su equivalente el percentil 90.

Competencias a reforzar

Calcula, analiza e interpreta las medidas de posición no central, para tomar decisiones acertadas ante una situación real.

Presaberes

- Operaciones básicas.
- Nociones de Porcentajes
- Noción de partición, posición y orden

MEDIDAS DE POSICIÓN

Además de conocer los valores de las medidas de tendencia central para un conjunto de datos, puede resultar interesante localizar la posición de determinadas puntuaciones individuales en relación con el grupo. De esto se encargan las medidas de posición; ya que informan de la posición de determinadas puntuaciones individuales en relación con el grupo del que forman parte, es decir del valor de la variable que ocupará la posición (en tanto por cien) que nos interese respecto de todo el conjunto de observaciones de la variable.

La mediana, además de indicar una tendencia central, puede ser considerada una medida de posición, si recordamos es un punto convencional en un conjunto de datos puesto que el valor que toma representa justo el centro del conjunto de datos, dejando el mismo número de observaciones por encima y por debajo de ella.

En este sentido se podría estar interesado en un valor que tuviera sólo el 25% inferior o el valor que sólo tiene el 10% de los datos superior a él. Estas situaciones han llevado a la idea de **cuantil**: Diremos que un número es el cuantil de orden p en un conjunto de datos si el porcentaje de datos inferiores a él es igual a p y los superiores 100-p.

Los cuantiles constituyen una generalización del concepto de mediana. Así como la mediana divide a la serie estudiada en dos partes con el mismo número de elementos cada una, si la división se hace en cuatro partes, o en diez partes, o en cien partes, llegamos al concepto de **cuantil**.

Se llaman medidas de posición o cuantiles de orden *k* a aquellos que dividen el conjunto de datos en *k* partes, de tal forma que en cada una de esas partes haya el mismo número de observaciones.

En términos generales los cuantiles son medidas de posición que dividen el conjunto de datos en un determinado número de partes de manera que en cada una de ellas hay la misma cantidad de observaciones de la variable. En la vida cotidiana nos podemos encontrar con este tipo de particiones, así como se muestra en la siguiente figura.

Figura 2. Medidas de posición vida cotidiana

Como ejemplo, consideremos la distribución de peso de recién nacidos de sexo femenino y 38 semanas de gestación Si se informa que el percentil 10 de esta distribución es 2450 g y el percentil 90 es 3370 g, está indicando que un 10% de las niñas que nacen en la semana 38 de gestación pesan 2450g o y que el 90% de las niñas de esta edad gestacional nacen con peso menor o igual que 3370 g y sólo el 10% con peso mayor que 3370 g.

Se usan para describir la posición que tiene un valor de datos específico en relación con el resto de los datos. Las medidas de posición de las cuales mas se hace uso son: Cuartiles, Deciles y Percentiles.

Para el cálculo de estas tres medidas de posición es necesario ordenar las observaciones de forma creciente o decreciente.

Se emplean generalmente, en la determinación de estratos o grupos correspondientes a fenómenos socio-económicos, monetarios o teóricos.

Observaciones a los cuantiles

1. Los cuantiles, en particular los deciles y percentiles, son parámetros estadísticos muy usados en ciencias sociales y en el campo de la salud.
2. Algunos de los cuantiles no están cerca del centro de la distribución, a pesar de ser considerados medidas de centralización por su analogía con la mediana, por eso también se les llama *medidas de posición.*
3. El cuartil primero coincide con el percentil 25 y el cuartil tercero con el percentil 75.

CUARTILES

Se definen los cuartiles como tres valores de la variable que dividen las observaciones ordenadas en cuatro partes porcentualmente iguales, estando en cada una de ellas el 25% de sus observaciones y se denotan por Q_i (Q_1, Q_2 y Q_3); de manera que el cuartil Q_m deja por debajo de sí m cuartas partes de las observaciones totales del conjunto de datos.

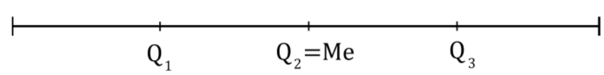


Figura 3. Representación de los Cuartiles

El primer cuartil (Q_1) es el valor de la variable que deja la cuarta parte de las observaciones menores o iguales a él y las tres cuartas partes superiores a él. Para ser más precisos, al menos el 25% de los valores son menores o iguales que Q_1, y al menos el 75% de los valores son mayores o iguales que Q_1.

El segundo cuartil (Q_2) es el valor de la variable que las dos cuartas partes de las observaciones son inferiores o iguales a él (la mitad) y las otras dos cuartas partes son mayores o iguales. Para ser más precisos es el valor que separa el 50% inferior de los valores del 50% superior. Este cuartil coincide con la mediana.

El Tercer cuartil (Q_3) es el valor de la variable que deja las tres cuartas partes de las observaciones inferiores o iguales a él y la cuarta parte de éstas superior a él. Para ser más precisos, al menos el 75% de los valores son menores o iguales que Q_3, y al menos el 25% de los valores son mayores o iguales a Q_3.

DECILES

Se definen los deciles como nueve valores de la variable que dividen las observaciones en diez partes porcentuales iguales, estando en cada una de ellas el 10% de sus observaciones y se denota por D_i (D_1, D_2, D_3, D_4, D_5, D_6, D_7, D_8, D_9); de manera que el decil D_m deja por debajo de sí m décimas partes de las observaciones totales del conjunto de datos.

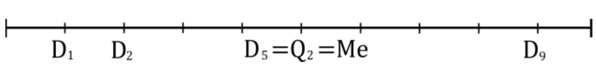


Figura 4. Representación de los Deciles

El primer decil (D_1) es el valor que al menos el 10% de los valores son menores o iguales que D_1, y al menos el 90% de los valores son mayores o iguales que D_1.

El segundo decil (D_2) es el valor que al menos el 20% de los valores son menores o iguales que D_2, y al menos el 80% de los valores son mayores o iguales que D_2.

El quinto decil (D_5) es el valor que al menos el 50% de los valores son menores o iguales que D_5, y al menos el 50% de los valores son mayores o iguales que D_5. Este decil coincide con la mediana.

El octavo decil (D_8) es el valor que al menos el 80% de los valores son menores o iguales que D_8, y al menos el 20% de los valores son mayores o iguales que D_8.

En términos generales el decil k-ésimo, se define como el valor de la variable que deja inferiores o iguales a él las $k/10$ partes de las observaciones, donde $k = 1, 2, 3, \ldots, 9$.

PERCENTILES

Se definen los percentiles como noventa y nueve valores de la variable, que dividen las observaciones en cien partes porcentuales iguales, estando en cada una de ellas el 1% de sus observaciones y se denota por P_i (P_1, P_2, P_3, ...,P_{99}); de manera que el percentil P_m deja por debajo de sí el m por ciento de las observaciones totales del conjunto de datos.

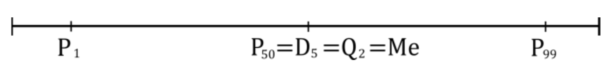


Figura 5. Representación de los Percentiles

Primer percentil (P_1) es el valor que al menos el 1% de los valores son menores o iguales que P_1, y al menos el 99% de los valores son mayores o iguales que P_1.

Décimo percentil (P_{10}) es el valor que al menos el 10% de los valores son menores o iguales que P_{10}, y al menos el 90% de los valores son mayores o iguales que P_{10}. Coincide con el primer decil.

25avo percentil (P_{25}) es el valor que al menos el 25% de los valores son menores o iguales que P_{25}, y al menos el 25% de los valores son mayores o iguales que P_{25}. Coincide con el primer cuartil.

50avo percentil (P_{50}) es el valor que al menos el 50% de los valores son menores o iguales que P_{50}, y al menos el 50% de los valores son mayores o iguales que P_{50}. Coincide con el segundo cuartil y del quinto decil y además coincide con la mediana.

En términos generales el percentil k-ésimo, se define como el valor de la variable que deja inferiores o iguales a él las $k/100$ partes de las observaciones, donde $k = 1, 2, 3,...,99$.

RELACIÓN DE CUARTILES, DECILES Y PERCENTILES

Muchas veces para facilitar los cálculos de estas medidas es importante considerar las siguientes relaciones:

$Q_1=P_{25}$				Me =	$Q_3=P_{75}$				
$D_1=P_{10}$	$D_2=P_{20}$	$D_3=P_{30}$	$D_4=P_{40}$	$Q_2=D_5=P_{50}$ $Q_4=D_{10}=P_{100}$	$D_6=P_{60}$	$D_7=P_{70}$	$D_8=P_{80}$	$D_9=P_{90}$	$D_{10}=P_{100}$

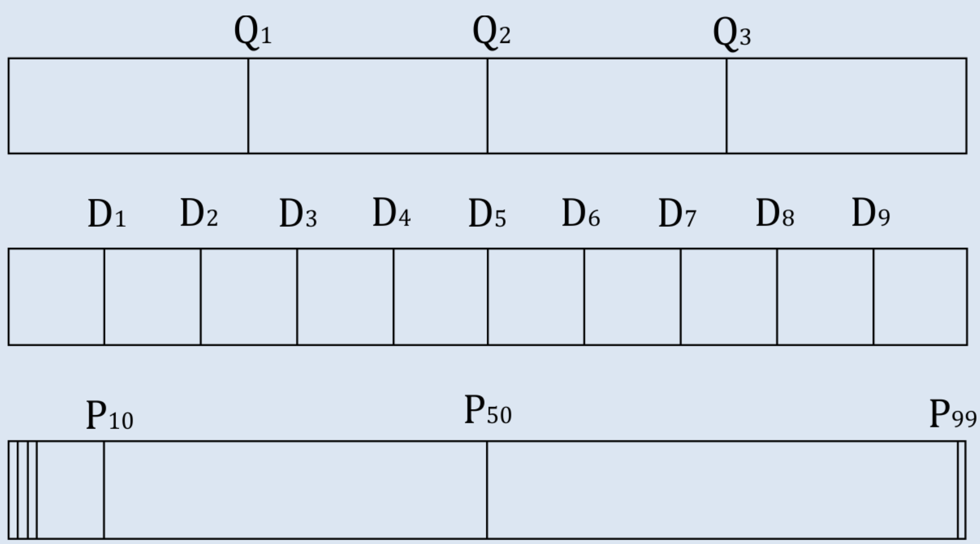


Figura 5: Relación de las medidas de posición

CÁLCULO DE LAS MEDIDAS DE POSICIÓN

Caso 1: Datos sin agrupar

El procedimiento para determinar el valor de cualquier cuantil se presenta a continuación.

1. Ordenar los datos de menor a mayor.
2. Calcular la posición del cuantil contada a partir del dato menor:

Posición $\longrightarrow \frac{r*n}{k}$

Donde:

r: Número del cuartil, decil o percentil a calcular.

n: Cantidad de datos o tamaño de la muestra.

k: Número de partes en que el cuantil divide al conjunto de datos.

Valores	Cuartiles	Deciles	Percentiles
r	1,2,3	1,2,3,4,5,6,7,8,9	1,2,3,…….,99
k	4	10	100

Tabla 1: Cálculo para datos sin agrupar

3. Si la posición $\frac{r*n}{k}$ es un número entero entonces el valor del cuantil se encuentra a la mitad del valor de la posición encontrada y el valor de la siguiente posición, luego obtener el valor del cuantil buscado como el promedio de ambos valores. Es decir el cuantil buscado es: $\frac{X_i + X_{i+1}}{2}$

Si la posición $\frac{r*n}{k}$ no es un número entero redondear al siguiente entero más grande y el valor del cuantil será el valor ubicado en dicha posición.

Ejemplo 1: Las temperaturas promedios (^{0}C) registradas durante el mes de noviembre de 2011, reportadas por la estación meteorológica: 786630 (MSSS) con sede en Ilopango son: 24.4, 24.6, 25.4, 25.6, 25.5, 25.4, 25.1, 24.7, 24.4, 25.6, 24.4, 23.8, 24.5, 24.6, 25.0, 23.8, 24.1, 25.0, 24.5, 24.8, 25.8, 25.5, 24.9, 23.6, 24.4, 23.6, 25.1, 22.8, 20.9, 21.2

a) ¿Qué valor de la temperatura promedio supera la primera cuarta parte de ellas?

b) ¿Qué valor de la temperatura promedio que supera el cuarenta por ciento de ellas?

c) ¿Qué valor de la temperatura promedio es superado por el treinta y cinco por ciento de ellas?

Solución

Datos ordenados de manera ascendente: 20.9, 21.2, 22.8, 23.6, 23.6, 23.8, 23.8, 24.1, 24.4, 24.4, 24.4, 24.4, 24.5, 24.5, 24.6, 24.6, 24.7, 24.8 ,24.9, 25.0, 25.0, 25.1, 25.1, 25.4, 25.4, 25.5, 25.5, 25.6, 25.6, 25.8

Solución a) Corresponde a Q_1 o P_{25}

Calculando Q_1

En este caso r = 1 k = 4 y n = 30

Posición $\longrightarrow \frac{r*n}{k}$

Encontrando la posición de Q_1: $\frac{1*30}{4}$= 7.50, como este valor no es un número entero se aproxima al siguiente entero más próximo; entonces en la posición 8 se encuentra el Q_1 contada del menor valor. Es decir, Q_1 =24.1^{0}C. Por lo tanto, de los 30 días que tiene el mes de noviembre, el 25% de los ellos se tuvo una temperatura promedio menor a 24.1^{0}C y el 75% de los días una temperatura promedio mayor a 24.1^{0}C.

Solución b) Corresponde a D_4 o P_{40}

Calculando D_4

En este caso r = 4 k = 10 y n = 30

Posición $\longrightarrow \frac{r*n}{k}$

Encontrando la posición de D_4: $\frac{4*30}{10} = 12$, como este valor es un número entero, entonces D_4 se encuentra a la mitad de la posición 12 y la posición 13. Entonces corresponde al promedio de los valores de la posición 12 y 13. Es decir,

$$D_4 = \frac{24.4 + 24.5}{2} = 24.45\,^0\text{C}$$

Por lo tanto, de los 30 días que tiene el mes de noviembre, el 40% de los ellos se tuvo una temperatura promedio menor a 24.45⁰C y el 60% de los días una temperatura promedio mayor a 24.45⁰C.

Solución c) Corresponde al P_{65}

En este caso r = 65 k = 100 y n = 30

Posición $\longrightarrow \frac{r*n}{k}$

Encontrando la posición de P_{65}: $\frac{65*30}{100} = 19.5$, como este valor no es un número entero se aproxima al siguiente entero más próximo, entonces la posición 20 es donde se encuentra el valor del P_{65} contada a partir del menor valor. Es decir, P_{65} = 25.0 ⁰C.

Por lo tanto, de los 30 días que tiene el mes de noviembre, el 65% de los ellos se tuvo una temperatura promedio menor a 25.0 ⁰C y el 35% de los días una temperatura promedio mayor a 25.0 ⁰C.

Caso 2: Datos agrupados en tablas de frecuencias.

El procedimiento para determinar el valor de cualquier cuantil es similar al que se utilizó para el cálculo de la mediana y se presenta a continuación.

En términos de frecuencias Absolutas Acumuladas.

1. Obtener la tabla de frecuencias absolutas y relativas acumuladas.
2. Calcular posición $\longrightarrow \frac{r*n}{k}$

Donde:

r: Número del cuartil, decil o percentil a calcular.

n: Cantidad de datos o tamaño de la muestra.

k: Número de partes en que el cuantil divide al conjunto de datos.

Valores	Cuartiles	Deciles	Percentiles
r	1,2,3	1,2,3,4,5,6,7,8,9	1,2,3,.......,99
k	4	10	100

Tabla 2: Cálculo para frecuencias absolutas acumuladas

3. Comparar el valor de $\frac{r*n}{k}$ con los valores de la frecuencia absoluta acumulada (N_i):

- Para encontrar el lugar que ocupa el cuantil, se busca en la columna de frecuencias absolutas acumuladas el valor que sea igual o inmediatamente superior a $\frac{r*n}{k}$; luego el cuantil buscado será el valor X_i de la variable que corresponde a la frecuencia absoluta acumulada N_i. Es decir el cuantil buscado es X_i que corresponde a N_i si

$$N_{i-1} < \frac{r*n}{k} < N_i$$

- Si un valor de la frecuencia absoluta acumulada coincide con $\frac{r*n}{k}$, entonces el cuantil será el promedio del valor de X_i que corresponde a N_i y el valor siguiente X_{i+1}. Es decir el cuantil buscado es $\frac{X_i + X_{i+1}}{2}$ si $\frac{r*n}{k} = N_i$

En términos de Frecuencias Relativas Acumuladas.

1. Obtener la tabla de frecuencias absolutas y relativas acumuladas.
2. Calcular la posición $\longrightarrow \frac{r}{k}$.

Donde:

r: Número del cuartil, decil o percentil a calcular.

k: Número de partes en que el cuantil divide al conjunto de datos.

Valores	Cuartiles	Deciles	Percentiles
R	1,2,3	1,2,3,4,5,6,7,8,9	1,2,3,……,99
K	4	10	100

Tabla 3: Cálculo para frecuencias relativas acumuladas

1. Comparar el valor de $\frac{r}{k}$ con los valores de la frecuencia relativas acumulada (F_i):

- Para encontrar el lugar que ocupa el cuantil se busca en la columna de frecuencias relativas acumuladas (F_i), el valor que sea igual o inmediatamente superior a $\frac{r}{k}$, entonces el cuantil será el valor de la variable X_i que corresponde a F_i . Es decir el cuantil buscado es X_i si $F_{i-1} < \frac{r}{k} < F_i$
- Si un valor de la frecuencia relativa acumulada (F_i) coincide con $\frac{r}{k}$, entonces el cuantil será el promedio del valor de X_i que corresponde a (F_i) y el valor siguientes X_{i+1}.

  Es decir el cuantil buscada es $\frac{X_i + X_{i+1}}{2}$ si $\frac{r}{k} = F_i$

Ejemplo 2: Retomando el ejemplo 14 de la unidad anterior: Se les pregunto a 60 alumnos de primer año de bachillerato general, el número de asignaturas reprobadas en el primer trimestre de este año; y obtuvo la siguiente tabla de distribución de frecuencias

Número Asignaturas Reprobadas x_i	Cantidad de Alumnos n_i	N_i
0	8	8
1	11	19
2	13	32
3	15	47
4	10	57
5	3	60
n	60	

a) Encontrar el número de asignaturas reprobadas que supera y es superado por la mitad de los alumnos.
b) Encontrar el quinto decil y realizar su respectiva interpretación.
c) Encontrar el número de asignaturas reprobadas que es superada por el quince por ciento de los alumnos.

Solución

En términos de frecuencias absolutas acumuladas.

Solución a) Corresponde a Q_2 o D_5 o P_{50}

Calculando Q_2

En este caso: r=2 k=4 y n=60
Posición $\longrightarrow \frac{r*n}{k}$

Encontrando la posición de Q_2: $\frac{2*60}{4} = 30$, al comparar este valor en la columna de N_i se observa que $19 < 30 < 32$, entonces $Q_2 = 2$; ya que $N_3 = 32$ es el primer valor que supera a 30.

Por lo tanto, el 50% de los alumnos han reprobado dos o menos asignaturas y el otro 50% ha reprobado dos o más asignaturas.

Solución b) Corresponde a D_5

En este caso: r=5 k=10 y n=60
Posición $\longrightarrow \frac{r*n}{k}$

Encontrando la posición de D_5: $\frac{5*60}{10} = 30$, al comparar este valor en la columna de N_i se observa que $19 < 30 < 32$, entonces $D_5 = 2$; ya que $N_3=32$ es el primer valor que supera a 30.

Por lo tanto, el 50% de los alumnos han reprobado dos o menos asignaturas y el otro 50% ha reprobado dos o más asignaturas (igual que en el literal anterior).

Solución c) Corresponde al P_{85}

En este caso: r = 85 k = 100 y n = 60

Posición $\longrightarrow \frac{r*n}{k}$

Encontrando la posición de P_{85}: $\frac{85*60}{100} = 51$, al comparar este valor en la columna de N_i se observa que $47 < 51 < 57$, entonces $P_{85} = 4$; ya que $N_5=57$ es el primer valor que supera a 51.

Por lo tanto, el 85% de los alumnos han reprobado cuatro o menos asignaturas y el otro 15% ha reprobado cuatro o más asignaturas.

Caso 3: Datos agrupados en tablas de frecuencia por intervalos.

El procedimiento para determinar el valor de cualquier cuantil es similar al que se utilizó en el caso anterior, con la diferencia que en un primer momento se encontrara el intervalo

$[l_{i-1}, l_i[$, donde se encuentra el cuantil y se presenta a continuación.

En términos de frecuencias Absolutas Acumuladas.

1. Obtener la tabla agrupada de frecuencias absolutas y relativas acumuladas.
2. Calcular la posición $\longrightarrow \frac{r*n}{k}$

Donde:

r: Número del cuartil, decil o percentil a calcular.

n: Cantidad de datos o tamaño de la muestra.

k: Número de partes en que el cuantil divide al conjunto de datos.

Valores	Cuartiles	Deciles	Percentiles
r	1,2,3	1,2,3,4,5,6,7,8,9	1,2,3,……..,99
k	4	10	100

Tabla 4: Cálculo para frecuencias absolutas acumuladas

1. Comparar el valor de $\frac{r*n}{k}$ con los valores de la frecuencia absoluta acumulada N_i:

- Para encontrar el intervalo $[l_{i-1}, l_i[$, donde se encuentra el cuantil, se busca en la columna de las frecuencias absolutas acumuladas (N_i) el valor que sea igual o inmediatamente superior a $\frac{r*n}{k}$.
- Si $N_{i-1} < \frac{r*n}{k} \leq N_i$, entonces el cuantil se encuentra en $[l_{i-1}, l_i[$ que corresponde a N_i y luego se utiliza la siguiente fórmula que representa el cuantil r de orden k:

$$Q_{r/k} = L_{i-1} + \frac{\frac{r*n}{k} - N_{i-1}}{n_i} * c_i$$

Donde:

Li-1: Límite inferior del intervalo donde se ubica el cuantil.

N : Número total de observaciones.

Ni-1: Frecuencia absoluta acumulada del intervalo anterior de donde se ubica el cuantil.

Ni : Frecuencia absoluta del intervalo donde se ubica el cuantil.

Ci : Amplitud del intervalo donde se ubica el cuantil.

Si Ni-1 $= \frac{r*n}{k} <$ Ni entonces el cuantil se encuentra en [li-1,li[que corresponde a Ni y el cuantil buscados es igual a li-1. Es decir,

En términos de frecuencias relativas acumuladas.

1. Obtener la tabla de frecuencias absolutas y relativas acumuladas.
2. Calcular la posición $\longrightarrow \frac{r}{k}$

Donde:

r : Número del cuartil, decil o percentil a calcular

k : Número de partes en que el cuantil divide al conjunto de datos.

Valores	Cuartiles	Deciles	Percentiles
R	1,2,3	1,2,3,4,5,6,7,8,9	1,2,3,……..,99
K	4	10	100

Tabla 5: Cálculo para frecuencias relativas acumuladas

Comparar el valor de $\frac{r}{k}$ con los valores de la frecuencia relativas acumuladas F_i.

Para encontrar el intervalo $[l_{i-1},l_i[$ donde se encuentra el cuantil, se busca en las frecuencias relativas acumuladas (F_i), el valor que sea igual o inmediatamente superior a $\frac{r}{k}$.

Si $F_{i-1} < \frac{r}{k} \leq F_i$, entonces el cuantil se encuentra en $[l_{i-1},l_i[$ que corresponde a F_i y luego se utiliza la siguiente fórmula que representa el cuantil r de orden k:

$$Q_{r/k} = L_{i-1} + \frac{\frac{r}{k} - F_{i-1}}{f_i} * c_i$$

Donde:

L_{i-1}: Límite inferior del intervalo donde se ubica el cuantil.

F_{i-1}: Frecuencia relativa acumulada del intervalo anterior de donde se ubica el cuantil.

f_i: Frecuencia relativa del intervalo donde se ubica el cuantil.

c_i : Amplitud del intervalo donde se ubica el cuantil.

- **Si** $F_{i-1} = \frac{r}{k} < F_i$ entonces el cuantil se encuentra en $[l_{i-1},l_i[$ que corresponde a F_i y el cuantil buscado es igual a l_{i-1}.
  Es decir, $Q_{r/k} = L_{i-1}$

Ejemplo 3: En un programa de autocontrol personal del peso, aplicado a 90 personas, los kilogramos que estas perdieron al terminar dicho programa se muestran en la siguiente tabla.

Peso Perdido X_i	Número de personas n_i	N_i	f_i	F_i
5 - 9	9	9	0.1	0.1
10 - 14	19	28	0.21	0.31
15-19	33	61	0.37	0.68
20 – 24	15	76	0.17	0.85
25 – 29	10	86	0.11	0.97
30 -34	2	88	0.02	0.98
35 - 39	0	88	0	0.98
40 - 44	2	90	0.02	1
n	90		1	

a) ¿Cuál es el valor del peso perdido que supera la tercera parte de las personas?
b) ¿Cuál es el valor del peso perdido que es superado por el setenta por ciento de las personas?
c) ¿Cuál es el valor del peso perdido que supera el noventa por ciento de las personas?

Solución

Solución a) Corresponde al Q_3 o P_{75}

Calculando Q_3

En términos de frecuencias absolutas acumuladas.

En este caso: r=3 k=4 y n=90
Posición ⟶ $\frac{r*n}{k}$

Encontrando la posición de Q_3: $\frac{3*90}{4} = 67.5$, al comparar este valor en la columna de N_i se observa que $61 < 67.25 < 76$, entonces C_3 se encuentra en el intervalo $[20,24]$; ya que $N_4=76$ es el primer valor que supera a 67.25.

Entonces: $N_3 = 61$ $n_4=15$ $c_4=5$ $L_3= 20$

Sustituyendo se tiene:

$$Q_3 = 20 + \frac{67.25-61}{15} * 5 = 20 + 2.08$$
$$Q_3 = 22.08$$

Por lo tanto, el 75% de las personas han perdido 22.08 kg o menos del peso en el programa, y el 25% de las personas han perdido 22.8 kg o más del peso en el programa.

Solución b) Corresponde al D_3 o P_{75}

Calculando D_3

En términos de frecuencias relativas acumuladas.

En este caso: r=3 k=10 y n=90

Posición ⟶ $\frac{r}{k}$

Encontrando la posición de D_3: $\frac{3}{10} = 0.3$, al comparar este valor en la columna de F_i se observa que $0.1 < 0.30 < 0.31-$, entonces D_3 se encuentra en el intervalo $[10,14]$; ya que $F_2=0.31$ es el primer valor que supera a 0.30.

Entonces: $L_1= 10$ $F_1= 0.1$ $f_2=0.21$ $c_1= 5$

Sustituyendo se tiene:

$$D_3 = 10 + \frac{\frac{3}{10} - 0.1}{0.21} * 5 = 10 + 4.76$$

$$D_3 = 14.76$$

Por lo tanto, el 30% de las personas han perdido 14.76 kg o menos del peso en el programa y el 70% de las personas han perdido 14.76 kg o más del peso en el programa.

Solución c) Corresponde a P_{90} o D_9

Calculando P_{90}

En términos de frecuencias absolutas acumuladas.

En este caso: r=90 k=100 y n=90

Encontrando la posición P_{90}: $\frac{90*90}{100} = 81$, al comparar este valor en la columna de N_i se observa que $76 < 81 < 86$, entonces P_{90} se encuentra en el intervalo $[25,29]$; ya que $N_5=86$ es el primer valor que supera a 81.

Entonces: $N_4 = 76$ $n_5=10$ $c_5=5$ $L_4= 25$

Sustituyendo se tiene:

$$P_{90} = 25 + \frac{81-76}{10} * 5 = 25 + 2.5$$

$$P_{90} = 27.5$$

Por lo tanto, el 90% de las personas han perdido 27.50 kg o menos del peso en el programa y el 10% de las personas han perdido 27.50 kg o más del peso en el programa.

PROCEDIMIENTO PARA CALCULAR LOS CUANTILES

(Datos agrupado en tablas de frecuencias)

1. Encontrar las frecuencias absolutas y relativas acumuladas.
2. Determinar la posición del cuantil a calcular.
3. Identificar la posición del cuantil en el valor o próximo mayor de la frecuencia absoluta o relativa acumulada.
4. Obtener los datos respectivos del intervalo o valor correspondiente de la variable.

5. Aplicar fórmula para obtener el cuantil deseado.

PROCEDIMIENTO PARA CALCULAR LOS CUANTILES

Datos sin agrupar

1. Ordenar los valores de la variable o datos de menor a mayor (orden ascendente).
2. Determinar la posición del cuantil a calcular.
3. Identificar la posición del cuantil según orden de los datos.
4. Obtener el valor del cuantil buscado.

UTILIDAD DE LOS CUANTILES

Los cuartiles se utilizan:

- Para identificar el porcentaje igual o menor que el valor de un cuartil.
- Para construir la curva endémica.
- Para describir el 50% central de las observaciones.
- Elaboración de gráficos de caja.

Los percentiles se utilizan:

- Para comparar un valor de un individuo con un conjunto de normas
- Para determinar rangos normales de análisis de laboratorio, los límites normales de muchos análisis se ubican entre el percentil 2.5 y 97.5
- También se usa para establecer el rango intercuartílico.

APLICANDO LO APRENDIDO

1. En la columna vacía escriba una C o una I si el enunciado es correcto o incorrecto.

1	El cuartil 2 divide a la serie en dos partes iguales	
2	El decil 5 de la siguiente serie: 18,17,15,14,13,12 es 14	
3	El Percentil 50 de la serie anterior es 3.5	
4	El cuartil 3 de la serie anterior es 12.5	
5	El valor del percentil 80 es igual al decil 8	
6	El cuartil 3 es diferente al percentil 75	
7	El valor de la mediana es igual al cuartil 2	
8	El valor de decil 6 es igual al percentil 6	
9	E valor de la mediana es igual al percentil 50	
10	El percentil 99 deja a la derecha un 10%	

2. En una colonia de San Salvador se investigo sobre el número de horas que 60 niños ven televisión diariamente. En la siguiente tabla se presenta la información de esta investigación.

No. Horas	1	2	3	4	5	6
No. Niños	10	12	15	8	6	9

a) ¿Cuántas horas ven televisión diariamente la mitad de los niños? R/ 3 horas

b) ¿Cuál es el número de horas que ven televisión diariamente la cuarta parte de los niños? R/2 horas
c) ¿Cuál es el número de horas que ven televisión diariamente los niños que supera el cuarenta por ciento de ellos? R/3 horas
d) ¿Cuál es el número de horas que ven televisión diariamente los niños que es superado por el treinta por ciento de ellos? R/4 horas
e) ¿Cuántas horas ven televisión diariamente los niños que es superado por el treinta y ocho por ciento de ellos? R/4 horas

3. En la siguiente tabla se encuentran registrados el número de lesionados por accidentes de tránsito registrados en San Salvador durante el primer semestre del año 2008, organizados por rango de edades reportados por CNIP policía nacional civil.

Edad	[0 – 11]	[12 – 17]	[18 – 25]	[26 – 30]	[31 – 40]	[41 – 50]	[51-60]
No. Lesionados	16	19	32	27	39	18	7

a) ¿Cuál es la edad que supera las tres cuartas partes de los lesionados?
b) Encontrar la edad que supera a la mitad de los lesionados
c) ¿Cuál es la edad que supera los ochenta por ciento de los lesionados?
d) ¿Cuál es la edad que es superada por el cincuenta y cinco de los lesionados?

En la siguiente tabla se encuentran registrados el número de fallecidos por accidentes de tránsito registrados en Sonsonate durante el primer semestre del año 2008, organizados por rango de edades reportados por CNIP policía nacional civil.

Edad	[0 – 11]	[12 – 17]	[18 – 25]	[26 – 30]	[31 – 40]	[41 – 50]	[51-60]
No. fallecidos	4	3	9	4	14	10	6

e) Calcular el cuartil 3 e interprételo.
f) ¿Cuál es la edad que supera las tres cuartas partes de los fallecidos? Compare con el resultado de b) que observa.
g) Encontrar la edad que es superada por el noventa por ciento de los fallecidos
h) ¿Cuál es la edad que supera el setenta por ciento de los fallecidos?

4. En una competencia de tiro olímpico, 30 tiradores han obtenido las siguientes puntuaciones: 10, 9, 6, 8, 9, 5, 3, 8, 9, 7, 10, 10, 9, 6, 8, 7, 6, 10, 9, 8, 5, 3, 1, 8, 8, 9, 7, 8, 9 y 10.
a) ¿Cuál es la puntuación mediana?
b) Hallar el percentil 30 y 60 e interprételo.
c) ¿Qué valor toma la puntuación que es superada por el veinte por ciento?
d) Encontrar el valor de la puntuación que supera las tres terceras partes de los datos.

5. Las edades de veinte jóvenes son: 12, 13, 14, 10, 11, 12, 11, 13, 14, 12, 10, 12, 11, 13, 12, 11, 13, 12, 10 y 15. Organiza los datos en una tabla de frecuencias y calcula:

1. El cuartil 1

2. Los deciles 1 y 6
3. Los percentiles 35 y 80.

6. El número de turistas que visitaron un parque de diversiones en distintas fechas es: 12, 14, 17, 16, 19, 15, 15, 21, 24, 26, 28, 24, 25, 26, 20, 21, 34, 35, 33, 32, 34, 38, 40, 43, 41, 45, 50, 53, 58. Calcular e interpretar:

   a) Los cuartiles 2 y 3   b) Los deciles 2 y 7   c) Los percentiles 35, 60 y 95

7. A los 50 alumnos de primer año de bachillerato se les realizó una prueba de matemática y se obtuvieron los siguientes resultados.

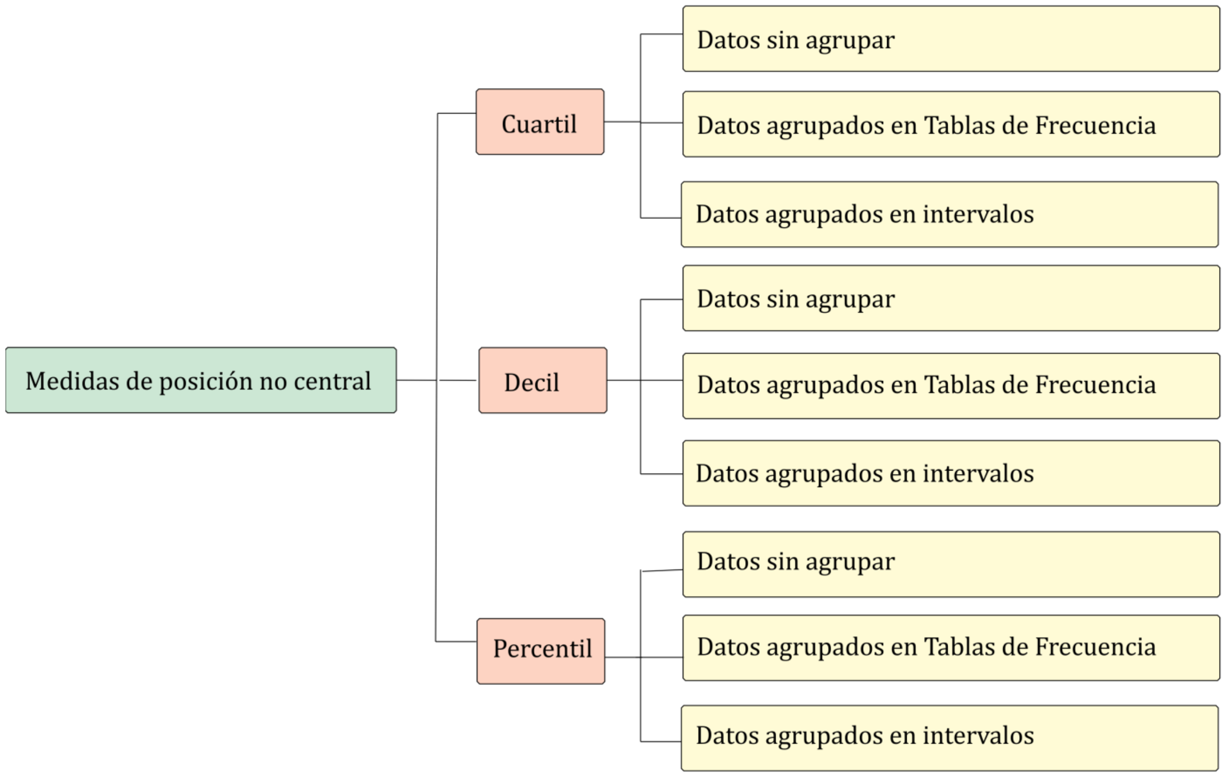
Puntajes (Xi)	Número de alumnos (ni)
[60 – 65[	5
[65 – 70[	5
[70 – 75[	8
[75 – 80[	12
[80 – 85[	16
[85 – 90[	4
n	50

   a) Calcular e interpretar: Q_1, D_4, P_{65} y P_{80}.
   b) El puntaje mínimo del 25% que obtuvo los mejores resultados.
   c) El puntaje mínimo del 10% que obtuvo los mejores resultados y ganará una disminución de su cuota escolar.
   d) El puntaje que debe superar el 20% que obtuvo las notas más bajas, para no asistir a un taller de refuerzo.
   e) El puntaje que separa la serie en dos partes iguales (50% inferior y 50% superior).

BIBLIOGRAFIA

1. Johnson R., Cuby P.(1999), *Estadística Elemental*. México: Internacional Thomson Editores, S.A de S.V.
2. Pérez C., (2003).*Estadística. Problemas Resueltos y Aplicaciones*. Madrid: Pearson Educación, S.A.
3. Pérez-T. H.E (2007), Estadística para las Ciencias Sociales, del Comportamiento y de la Salud. (3ª. ed.)México: Impreso Edamsa Impresiones, S.A. de C.V.
4. Sarabia J.M. (2000), *Curso Práctico de Estadística*. (2da ed.). España: Impreso por Gráficas Rogar, S.A Navalcarnero (Madrid).
5. Triola, M., (2009). *Estadística.* (10a ed.). México: Pearson Educación, S.A.

DIAGRAMA DE CONTENIDO



Lección 5 **Primer año de Bachillerato** **Unidad VIII** **Tiempo: 8 horas clase**

Medidas de Variabilidad

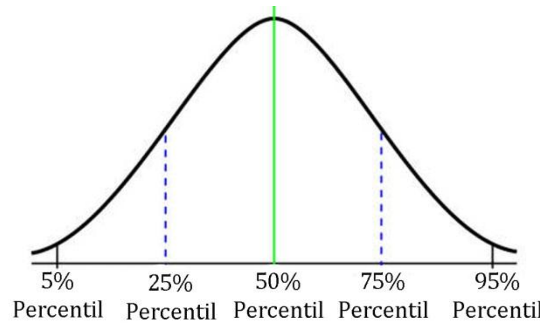


Figura 1. El Salvador. Matrimonios, por sexo, según grupos de edad: 2006.
Fuente:DIGESTYC

Introducción

Las medidas de tendencia central o posición indican donde se sitúa un conjunto de valores. Sin embargo, los de variabilidad o dispersión indican si estos valores están próximás entre sí o si por el contrario están alejados o muy dispersos. Por consiguiente, además de las medidas de tendencia central, siempre es importante contar con indicadores que midan la dispersión de los datos. Una medida de tendencia central, casi nunca es suficiente por sí sola, para resumir adecuadamente las características de un conjunto de datos. Por lo general, es necesario, adicionalmente, una medida de la *dispersión* de los datos para comprender el comportamiento y además, conocer que tan alejados están esos datos respecto a ese punto de concentración.

Entre las medidas de dispersión más utilizadas se encuentran: el rango, la desviación media, varianza, la desviación estándar y coeficiente de variabilidad. Muchas veces se suele diferenciar entre medidas de dispersión absolutas y relativas. Las medidas de dispersión absolutas son aquellas que vienen expresadas en las mismas unidades que los datos; mientras que las medidas de dispersión relativas no vienen expresadas en las unidades de los datos sino en porcentaje.

Estas medidas de dispersión indican la distancia promedio de los datos respecto a las medidas de tendencia central. Así se podrá diferenciar dos conjuntos de datos que poseen iguales medias, siendo los datos de uno más dispersos del otro.

Objetivos

- Identificar la variabilidad existente en la naturaleza.
- Calcular las medidas de dispersión más importantes.
- Diferenciar las medidas de dispersión, y calcularlas estando agrupadas en intervalos.
- Interpretar las medidas de dispersión.

Importancia

Medir la variabilidad resulta muy importante en diversas situaciones prácticas, pues a través de su medición se podrá establecer cuando existe una mayor concentración de ellos en la región central. Así por ejemplo, en estudios sociales las medidas de dispersión proporcionan la información requerida para analizar cómo es la distribución de los ingresos dentro de la sociedad; en los estudios de calidad industrial, estas mismas medidas de dispersión se utilizan para medir la precisión de las máquinas utilizadas en el proceso de producción. También en la metodología seis sigma (σ): centrada en la reducción de la variabilidad de los procesos, se consigue reducir o eliminar los defectos o fallas en la entrega de un producto o servicio al cliente.

MEDIDAS DE VARIABILIDAD

Los estudios estadísticos permiten hacer inferencias de una característica de la población a partir de la información contenida en una muestra. Los métodos numéricos que describen a los conjuntos de datos tienen como objetivo dar una imagen mental de la distribución de frecuencias.

No sólo basta con determinar las medidas de tendencia central para comprender el comportamiento de una serie de datos, es importante además, conocer que tan alejados están esos datos respecto a ese punto de concentración.

Un promedio puede ser engañoso a menos que sea identificado y vaya acompañado por otra información que informe las desviaciones de los datos respecto a la medida de tendencia central seleccionada. Una vez localizado el centro de la distribución de un conjunto de datos, lo que procede es buscar una medida de dispersión de los datos.

La variación o dispersión de un conjunto se refiere a la variedad que exhiben las observaciones, si todos los valores son iguales no hay dispersión, de forma contraria si no todos los valores son iguales existe dispersión de los datos. La dispersión será pequeña cuando los valores estén próximos entre sí y será muy grande si los valores se encuentran ampliamente diseminados.

Así, cuanto menor es la variabilidad, más homogénea es la muestra de sujetos en la variable. En el caso de máxima homogeneidad, todos los valores de la variable serán iguales. De otro modo, cuanta más o menos dispersión en los datos, la muestra es más o menos heterogénea y las puntuaciones difieren entre sí.

Competencias a reforzar.

Calcula, analiza e interpreta las medidas de dispersión, para tomar decisiones acertadas ante una situación real.

Presaberes

- Operaciones básicas
- Potencia de orden 2
- Raíz cuadrada
- Valor absoluto de un número

Las medidas de dispersión nos indican, la distancia promedio de los datos respecto a las medidas de tendencia central. Así se podrá diferenciar dos conjuntos de datos que poseen iguales medias, siendo los datos de uno más dispersos del otro.

Las medidas de tendencia central por sí solas carecen de significado, pues de nada sirve saber el promedio sin conocer la dispersión, qué significa esto, saber cuánto se alejan las observaciones de su propio promedio

Aunque "dispersión" y "concentración" tengan significados opuestos en el lenguaje coloquial, en estadística no coincide el concepto de concentración con la acepción normal del vocablo.

La "dispersión" hace referencia a la variabilidad de los datos, a las diferencias existentes entre ellos y la representatividad de los promedios.

La "concentración", por su parte, se refiere al mayor o menor grado de igualdad en el reparto de todos los valores de la variable.

Cuando la dispersión es alta, es decir, cuando los valores se separan mucho entre sí, el promedio se vuelve de poco significativo. Por el contrario, si la dispersión es baja, el promedio es representativo del conjunto de datos.

En la vida cotidiana se puede hacer referencia a este tipo de medidas, por ejemplo, para describir el perfil de una colina o montaña que se observa a través de la ventana, no será suficiente fijar la mirada en la altura de la colina, sino que también es necesario observar su forma, es decir, si es más bien plana o empinada, si es más bien simétrica o no, si tiene una o varias cumbres y dónde, etc.

Desde este punto de vista, una distribución no es diferente a una colina.

Figura 2. Colinas

La variabilidad de un conjunto de datos puede medirse a través de las siguientes medidas: Rango, Desviación media, Varianza, Desviación estándar y el Coeficiente de variación, de estos los más usados son la varianza, desviación estándar y el coeficiente de variación. El cálculo de cada uno de ellos se toma basado en la media aritmética.

Por ejemplo, si se tienen dos localidades, A y B, con la misma renta media por habitante. ¿Nos permitirá este simple hecho de igualdad de los dos medios concluir que la situación económica de las dos localidades es la misma?

Realmente no, pues esta igualdad podría existir aún si en A fuese perfectamente estabilizada, en el sentido que todos sus habitantes tuviesen prácticamente la misma renta y B tuviese unos pocos individuos con renta extraordinariamente alta y la mayoría con rentas muy bajas.

Con este ejemplo tan sencillo, basta para indicar que el conocimiento de la intensidad de los valores asumidos por una grandeza, es decir, de las medidas de posición de una distribución, no es suficiente para su completa caracterización. El hecho de que en A todos los individuos tienen la misma renta, es sinónimo de que en A las rentas no varían de un individuo para otro, o aún más, que la distribución de rentas no presenta variabilidad.

Análogamente, el hecho de que en B algunos individuos tienen rentas muy elevadas en detrimento de la gran mayoría que tienen rentas muy bajas, puede ser expresado diciendo que en B las rentas varían o que la distribución de las rentas presenta variabilidad

ACTIVIDAD INTRODUCTORIA

Si se tienen 9 rectángulos cuya altura es de 8 centímetros (y todos tienen la misma base).

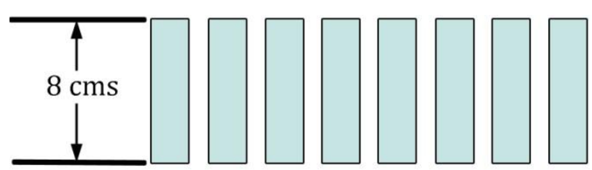


Figura 3. Rectángulos con longitudes constantes

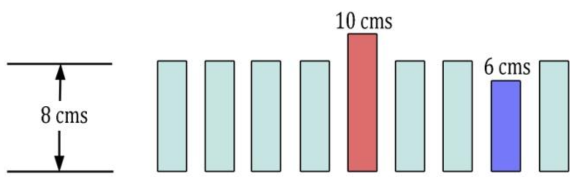
¿Existe alguna variación respecto de su altura entre estos rectángulos?

¿Cuál es el promedio de la altura de estos rectángulos?

Calculando el promedio:

$$\frac{8+8+8+8+8+8+8+8+8}{9}=\frac{72}{9}=8\ \text{cms}$$

Obsérvese ahora el quinto rectángulo y el octavo rectángulo en un acto de rebeldía cambiaron su altura. El quinto rectángulo, ahora de color rojo, mide 10 centímetros, y el octavo rectángulo, de color azul, mide 6 centímetros.



¿Cuál es el nuevo promedio de estos 9 rectángulos?

Calculando el promedio:

$$\frac{8+8+8+8+10+8+8+6+8}{9}=\frac{72}{9}=8\ \text{cms}$$

... ¡Se obtiene el mismo promedio! Pero...... ¿ha habido variación?

El rectángulo rojo tiene (+2 centímetros) sobre el promedio, y el rectángulo azul tiene (–2 centímetros) bajo el promedio. Los otros rectángulos tienen cero de diferencia respecto del promedio.

Si sumamos estas diferencias de la altura respecto del promedio, se tiene:

$0 + 0 + 0 + 0 + 2 + 0 + 0 - 2 + 0 = 0$

Este valor parece indicar que ¡no ha habido variabilidad! Y sin embargo, ante nuestros ojos, se sabe que hay variación.

Una forma de eliminar los signos menos de aquellas diferencias que sean negativas, esto es de aquellos mediciones que estén bajo el promedio, es elevar al cuadrado todas las diferencias, y luego sumar... $0^2 + 0^2 + 0^2 + 0^2 + 2^2 + 0^2 + 0^2 + (-2)^2 + 0^2 = 8$

Y este resultado repartirlo entre todos los rectángulos, es decir lo dividimos por el número de rectángulos que es 9. Es decir:

$$\frac{0^2+0^2+0^2+0^2+2^2+0^2+0^2+(-2)^2+0^2}{9}$$

$$=\frac{8}{9}=0.89\ \text{cms}^2$$

Se dice entonces que la *varianza* fue de 0,89. Observemos que las unidades involucradas en el cálculo de la varianza están al cuadrado. En rigor la varianza es de 0,89 centímetros cuadrados.

De manera que se define:

$$\sqrt{0.89cms^2}=0.943\ \text{cms}$$

La raíz cuadrada de la varianza se llama ***desviación estándar.***

Que la desviación estándar haya sido de 0,943 significa que en promedio la altura de los rectángulos variaron (ya sea aumentando, ya sea disminuyendo) en 0,943 centímetros.

Es claro que esta situación es "en promedio", puesto que sabemos que los causantes de la variación fueron los rectángulos quinto y octavo.

Esta variación hace repartir la "culpa" a todos los demás rectángulos que se "portaron bien". La desviación estándar mide la dispersión de los datos respecto del promedio.

RANGO O AMPLITUD

La amplitud o rango de un conjunto de valores de una variable se define como la distancia que hay en la escala numérica entre los valores que representa el valor máximo y mínimo. Es decir, la diferencia entre el mayor y el menor valor de los datos. Y se simboliza por R.

- **Caso 1: Datos sin agrupar**
  $R = X_{Máximo} - X_{Mínimo}$
- **Caso 2: Datos agrupados en tablas de frecuencias.**
  $R = X_{último} - X_{primer}$
- **Caso 3: Datos agrupados en tablas de frecuencia por intervalos.**

Cuando se tienen intervalos en una tabla de frecuencias, se hace un cálculo aproximado usando el límite inferior del intervalo de clase menor (Primera clase) y el límite superior del intervalo de clase más alto (Última clase).

Uno de los inconvenientes es que únicamente utilizar los valores extremos del conjunto de datos; de esta forma, no recoge la poca o mucha dispersión que pueda existir entre los restantes valores, que son la mayoría. La principal desventaja del rango es que al basarse su cálculo en los valores mínimo y máximo, si la distribución tiene valores atípicos, su cálculo se verá muy influido por los mismos.

Es muy fácil de calcular, pero es poca su utilidad.

La única utilidad es en la construcción de tablas de distribuciones de frecuencias en la cual los datos están organizados en intervalos.

En lo que respecta a la interpretación del rango, tanto éste como el resto de índices de variabilidad ofrecen resultados que no tienen una interpretación directa en términos absolutos. ¿Qué significa un rango de 4 o un rango de 10, mucha o poca dispersión?

El único caso en que la interpretación de estos índices es absoluta es cuando dan igual a 0, indicando la ausencia de variabilidad en los datos de la variable que es un caso excepcional; y valores mayores que 0 indicarán dispersión en los datos, tanto mayor cuanto mayor sea ese valor, pero sin existir un techo que nos permita establecer interpretaciones en términos absolutos. La interpretación de estos índices depende de la naturaleza de la variable considerada y de la escala utilizada al ser medida.

Pero es posible realizar interpretaciones en términos relativos, por ejemplo, establecer en dos muestras de las que se tiene datos en una misma variable, cuál de los dos tiene una mayor dispersión en sus datos o, también, comparar la dispersión de los datos de una misma variable medida en dos momentos temporales distintos.

Ejemplo 1: Se desea comparar la variabilidad de las temperaturas promedios registradas en el mes de octubre y noviembre del año 2011, reportadas por la estación meteorológica: 786630 (MSSS) con sede en Ilopango.

Temperaturas registradas en el mes de Octubre: 21.1, 21.3, 21.7, 21.7, 21.8, 21.8, 21.8, 22.1, 22.2, 22.3, 22.9, 23.0, 23.3, 23.4, 23.4, 23.4, 24.1, 24.2, 24.3, 24.4, 24.4, 24.4, 24.5, 24.7, 24.7, 24.7, 25.1, 25.1, 25.1, 25.3, 25.5.

Cálculo del rango: $R = X_{Máximo} - X_{Mínimo}$

$R = 25.5 - 21.1 = 4.4\ ^{0}C$

Significa que las temperaturas promedio registradas durante el mes de octubre se encuentran en un intervalo de 4.4 ^{0}C .

Temperaturas registradas en el mes de Noviembre: 20.9, 21.2, 22.8, 23.6, 23.6, 23.8, 23.8, 24.1, 24.4, 24.4, 24.4, 24.4, 24.5, 24.5, 24.6, 24.6, 24.7, 24.8 ,24.9, 25.0, 25.0, 25.1, 25.1, 25.4, 25.4, 25.5, 25.5, 25.6, 25.6, 25.8

Cálculo del rango: $R=X_{Máximo} - X_{Mínimo}$

$R= 25.8 - 20.9 = 4.9\ ^0C$

Significa que las temperaturas promedio registradas durante el mes de noviembre se encuentran en un intervalo de 4.9 ^{0}C

Por lo tanto, se observa mayor variabilidad de las temperaturas promedios en el mes de noviembre.

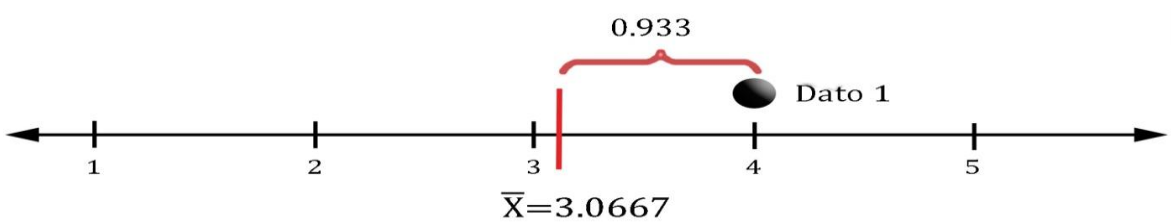
Características del rango

1. A medida que el rango es menor, el grado de representatividad de los valores centrales se incrementa.
2. A medida que el rango es mayor, la distribución está menos concentrada o más dispersa.
3. Su cálculo es extremadamente sencillo.
4. Tiene gran aplicación en procesos de control de calidad.
5. Tiene el inconveniente de que sólo depende de los valores extremos. De esta forma basta que uno de ellos se separe mucho para que el recorrido se vea sensiblemente afectado.

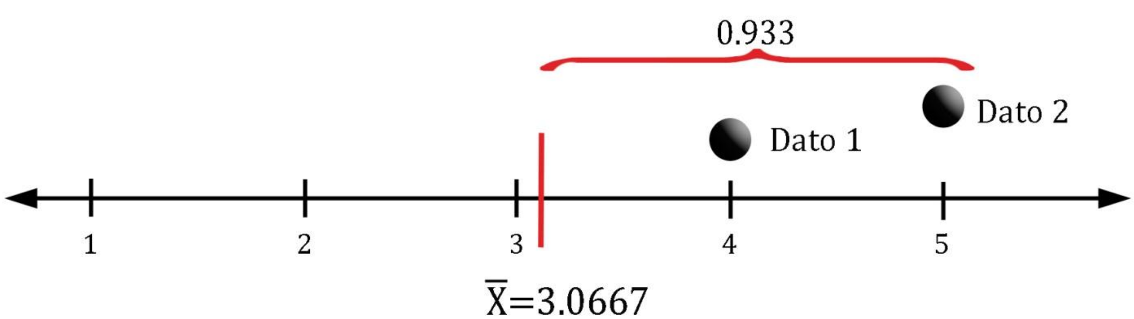
DESVIACIÓN MEDIA

Para conocer con un solo indicador que tan disperso se encuentran un conjunto de datos a un punto de concentración, se debe como primera medida, calcular la distancia de cada dato respecto a una medida de tendencia central. Por ejemplo, si se tienen los siguientes datos: 4, 5, 3,5, 3, 2,2, 2, 2,3, 5, 1,4, 1, 4.

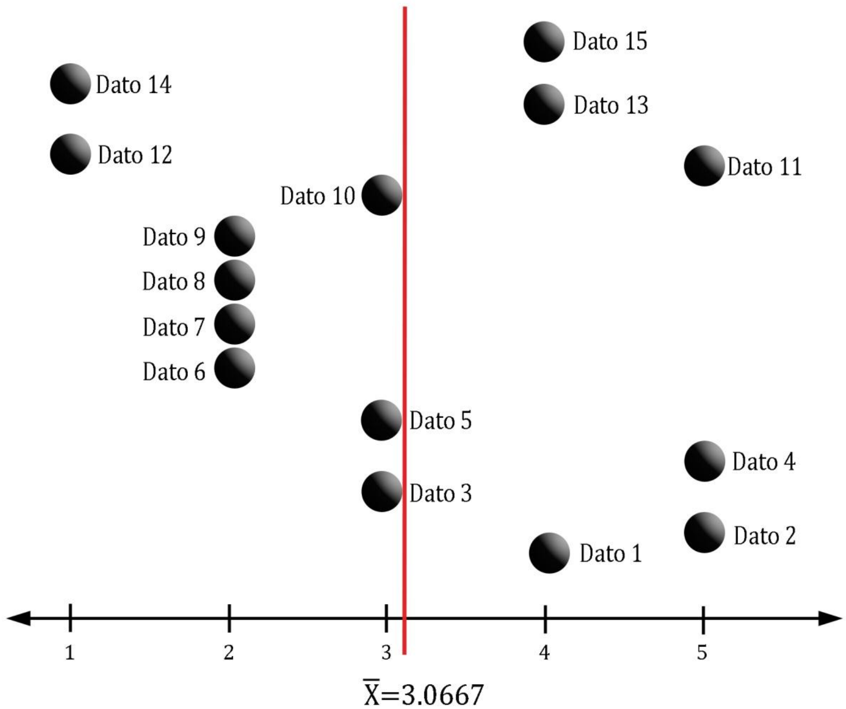
Tenemos que la media aritmética es de aproximadamente 3,0667 (indicador de tendencia central por excelencia). El primer dato (4) se aleja de la media aritmética en 0,9333 hacia la derecha. Gráficamente se representa:



Y el segundo dato (5) la distancia es de 1.9333 respecto a la media aritmética. Gráficamente:



El caso del tercer dato (3) posee una distancia de 0.0667 hacia la izquierda de la media aritmética. Para indicar las distancias a la izquierda se utiliza el signo negativo, por tanto, la distancia del tercer dato sería −0.0667. La representación gráfica de todos los puntos quedaría:



La suma de todas las distancias de los puntos que están a la izquierda respecto a la media es de -8.6, y la sumatoria de las distancias de los puntos que están a la derecha respecto a la media 8.6, que son iguales. Al sumar todas las distancias de cada punto respecto a la media aritmética es igual a cero (las distancias se anulan); es decir:

$$\sum_{i=1}^{n}(X_i - \overline{X}) = 0 \qquad \textbf{Ecuación 1}$$

Demostración

$$\sum_{i=1}^{n}(X_i - \overline{X}) = \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \overline{X} = \sum_{i=1}^{n} X_i - n\overline{X} = \sum_{i=1}^{n} X_i - n\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} X_i = 0$$

Ya que:

Pero ¿qué tan disperso están los datos respecto a la media aritmética?, para contestar esta pregunta se recurre al promedio simple. Para llegar a una fórmula básica de dispersión, en que las distancias positivas y negativas no se eliminen, se modifica la fórmula anterior para trabajar solo con distancias positivas mediante el valor absoluto:

$$\sum_{i=1}^{n} | X_i - \overline{X} | \qquad \textbf{Ecuación 2}$$

Luego se divide entre la distancia total absoluta y el total de datos. A esta distancia promedio se le conoce con el nombre de **desviación media**.

El valor absoluto de un número es el número sin signo y se denota con dos barras verticales.

Para el ejemplo, la distancia promedio sería de aproximadamente 1.15 y significa que en promedio, los datos se separan de la media aritmética en 1.15 unidades.

La desviación media, se simboliza por (Dm), y se obtiene por la división de la sumatoria del valor absoluto de las distancias existentes entre cada dato y su media aritmética, y el número total de datos.

El considerar la diferencia de cada valor respecto de la media aritmética en valor absoluto, significa que no se toma en cuenta el signo negativo, es decir, no habrá nunca valores negativos, por lo cual, la Desviación Media siempre será positiva. Y se expresa en las mismas unidades de la variable que se está estudiando (m, dólares, g, etc.).

La desviación media tiene dos ventajas: Utiliza para su cómputo todos los elementos de la serie de datos y es fácil de entender. Sin embargo, es difícil trabajar con valores absolutos y por ello la desviación media no es usada frecuentemente.

CÁLCULO DE LA DESVIACIÓN MEDIA

Caso 1: Datos sin agrupar

Se utiliza la siguiente fórmula:

$$Dm = \sum_{i=1}^{n} \frac{| X_i - \overline{X} |}{n} \qquad \textbf{Ecuación 3}$$

Donde:

$\bar{X}$: Es la media aritmética de los números

n : Cantidad de datos o tamaño de la muestra.

$|X_i - \bar{X}|$: es el valor absoluto de la desviación de X_i respecto de $\bar{X}$.

Caso 2: Datos agrupados en tablas de frecuencias.

Se utiliza la siguiente fórmula:

$$Dm = \sum_{i=1}^{n} \frac{| X_i - \overline{X} | * n_i}{n} \qquad \textbf{Ecuación 4}$$

Donde:

$\bar{X}$: Es la media aritmética de los números

n: Cantidad de datos o tamaño de la muestra.

n_i : Frecuencia absoluta de cada dato.

$|X_i - \bar{X}|$: Valor absoluto de la desviación de X_i respecto de $\bar{X}$.

Caso 3: Datos agrupados en tablas de frecuencia por intervalos.

Se utiliza la siguiente fórmula:

$$Dm = \sum_{i=1}^{n} \frac{| c_i - \overline{X} | * n_i}{n} \qquad \textbf{Ecuación 5}$$

Donde:

$\bar{X}$: Es la media aritmética de los números.

n : Cantidad de datos o tamaño de la muestra.

n_i : Frecuencia absoluta de cada intervalo.

c_i : Punto medio de cada intervalo o clase.

$|c_i - \bar{X}|$: Valor absoluto de la desviación de c_i respecto de $\bar{X}$.

Procedimiento de cálculo

1. Auxiliarse de una tabla para obtener los valores requeridos por la fórmula.
2. Determinar la Media Aritmética de los valores considerados.
3. Obtener las diferencias de cada valor respecto a la Media Aritmética en valor absoluto, es decir, sin considerar el signo, por lo cual serán todas positivas.
4. Sumar las diferencias absolutas de cada valor respecto a la Media Aritmética (Datos sin agrupar).
5. Multiplicar las diferencias por su correspondiente frecuencia en cada clase y sumar estos resultados (Datos agrupados).
6. Sustituir los valores en la fórmula y determinar el valor de la Desviación Media (Dm).

VARIANZA

Si las desviaciones con respecto a la media se consideran al cuadrado, $(X_i - \overline{X})^2$ de nuevo se obtiene que todos los sumandos tengan el mismo signo (positivo). Esta es además la forma de medir la dispersión de los datos de tal manera que sus propiedades matemáticas son más fáciles de utilizar. Se pueden definir, entonces, dos estadísticos fundamentales: La varianza y la desviación estándar (o típica).

La varianza de un conjunto de datos o mediciones, se representa por (s^2) y se define como la suma de los cuadrados de las desviaciones de las observaciones con respecto a su media, dividida entre el número de observaciones menos uno; es decir, como la media de las diferencias cuadráticas de n puntuaciones con respecto a su media aritmética.

La varianza es una medida que proporciona información sobre la mayor o menor dispersión de los valores de la variable respecto a su media aritmética, de tal modo que, mientras que cuando mayor sea el valor de la varianza, mayor dispersión existirá y por tanto, menor representatividad tendrá la media aritmética, y cuanto más pequeña sea la varianza, menor es la dispersión, lo que significa que mayor es la concentración de los datos o valores alrededor de su media.

La varianza tiene una gran aplicación en el análisis estadístico avanzado, pero tiene el inconveniente de que sus unidades se expresan en las mismas de la variable analizada, pero elevadas al cuadrado lo cual complica la interpretación de dicho resultado.

CÁLCULO DE LA VARIANZA

Caso 1: Datos sin agrupar

Se utiliza la siguiente fórmula:

$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n} \qquad \textbf{Ecuación 6}$$

Donde:

$\overline{X}$: Es la media aritmética de los números

n : Cantidad de datos o tamaño de la muestra.

$X_i - \bar{X}$: Es la desviación de X_i respecto de $\bar{X}$.

Caso 2: Datos agrupados en s de frecuencias.

Se utiliza la siguiente fórmula:

$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2 * n_i}{n} \qquad \textbf{Ecuación 7}$$

Donde:

$\bar{X}$: Es la media aritmética de los números

n: Cantidad de datos o tamaño de la muestra.

n_i : Frecuencia absoluta de cada dato.

$X_i - \bar{X}$: Es la desviación de X_i respecto de $\bar{X}$.

c_i : Punto medio de cada intervalo o clase.

$|c_i - \bar{X}|$: Es la desviación de c_i respecto de $\bar{X}$.

Caso 3: Datos agrupados en tablas de frecuencia por intervalos.

Se utiliza la siguiente fórmula:

$$S^2 = \frac{\sum_{i=1}^{n}(c_i - \overline{X})^2 * n_i}{n} \qquad \textbf{Ecuación 8}$$

Donde:

$\bar{X}$: Es la media aritmética de los números.

n : Cantidad de datos o tamaño de la muestra.

n_i : Frecuencia absoluta de cada intervalo.

Importante: La varianza muestral también puede definirse como:

$$S'^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1} \qquad \textbf{Ecuación 9}$$

Esta expresión divida entre n-1 se utiliza con la finalidad de, además de tener fines descriptivos, realizar inferencias sobre una población usando S'^2 y no S^2; por cuanto se demuestra que S'^2 es un mejor estimador de la varianza poblacional σ^2 que como se verá en el tema de estimación.

Fórmulas Abreviadas para el Cálculo de la Varianza

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n} = \frac{1}{n}\left[\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right] = \frac{1}{n}\left[\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}\right] = \frac{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}{n-1} \qquad \textbf{Ecuación 10}$$

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2 * n_i}{n} = \frac{\sum_{i=1}^{n} X_i^2 * n_i}{n} - \overline{X}^2 = \frac{n\sum_{i=1}^{n} X_i^2 * n_i - \left(\sum_{i=1}^{n} X_i * n_i\right)^2}{n} \qquad \textbf{Ecuación 11}$$

$$S^2 = \frac{\sum_{i=1}^{n}(c_i - \overline{X})^2 * n_i}{n} = \frac{n\sum_{i=1}^{n} c_i^2 * n_i - \left(\sum_{i=1}^{n} c_i * n_i\right)^2}{n} = \frac{\sum_{i=1}^{k} n_i * C_i^2}{n} - \overline{X}^2 \qquad \textbf{Ecuación 12}$$

Demostración (Ecuación 10)

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

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

$$= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2\overline{X}\frac{1}{n}\sum_{i=1}^{n} X_i + n\frac{1}{n}\overline{X}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2\overline{X}\,\overline{X} + \overline{X}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2\overline{X}^2 + \overline{X}^2$$

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

$$S^2 = \frac{1}{n}\left[\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right] \quad \text{l.q.q.d}$$

PROPIEDADES DE LA VARIANZA

1. El valor generalmente es positivo. Solo es igual a cero cuando todos los valores de los datos son el mismo número, es decir, que no existe variabilidad, y recíprocamente cuando todos los datos son iguales entonces es cero. Nunca es negativa porque el numerador incluye diferencias al cuadrado.
2. Si a todos los valores de la variable se les suma un número la varianza no varía.
3. Si todos los valores de la variable se multiplican por un número la varianza queda multiplicada por el cuadrado de dicho número.
4. Si tenemos varias distribuciones con la misma media y conocemos sus respectivas varianzas se puede calcular la varianza total.

Si todas las muestras tienen el mismo tamaño:

$$S^2 = \frac{S_1^2 + S_2^2 + \ldots + S_n^2}{n}$$

Si las muestras tienen distinto tamaño:

$$S^2 = \frac{k_1 * S_1^2 + k_2 * S_2^2 + \ldots + k_n * S_n^2}{k_1 + k_2 + \ldots + k_n}$$

Observaciones sobre la varianza

1. Al igual que la media, es un índice muy sensible a las puntuaciones extremas.
2. En los casos que **no se pueda hallar la media** tampoco será posible hallar la **varianza**.
3. No viene expresada en las mismas unidades que los datos, ya que las desviaciones están elevadas al cuadrado.

DESVIACIÓN ESTÁNDAR

La desviación estándar se define como la raíz cuadrada de la varianza, considerada siempre positiva y por lo tanto se representa por (S); también recibe el nombre de desviación típica. Al igual que la Desviación Media, es considerada como una medida de la fluctuación (dispersión) que hay en los datos; es decir, determina la dispersión (variación) alrededor de la media Aritmética. Y es matemáticamente lógica; ya que toma en cuenta los signos de las diferencias de cada valor respecto a la Media Aritmética.

La desviación típica por sus propiedades algebraicas es la medida de dispersión de mayor utilidad, al obtenerse como raíz cuadrada de la varianza tiene la ventaja que sus unidades se expresen en las mismas de la variable a partir de la que se haya obtenido. En términos generales:

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

CÁLCULO DE LA DESVIACIÓN TÍPICA

Caso 1: Datos sin agrupar

Se utiliza la siguiente fórmula:

$$S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}} \qquad \textbf{Ecuación 13}$$

Caso 2: Datos agrupados en tablas de frecuencias.

Se utiliza la siguiente fórmula:

$$S = \sqrt{\frac{\sum_{i=1}^{k}(X_i - \overline{X})^2 * n_i}{n}} \qquad \textbf{Ecuación 14}$$

Caso 3: Datos agrupados en tablas de frecuencia por intervalos.

Se utiliza la siguiente fórmula:

$$S = \sqrt{\frac{\sum_{i=1}^{k}(c_i - \overline{X})^2 * n_i}{n}} \qquad \textbf{Ecuación 15}$$

Propiedades de la Desviación Estándar

1. El valor generalmente es positivo. Solo es igual a cero cuando todos los valores de los datos son el mismo número, es decir, que no existe variabilidad, y recíprocamente cuando todos los datos son iguales entonces es cero. Nunca es negativa porque el numerador incluye diferencias al cuadrado.
2. Si a todos los valores de la variable se les suma un número la desviación estándar no varía.
3. Si todos los valores de la variable se multiplican por un número la desviación estándar queda multiplicada por dicho número.
4. Si tenemos varias distribuciones con la misma media y conocemos sus respectivas desviaciones estándar se puede calcular la desviación estándar total.

Si todas las muestras tienen el mismo tamaño:

$$S^2 = \sqrt{\frac{S_1{}^2 + S_2{}^2 + \ldots + S_n{}^2}{n}}$$

Si las muestras tienen distinto tamaño:

$$S^2 = \sqrt{\frac{k_1 * S_1{}^2 + k_2 * S_2{}^2 + \ldots + k_n * S_n{}^2}{k_1 + k_2 + \ldots + k_n}}$$

Observaciones sobre la desviación típica

1. Al igual que la media y la varianza, es un índice muy sensible a las puntuaciones extremas.
2. En los casos que no se pueda hallar la media tampoco será posible hallar la desviación típica.
3. Cuanta más pequeña sea la desviación típica mayor será la concentración de datos alrededor de la media.

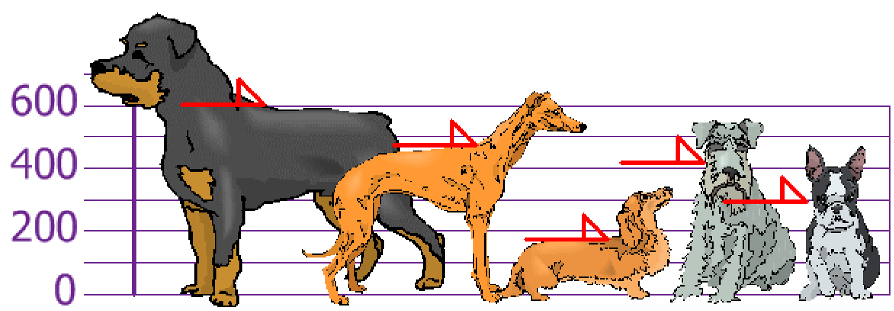
Procedimiento para el cálculo de Varianza y Desviación Estándar (Datos sin agrupar)

1. Auxiliarse de una tabla para obtener los valores requeridos por la fórmula.
2. Determinar la Media Aritmética de los datos.
3. Obtener las diferencias de cada valor respecto a la Media Aritmética.
4. Elevar al cuadrado cada diferencia.
5. Obtener la suma de las diferencias elevadas al cuadrado.
6. Sustituir los valores en la fórmula de la varianza (S^2 o S).

Procedimiento para el cálculo de Varianza y Desviación Estándar (Datos agrupados)

1. Auxiliarse de una tabla para obtener los valores requeridos por la fórmula.
2. Determinar la Media Aritmética de los datos según caso.
3. Obtener las diferencias de cada valor respecto a la Media Aritmética.
4. Elevar al cuadrado cada diferencia.
5. Multiplicar cada diferencia al cuadrado por su respectiva frecuencia.
6. Obtener la suma.
7. Sustituir los valores en la formula de la varianza (S^2 o S).

Ejemplo 2: Tú y tus amigos miden las alturas de sus perros (en milímetros):

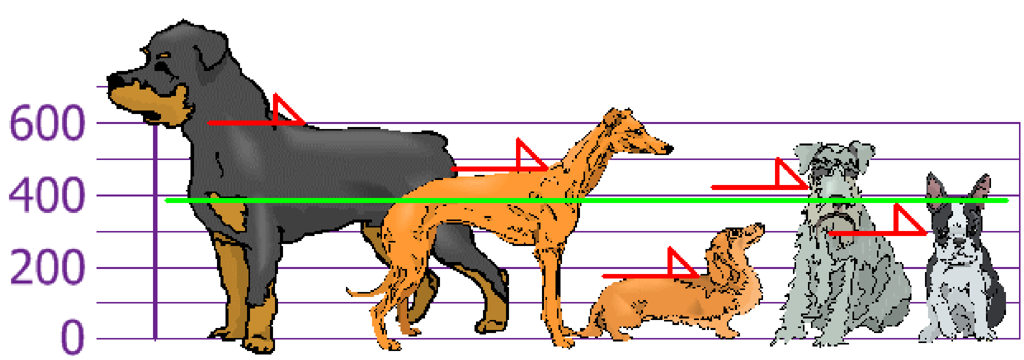


De la gráfica se observa que las alturas de los hombros de los perros son: 600 mm, 470 mm, 170 mm, 430 mm y 300mm. Calcula la media, la varianza y la desviación estándar.

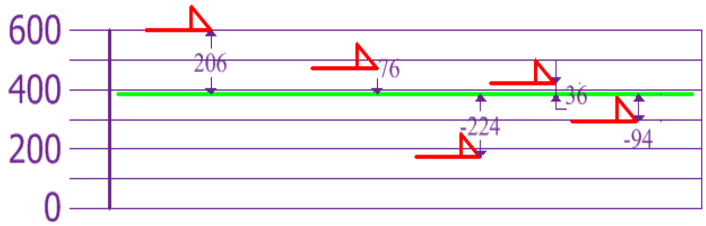
Solución:

Cálculo la Media Aritmética: $\overline{X} = \dfrac{600+470+170+430+300}{5} = \dfrac{1970}{5} = 394 \text{ mm}$

Por tanto, la altura media es 394 mm. Si se dibuja este resultado en el gráfico queda de la siguiente forma:



Al calcular la diferencia de cada altura con la Media Aritmética se tiene:



Para calcular la varianza, se toma cada diferencia, elévala al cuadrado, y hace la media:

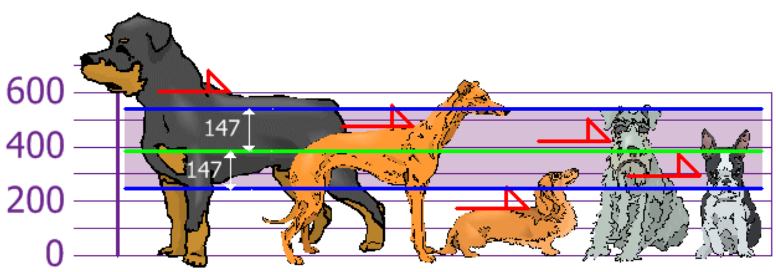
$$\sigma^2 = \frac{(600-394)^2 + (470-394)^2 + (170-394)^2 + (430-394)^2 + (300-394)^2}{5}$$

$$= \frac{(206)^2 + (76)^2 + (-224)^2 + (36)^2 + (-94)^2}{5} = \frac{108,520}{5} = 21,704 \ \text{mm}^2$$

Y la desviación estándar es la raíz de la varianza, así que:

$$S = \sqrt{21.704 \ \text{mm}^2} = 147 \ \text{mm}$$

En la siguiente gráfica se observa qué alturas están a distancia menos de la desviación estándar (147mm) de la media:



Así que la desviación estándar permite determinar una manera "estándar" de saber qué es normal, o extra grande o extra pequeño. Por lo tanto, los Rottweilers son perros grandes y los Dachsunds son un poco pequeños.

COEFICIENTE DE VARIACIÓN

Las unidades de la desviación estándar son las mismas que las unidades de los datos originales, y por eso es más fácil comprender la desviación estándar que la varianza. Sin embargo, esta propiedad dificulta o impide comparaciones entre conjuntos de datos que tengan diferente naturaleza o comparar la variación de valores tomados de distintas muestras o poblaciones.

Así por ejemplo, si se quisiera saber cual variable tiene un comportamiento más homogéneo, el peso o la estatura de un conjunto de personas, no es posible comparar las desviaciones típicas entre ellas, porque se encuentran expresadas en diferentes unidades.

Para solucionar esta desventaja o inconveniente que presentan las medidas de dispersión estudiadas, se utiliza el ***coeficiente de variación.***

El coeficiente de variación se representa por CV, y se define como el cociente o razón entre la media aritmética y la desviación estándar. Es decir:

$$CV = \frac{S}{\overline{X}} \qquad ó \qquad CV = \frac{S}{\overline{X}} * 100\%$$

La mayoría de las veces es expresado en términos de porcentaje; ya que es la proporción o porcentaje de la media que representa la desviación estándar. La fórmula anterior proviene de una regla de tres simple:

$$\overline{X} \longrightarrow 100\%$$
$$S \longrightarrow CV$$

Si por ejemplo el CV = 20%, significa que la desviación estándar representa el 20% del valor de la media aritmética.

Si la dispersión que se quiere estudiar esta referido a la mediana, se obtiene el coeficiente de variación mediana de la siguiente forma:

$$CV = \frac{S}{Me} \qquad ó \qquad CV = \frac{S}{Me} * 100\%$$

El coeficiente de variación permite realizar comparaciones entre dos muestra o poblaciones distintas e incluso, comparar la variación producto de dos variables diferentes (que pueden provenir de una misma población); porque este proporciona un valor libre de unidades de medida especifica (valor adimensional); es decir elimina la dimensionalidad de las variables y tiene en cuenta la proporción existente entre una medida de tendencia y la desviación típica o estándar.

Como la dispersión se utiliza para medir la representatividad de la media, el CV también se puede utilizar para comparar la representatividad de dos medias. Valores bajos del CV indicarán poca dispersión/mucha representatividad y valores altos indicarán mucha dispersión/poca representatividad.

No hay criterios universales para decir que un valor del CV es "bajo" o "alto", aunque en la práctica se suele considerar los siguientes criterios:

Valor de CV	Criterio
CV $< 10\%$	poca dispersión
$10\% \leq CV \leq 33\%$	aceptable
$34\% \leq CV < 50$	alta dispersión
$CV > 50\%$	muy alta dispersión

Si para un conjunto de datos el coeficiente de variabilidad es menor que 10% entonces se considera que este conjunto de datos es homogéneo, es decir que casi no existe variabilidad entre ellos y que por lo tanto la media aritmética es representativa de dichos datos. O sea que, en este caso, la media da una idea clara de dichos datos.

Propiedades del Coeficiente de Variación

1. Representa un porcentaje de razón entre la desviación típica y la media, de manera que representa cuantas veces es la desviación típica con relación a la media.
2. Se considera un número abstracto, es decir sin unidades, pues tanto S como $\overline{X}$ vienen en las mismas unidades de los datos, y al hacer la división se simplifican.
3. No es alterado cuando los datos son multiplicados por una constante, pues en virtud de las propiedades de $\overline{X}$ y de "S" ambos quedan multiplicados por esa constante, sin alterar al cociente.

De las propiedades anteriores se verifica que:

- Si CV= 50% significa que la desviación típica es la mitad de la media, lo que revela una alta variabilidad. Valores del CV menores al 10 % revelan poca variabilidad de los datos; y así pues en Control de Calidad, es frecuente exigir un CV menor al 5% entre las muestras, a fin de garantizar su homogeneidad.
- Cuando se utiliza el CV para hacer comparaciones entre varios conjuntos de datos, y concluir que cuanto más pequeño sea su valor, más homogéneo es el comportamiento.
- El CV es invariante frente a cambios de unidades, como por ejemplo, pasar de libras a kilogramos o de pies a centímetros, etc.

EJEMPLOS DE APLICACIÓN

Caso 1: Datos sin agrupar

Ejemplo 1: Tres alumnos son sometidos a una competencia para probar sus conocimientos en 10 materias diferentes, cada una sustentada con 10 preguntas. La idea del concurso es encontrar al alumno más idóneo para representar al Instituto en una competencia a nivel nacional. El número de preguntas buenas por materia se muestra a continuación:

Materias	1	2	3	4	5	6	7	8	9	10
Carlos	2	9	20	2	3	1	9	9	1	4
Pedro	7	2	2	6	6	3	6	7	6	5
Juan	5	6	5	5	5	5	4	5	6	4

a) ¿Cuál deberá ser el alumno seleccionado?
b) ¿Encuentre la desviación estándar e interprételа?
c) ¿Determine el porcentaje de variabilidad que tienen el número de pruebas aprobadas?

Solución a)

Lo primero que analizaremos es la media de los puntajes para cada uno de los alumnos, con el fin de determinar el alumno con mayor promedio de preguntas buenas.

Media Aritmética $\overline{X} = \dfrac{\sum_{i=1}^{10} X_i}{10}$		
Carlos	Pedro	Juan
$\overline{X} = \dfrac{\sum_{i=1}^{10} X_i}{10} = \dfrac{50}{10} = 5 \text{ pruebas}$	$\overline{X} = \dfrac{\sum_{i=1}^{10} X_i}{10} = \dfrac{50}{10} = 5 \text{ pruebas}$	$\overline{X} = \dfrac{\sum_{i=1}^{10} X_i}{10} = \dfrac{50}{10} = 5 \text{ pruebas}$

Se observa que el valor de la media aritmética es igual para los tres alumnos: los tres alumnos responden en promedio 5 preguntas correctas por cada prueba. Por lo tanto, no es una medida que nos ayuda a tomar una decisión.

¿Cuál sería entonces el indicador diferenciador entre los alumnos? La medida que responde a esta pregunta es la desviación media:

Carlos		Pedro		Juan	
X_i	$\mid X_i - \overline{X} \mid$	X_i	$\mid X_i - \overline{X} \mid$	X_i	$\mid X_i - \overline{X} \mid$
2	$\|2 - 5\| = 3$	7	$\|7 - 5\| = 2$	5	$\|5 - 5\| = 0$
9	$\|9 - 5\| = 4$	2	$\|2 - 5\| = 3$	6	$\|6 - 5\| = 1$
10	$\|10 - 5\| = 5$	2	$\|2 - 5\| = 3$	5	$\|5 - 5\| = 0$
2	$\|2 - 5\| = 3$	6	$\|6 - 5\| = 1$	5	$\|5 - 5\| = 0$
3	$\|3 - 5\| = 2$	6	$\|6 - 5\| = 1$	5	$\|5 - 5\| = 0$
1	$\|1 - 5\| = 4$	3	$\|3 - 5\| = 2$	5	$\|5 - 5\| = 0$
9	$\|9 - 5\| = 4$	6	$\|6 - 5\| = 1$	4	$\|4 - 5\| = 1$
9	$\|9 - 5\| = 4$	7	$\|7 - 5\| = 2$	5	$\|5 - 5\| = 0$
1	$\|1 - 5\| = 4$	6	$\|6 - 5\| = 1$	6	$\|6 - 5\| = 1$
4	$\|4 - 5\| = 1$	5	$\|5 - 5\| = 0$	4	$\|4 - 5\| = 1$
Suma	39		21		9
$Dm = \sum_{i=1}^{n} \frac{\mid X_i - \overline{X} \mid}{n}$	$\frac{39}{10} = 3.9$		$\frac{21}{10} = 2.1$		$\frac{9}{10} = 0.9$

Cálculo de la varianza y desviación estándar

Carlos			Pedro			Juan		
X_i	$X_i - \overline{X}$	$(X_i - \overline{X})^2$	X_i	$X_i - \overline{X}$	$(X_i - \overline{X})^2$	X_i	$X_i - \overline{X}$	$(X_i - \overline{X})^2$
2	2 - 5 = -3	$(-3)^2 = 9$	7	7 - 5 = 2	$(2)^2 = 4$	5	5 - 5 = 0	$(0)^2 = 0$
9	9 - 5 = 4	$(4)^2 = 16$	2	2 - 5 = -3	$(-3)^2 = 9$	6	6 - 5 = 1	$(1)^2 = 1$
10	10 - 5 = 5	$(5)^2 = 25$	2	2 - 5 = -3	$(-3)^2 = 9$	5	5 - 5 = 0	$(0)^2 = 0$
2	2 - 5 = -3	$(-3)^2 = 9$	6	6 - 5 = 1	$(1)^2 = 1$	5	5 - 5 = 0	$(0)^2 = 0$
3	3 - 5 = -2	$(-2)^2 = 4$	6	6 - 5 = 1	$(1)^2 = 1$	5	5 - 5 = 0	$(0)^2 = 0$
1	1 - 5 = -4	$(-4)^2 = 16$	3	3 - 5 = -2	$(-2)^2 = 4$	5	5 - 5 = 0	$(0)^2 = 0$
9	9 - 5 = 4	$(4)^2 = 16$	6	6 - 5 = 1	$(1)^2 = 1$	4	4 - 5 = 1	$(-1)^2 = 1$
9	9 - 5 = 4	$(4)^2 = 16$	7	7 - 5 = 2	$(2)^2 = 4$	5	5 - 5 = 0	$(0)^2 = 0$
1	1 - 5 = -4	$(-4)^2 = 16$	6	6 - 5 = 1	$(1)^2 = 1$	6	6 - 5 = 1	$(1)^2 = 1$
4	4 - 5 = 1	$(1)^2 = 1$	5	5 - 5 = 0	$(0)^2 = 0$	4	4 - 5 = -1	$(-1)^2 = 1$
Suma		128			34			4
$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}$	Poco significado por unidades al cuadrado $\frac{4}{10} = 0.4 \text{ pruebas}^2$		$\frac{34}{10} = 3.4 \text{ pruebas}^2$			$\frac{4}{10} = 0.4 \text{ pruebas}^2$		
$S = \sqrt{S^2}$	$S = \sqrt{12.8} = 3.58 \text{ pruebas}$		$S = \sqrt{3.4} = 1.84 \text{ pruebas}$			$S = \sqrt{0.4} = 0.63 \text{ pruebas}$		

De los resultados se observa que:

Carlos muestra una desviación media de 3,9 indicando que los datos se alejan en promedio de la media en 3.9 preguntas buenas. Pedro disminuye su variación (2.9), siendo Juan el que menos variación presenta con 0.9 preguntas tanto por arriba como por debajo de la media aritmética.

Recomendación

Elegir como ganador del concurso a Juan, pues presenta resultados más constantes que los otros dos alumnos, Juan en promedio acierta 5 preguntas buenas con una variación muy baja (rondando entre 4 y 6).

Solución b) Cálculos

Interpretación

Para Carlos se obtuvo una desviación estándar 3.58 pruebas, lo que significa que en promedio el número de pruebas buenas que contesto Carlos se distancian de la media aritmética (5 pruebas) aproximadamente tres y la mitad pruebas buenas.

Para Pedro se obtuvo una desviación estándar 1.84 pruebas, lo que significa que en promedió el número de pruebas buenas que contesto Pedro se distanciaron de la media aritmética (5 pruebas) cerca de dos pruebas buenas.

Para Juan se obtuvo una desviación estándar 0.63 pruebas, lo que significa que en promedió el número de pruebas buenas que contesto Juan se distanciaron de la media aritmética (5 pruebas) aproximadamente la mitad de una prueba buena.

En general el número de pruebas que contesto Juan se encontraron más cercanas a media aritmética, que el número de pruebas buenas que constaron Carlos y Pedro, lo que significa que en el número de pruebas contestadas por Carlos y Pedro hubo mayor dispersión.

Solución c) Coeficiente de variación

Coeficiente de Variación $CV = \frac{S}{\overline{X}} * 100\%$		
Carlos	Pedro	Juan
$CV = \frac{3.58}{5} * 100\% = 71.6\%$ Significa que el número de pruebas buenas que contesto Carlos tiene una variación del 71.6% alrededor de la media aritmética. Tiene una dispersión muy alta. La media no se consi-	$CV = \frac{1.84}{5} * 100\% = 36.8\%$ Significa que el número de pruebas buenas que contesto Pedro tiene una variación del 36.8% alrededor de la media aritmética. Tiene una dispersión alta. La media no es representa-	$CV = \frac{0.63}{5} * 100\% = 12.6\%$ Significa que el número de pruebas buenas que contesto Juan tiene una variación del 12.6% alrededor de la media aritmética. Tiene una dispersión aceptable acercándose a baja dis-

dera representativa del número de pruebas buenas.	tiva del número de pruebas buenas.	persión. La media es representativa del número de pruebas buenas que las que contesto Pedro.

La relación que existe entre los coeficientes de variación es: $CV_{Juan} < CV_{Pedro} < CV_{Carlos}$; por lo tanto, menor variación o dispersión se encuentra en el número de pruebas buenas que contesto Juan; es decir el conjunto de pruebas buena contestadas por Juan es más homogéneo que el de los otros dos.

Caso 2: Datos agrupados en tablas de frecuencias.

Ejemplo 2: Un pediatra del Hospital Benjamín Boom registro los meses de edad de 50 niños de su consulta en el momento de empezar a caminar, obteniéndose la siguiente tabla de frecuencias:

Meses (X_i)	9	10	11	12	13	14	15
Niños(n_i)	1	4	9	16	11	8	1

Haga un análisis de la dispersión de los datos y la representatividad de la media aritmética de estos datos.

Solución

- **Cálculo del rango:** $R = X_{último} - X_{primer}$

  $R = 15 – 9 = 6$ meses

  Interpretación: la variación de la edad de los niños que pasan consulta con el pediatra es de 6 meses, la cual oscila entre 9 y 15 meses.

- **Cálculo de la media aritmética:** $\overline{X} = \frac{\sum_{i=1}^{50} X_i * n_i}{50}$

$$\overline{X} = \frac{\sum_{i=1}^{50} X_i * n_i}{50} = \frac{1*9+4*10+9*11+16*12+11*13+8*14+1*15}{50} = \frac{610}{50} = 12.2 \text{ meses}$$

- **Cálculo de la varianza y desviación estándar.**

Mese (xi)	Niños(n_i)	$X_i - \overline{X}$	$(X_i - \overline{X})^2$	$(X_i - \overline{X})^2 * n_i$
9	1	9 – 12.2 = -3.2	10.24	10.24
10	4	10 – 12.2 = -2.2	4.84	19.36
11	9	11 –12.2 = -1.2	1.44	12.96
12	16	12 – 12.2 = -0.2	0.04	0.64
13	11	13 – 12.2 = 0.8	0.64	7.04

14	8	14 – 12.2 = 1.8	3.24	25.92
15	1	15 – 12.2 = 2.8	7.84	7.84
Suma	50			84
$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2 * n_i}{n} = \frac{84}{50} = 1.68 meses^2$		$CV = \frac{S}{\overline{\overline{X}}} * 100\%$	$S = \sqrt{S^2} = \sqrt{1.68} = 1.30$ meses	

Interpretación: El valor de la desviación estándar es igual a 1.3 meses, lo que significa que en promedio el número de meses que los niños aprenden a caminar se distancian de la media aritmética (12.2 meses) en 1.3 meses. Este considerablemente pequeño comparado con los valores que toma la variable, por lo tanto el valor de la media representa los meses en que los niños aprenden a caminar.

- **Cálculo del coeficiente de variación:**

$$CV = \frac{1.30}{12.2} * 100\% = 10.65\%$$

Interpretación: El valor del coeficiente de variabilidad esta cercano al 10% se puede considerar poca dispersión en los datos que representa la variable edad de los niños. Es decir, la media aritmética de 12.2 meses es representativa y da una idea clara del número meses en que estos niños aprenden a caminar.

Caso 3: Datos agrupados en tablas de frecuencia por intervalos.

Ejemplo 3: Se ha llevado a cabo una investigación sobre el número de panes consumidos por un grupo de familias en una colonia de San Salvador, durante una semana determinada. Los recopilados se presentan a continuación.

Clases	30-32	33-35	36-38	39-41	42-44	45-47	48-50
Número de familias n_i	10	18	60	100	80	14	6

Encuentre las medidas de dispersión e interprételas.

Solución

- **Cálculo del rango aproximado:**

R=50 – 30=20 panes

Interpretación: La variación del consumo de panes de las familias durante la semana es de 20 panes, el cual oscila entre 30 y 50 panes.

- **Organización de los datos.**

[li-1,li]	n_i	C_i	$n_i * c_i$	C_i^2	$C_i^2 * n_i$
[30-32]	10	31	310	961	9610
[33-35]	18	34	612	1156	20808

[36-38]	60	37	2220	1369	82140
[39-41]	100	40	4000	1600	160000
[42-44]	80	43	3440	1849	147920
[45-47]	14	46	644	2116	29624
[48-50]	6	49	294	2401	14406
n	288		11520		464508

- Cálculo de la media Aritmética: $\overline{X} = \sum_{i=1}^{k} \frac{c_i n_i}{n}$

$$\overline{X} = \sum_{i=1}^{7} \frac{c_i * n_i}{288} = \frac{11520}{288} = 40 \text{ panes}$$

- Cálculo de varianza y desviación estándar $S^2 = \frac{\sum_{i=1}^{k} C_i^2 * n_i}{n} - \overline{X}^2$

$$S^2 = \frac{\sum_{i=1}^{7} C_i^2 * n_i}{288} - (40)^2 = \frac{464508}{288} - 1600 = 1612.875 - 1600 = 12.875 \text{ panes}^2$$

$$S = \sqrt{S^2} = \sqrt{12.875} = 3.59 \text{ panes}$$

Interpretación: El resultado obtenido indica que en promedio, el consumo de pan del grupo de familias de una colonia de san salvador se dispersa con respecto a su media aritmética (40 panes) en una cantidad igual a 3.59 panes.

El valor de la desviación estándar se puede considerar pequeño comparado con los valores que toma la variable, por lo tanto la media aritmética es una medida fiable para representar esta variables (la media es representativas).

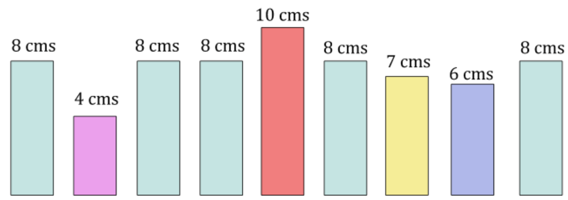
- Cálculo del coeficiente de variación: $CV = \frac{S}{\overline{X}} * 100\%$

$$CV = \frac{3.59}{40} * 100\% = 8.97\%$$

Interpretación: El valor del coeficiente de variabilidad es menor que 10% se puede considerar poca dispersión en los datos que representa la variable número de panes. Es decir, la media aritmética de 40 panes es una medida representativa y da una idea clara del número panes que estas familias consumieron en una semana.

APLICANDO LO APRENDIDO

1) Si se tienen 9 rectángulos cuya altura esta en centímetros (y todos tienen la misma base).



¿Cuál es la varianza y la desviación estándar de las alturas de los rectángulos?

2) La siguiente distribución de frecuencias corresponde a las horas semanales de trabajo de un grupo de docentes de educación media.

Horas de Trabajo	[10-15[	[15-20[	[20-25[	[25-30[	[30-35[	[35-40[	[40-45[	[45-50[	[50-55]
Cantidad de docentes	5	7	25	15	45	18	13	6	3

Encontrar las medidas de dispersión e interprételas.

3) Se tomó una prueba de nivel con un puntaje máximo de 20 a dos comisiones de estudiantes (C1 y C2) y los resultados obtenidos fueron los siguientes:

Comisión	Puntajes obtenidos				
	[15-16[	[16-17[	[17-18[	[18-19[	[19-20[
C1	8	20	19	13	10
C2	16	18	19	9	8

a) ¿Para cuál de los dos grupos el promedio de los puntajes resultó mayor?

b) ¿Para cuál de los dos grupos el puntaje resultó más homogéneo?

4) En la Dirección de Turismo de una determinada Ciudad, se buscaron los datos referidos a la entrada de turistas durante los últimos tres años, seleccionando sólo los meses de temporada alta. La finalidad de dicha búsqueda, es la de realizar una síntesis de la información que permita analizar posibles cambios con respecto a las medidas implementadas hasta el momento y de esta manera incrementar la afluencia turística en la ciudad. En la siguiente tabla se presenta la información obtenida.

Mes	Primer Año	Segundo Año	Tercer Año	Total
Diciembre	50	76	90	216
Enero	125	120	105	350
Febrero	89	110	96	295
Marzo	67	87	78	232
Total	331	393	369	1093

¿En qué mes de la temporada alta el ingreso de turistas a la Ciudad resultó más homogéneo? Comparar los coeficientes de variación de los cuatro meses considerados.

5) Consideremos los siguientes conjuntos de valores referidos a las edades de los jugadores de dos equipos de fútbol.

Equipo 1:24, 25, 26, 23, 26, 21, 27, 24, 23, 26, 25 **Equipo 2**:36, 18, 28, 17, 37, 15, 14, 44, 27, 21, 13

a) Calcula la media de las edades en los dos equipos.

b) ¿Qué puedes decir respecto de las edades del equipo 1 en relación a su media?

c) ¿Qué puedes decir respecto de las edades del equipo 2 en relación a su media?

d) En nuestro ejemplo, ¿cuál es el rango en el equipo 1? ¿Y en el equipo 2?

e) ¿Qué equipo es más disperso, es decir más heterogéneo? Si los equipos no tuvieran la misma cantidad de jugadores ¿podrías decir que equipo es más disperso?

f) ¿Cómo interpretarías el valor del rango?

6) Un profesor de matemática debe elegir entre sus dos mejores alumnos Andrés y Paula para una Olimpiada de matemática. Las notas de ambos son:

Andrés	6.5	6.6	6.4	6.6	6.5	6.7
Paula	7.0	6.0	6.3	6.0	7.0	7.0

¿A cuál alumno le aconsejas que presente a la Olimpiada? Justifica.

7) Camila obtuvo el primer semestre en matemática las siguientes notas: 3, 6, 5.9, 7.0, 4.7 ,6.2, 62.

a) Calcula el promedio de Camila el primer semestre en matemática.

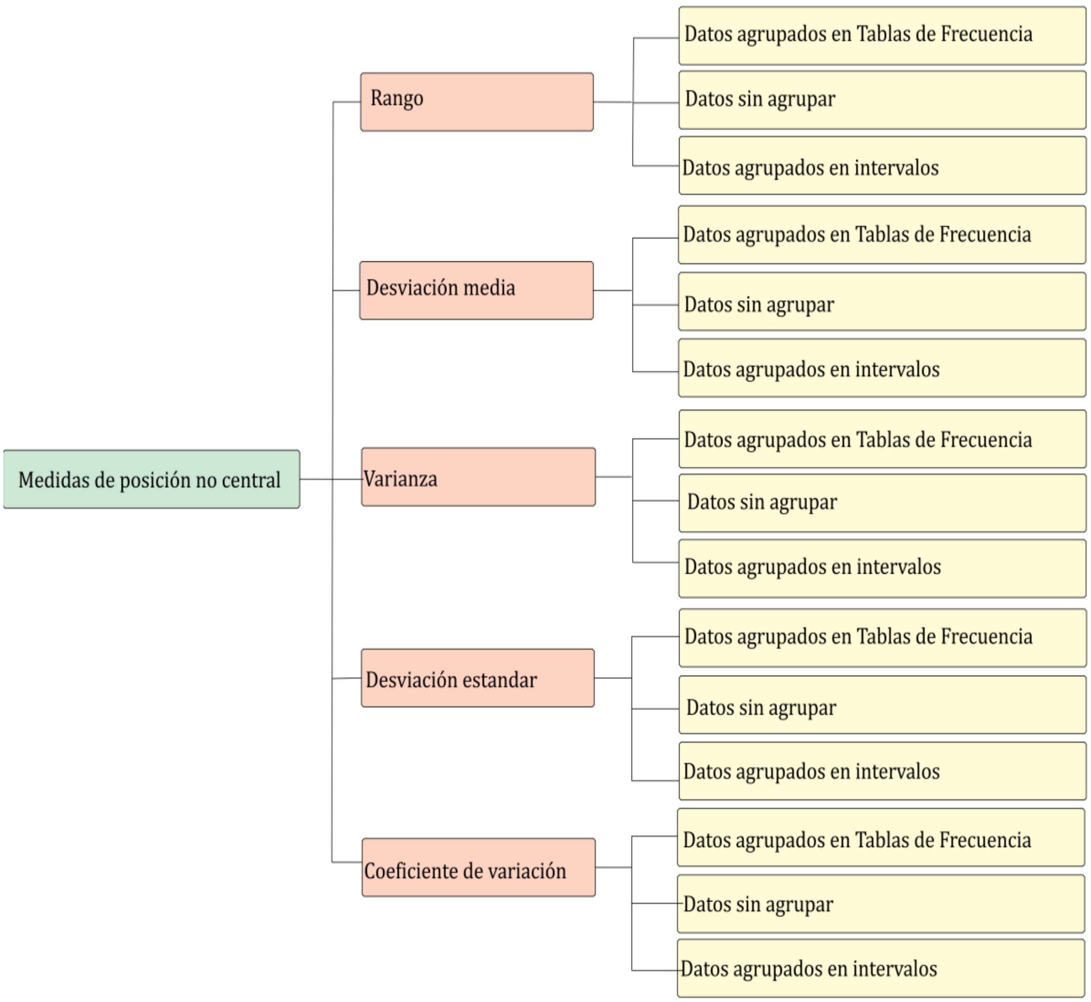
b) Calcula la desviación estándar.

c) ¿Consideras que el rendimiento académico en matemática de Camila fue parejo? ¿Qué información te permite justificar tu respuesta?

BIBLIOGRAFIA

1. Johnson R., Cuby P. (1999), *Estadística Elemental*. México: Internacional Thomson Editores, S.A de S.V.
2. Pérez C., (2003).*Estadística. Problemas Resueltos y Aplicaciones*. Madrid: Pearson Educación, S.A.
3. Pérez-T. H.E ((2007), Estadística para las Ciencias Sociales, del Comportamiento y de la Salud. (3ª. Ed.).México: Impreso Edamsa Impresiones, S.A. de C.V.
4. Sarabia J.M. (2000), *Curso Práctico de Estadística*. (2da ed.) . España: Impreso por Gráficas Rogar, S.A Navalcarnero (Madrid).
5. Triola, M., (2009). *Estadística.* (10a ed.). México: Pearson Educación.

DIAGRAMA DE CONTENIDOS



Lección 6 y 7 Segundo año de Bachillerato Unidad II Tiempo: 16 horas clase

TÉCNICAS DE CONTEO

Introducción

Con mucha frecuencia puede ser muy difícil y engorroso determinar el número posible de ordenaciones de un número finito de elementos, por medio de la enumeración directa. Las técnicas de conteo son las que permiten realizar estas ordenaciones, y se denominan también "Análisis Combinatorio".

La combinatoria puede considerarse tan vieja como la propia Matemática, ya que la operación básica de contar los elementos de un conjunto está ligada al origen mismo del concepto de número en los tiempos prehistóricos.

Desde tiempos muy remotos ha habido problemas de combinatoria que han llamado la atención de los matemáticos. Por ejemplo, el problema de los cuadrados mágicos que son arreglos de números con la propiedad de que la suma de los elementos de cualquier columna, renglón o diagonal es el mismo número, aparece en un viejo libro chino fechado 2200 A. C. Los cuadrados mágicos de orden 3 fueron estudiados con fines místicos. Los coeficientes binomiales, que son los coeficientes enteros de la expansión de $(a+b)^n$ fueron conocidos en el siglo XII. El triángulo de Pascal que es un arreglo triangular de los coeficientes binomiales fue desarrollado en el siglo XIII. La Combinatoria es la parte de la Matemática que se dedica a buscar procedimientos y estrategias para el recuento de los elementos de un conjunto o la forma de agrupar los elementos de un conjunto; y permite también contar los resultados posibles de un evento o suceso aleatorio sin tener la necesidad de hacer una lista de los posibles resultados que puedan suceder, ya que en la mayoría de los casos resulta tedioso hacer la lista completa y contar todas las posibilidades que se pueden dar.

Entre estas técnicas se pueden mencionar: las Permuta-

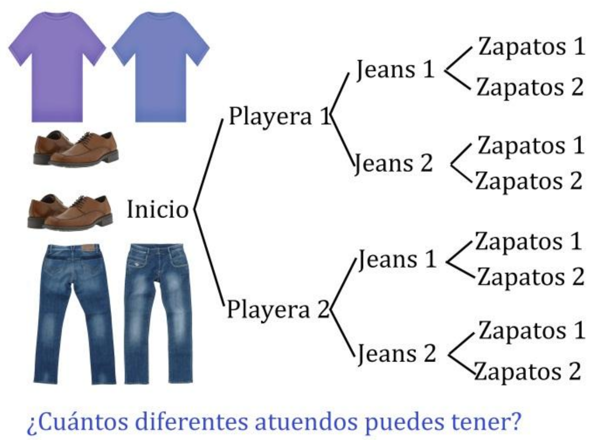


Figura 1. Distintas formas de vestirse.

Objetivos

- Reconocer las distintas formas de agrupar elemento.
- Calcular el número de permutaciones, variaciones o combinaciones, tanto si es posible repetir elementos en las agrupaciones, como si no es posible.
- Identificar las técnicas de conteo para la resolución de situaciones problemáticas que se le presenten en la vida cotidiana y/o escolar.

Importancia

Uno de los factores más importantes que han contribuido al gran desarrollo que ha tenido la combinatoria desde 1920 es la teoría de gráficas, la importancia de esta disciplina estriba en el hecho de que las gráficas pueden servir como modelos abstractos para modelar una gran variedad de relaciones entre objetos de un conjunto. Sus aplicaciones se extienden a campos tan diversos como la investigación de operaciones, química, mecánica, estadística, física teórica y problemas socio-económicos.

Entre sus aplicaciones prácticas se citan por una parte el cálculo de probabilidades, ya que para solucionar problemas de probabilidad en muchas ocasiones es fundamental llevar a cabo algún tipo de conteo, lo cual garantiza el éxito en la solución; este conteo se da para determinar los casos favorable y casos posibles, y por otra el cálculo de la complejidad o tiempo de ejecución de un algoritmo, por cuanto se cuenta el número de operaciones que se realizan en el procedimiento, número de comparaciones que realiza un programa para ordenar un conjunto de datos para determinar ¿Cuáles son "buenos" o "malos"?. Cuando se trata de estimar el número medio o esperado de operaciones que realiza un programa, se unen ambas aplicaciones, es decir, complejidad algorítmica y cálculo de probabilidades.

Competencias a reforzar.

- Aprenderás el principio de conteo para determinar los números de posibilidades.
- Aprenderás acerca de tipos especiales de disposiciones llamadas permutaciones y combinaciones
- Aplicar distintas técnicas de conteo, distinguiendo las adecuadas para la resolución de cada problema

Presaberes

- Elaboración de diagrama de árbol.
- Concepto de factorial de un número.
- Teoría de conjuntos.
- Operaciones básicas.
- Potenciación.

COMBINATORIA

La combinatoria es la parte de las Matemáticas que estudia las diversas formas de realizar agrupaciones con los elementos de un conjunto, formándolas y calculando su número.

Proporciona fórmulas que permiten conocer el número de elementos de aquellos conjuntos o la forma de realizar agrupaciones con sus elementos, en los que, por la extensión de los mismos, no es posible contar de uno en uno los elementos, pero que poseen algunas propiedades que permiten deducirlo utilizando algún procedimiento o fórmula.

La teoría combinatoria permite encontrar respuesta a muchas situaciones como las siguientes:

Si alguien nos pide que le digamos cuantos número de dos cifras se pueden formar con los dígitos 1 y 7, rápidamente respondemos que 4 y son (11, 17, 71, 77). Si por el contrario quisiera saber cuántos números de 15 cifras que pueden formar con esos mismos números, la respuesta no es tan inmediata.

Si se quisiera saber de cuantas formas se pueden sentar 20 personas en un autobús de 40 asientos, no se tendría una respuesta rápida e incluso si nos pusiésemos a contar acabaríamos por desistir.

Para darle respuesta a estas situaciones es esencial considerar lo siguiente:

- El número de elementos de que disponemos para formar los grupos.
- Los elementos que debe contener cada grupo.
- La posibilidad de repetir elementos o no en los grupos.
- La importancia o indiferencia en cuanto al orden en que aparecen los elementos en las agrupaciones.

En términos generales, la combinatoria ayudará a resolver problemas que consisten en:

- Dado un conjunto de **m** elementos.
- Se debe contar las maneras o formas de agruparlos.
- En grupos de **n** elementos.
- Considerando un criterio para el orden.
- Y un criterio para la posibilidad de repetición o no repetición de los elementos que forman los grupos.

Estas técnicas de conteo que permite contar los resultados posibles de un evento o suceso; o realizar agrupaciones sin tener la necesidad de hacer una lista de los posibles resultados que puedan suceder son:

a) Permutaciones
b) Variaciones
c) Combinaciones

Estas técnicas a su vez se pueden clasificar como: con repetición o sin repetición.

Actividad Introductoria: El dilema del taxista

Un taxista tiene que ir de un punto A de una ciudad a un punto B (ver figura 1). Para ir de A a B el taxista tomará las calles horizontales siempre en el sentido izquierda-derecha y las calles verticales siempre en el sentido arriba-abajo, esto significa que nunca retrocederá. ¿De cuántas formas puede el taxista realizar el trayecto?. Diseñe una estrategia para responder a esta pregunta y ayudar al taxista.

Solución

Diseño de una estrategia

Lo primero que se tiene que hacer es diseñar una estrategia que permita contar todos los casos sin olvidarse ninguno. Para ello una observación: el taxista no puede sino desplazarse en sentido horizontal o vertical, de tal forma que si arranca horizontalmente tendrá que seguir en este sentido hasta la primera intersección, donde podrá continuar sobre la misma calle o voltear a su derecha para bajar por la perpendicular, así sucesivamente. Estos sentidos de movimiento están señalizados en la figura anterior. Diseñe un método que le permita contar los posibles caminos y cuente estos; después siga con la lectura.

Primero se numeran las intersecciones de las calles donde debe pasar el taxista.

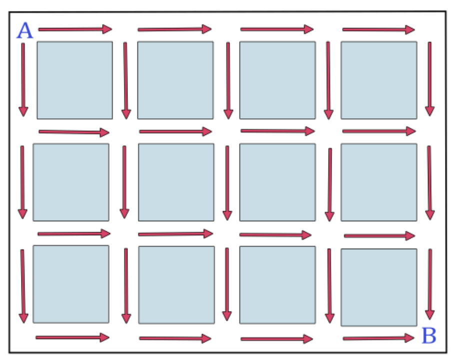


Figura 2. Sentidos de desplazamiento en ciudad

Luego se puede hacer es esquematizar el diagrama dado: lo importante son las intersecciones de las calles, puesto que entre ellas el taxista no puede variar la ruta. En la siguiente figura se muestra un esquema simplificado de la ciudad:

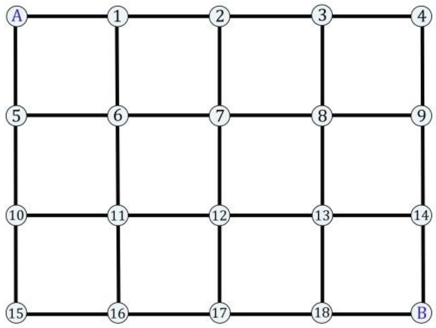


Figura 3. Intersecciones de ciudad

Cada intersección está señalada por un círculo; así, partiendo de la intersección A el taxista puede ir a la intersección 1 o 5; si ha avanzado horizontalmente, llegará al nodo 1 y puede decir ir a la intersección 2 o 6; si decide voltear, avanzará hasta el nodo 6 y, nuevamente, tendría que tomar una decisión.

¿Cuantos posibles caminos puede realizar?

Para hacer el recuento total de los caminos se puede seguir la siguiente estrategia: construir un árbol donde se indiquen las dos posibles opciones en cada caso, de tal forma que una ruta esté marcada por una secuencia de nodos.

Si se procede de esta forma, no es necesario escribir explícitamente todos los posibles caminos; por ejemplo, si se ha establecido que desde el nodo 7 hay 6 posibles caminos, siempre que lleguemos a él se tomara esté hecho como dato, sin establecer otra vez cuales son éstos.

Como se puede observar en la figura el número total de caminos es 35.

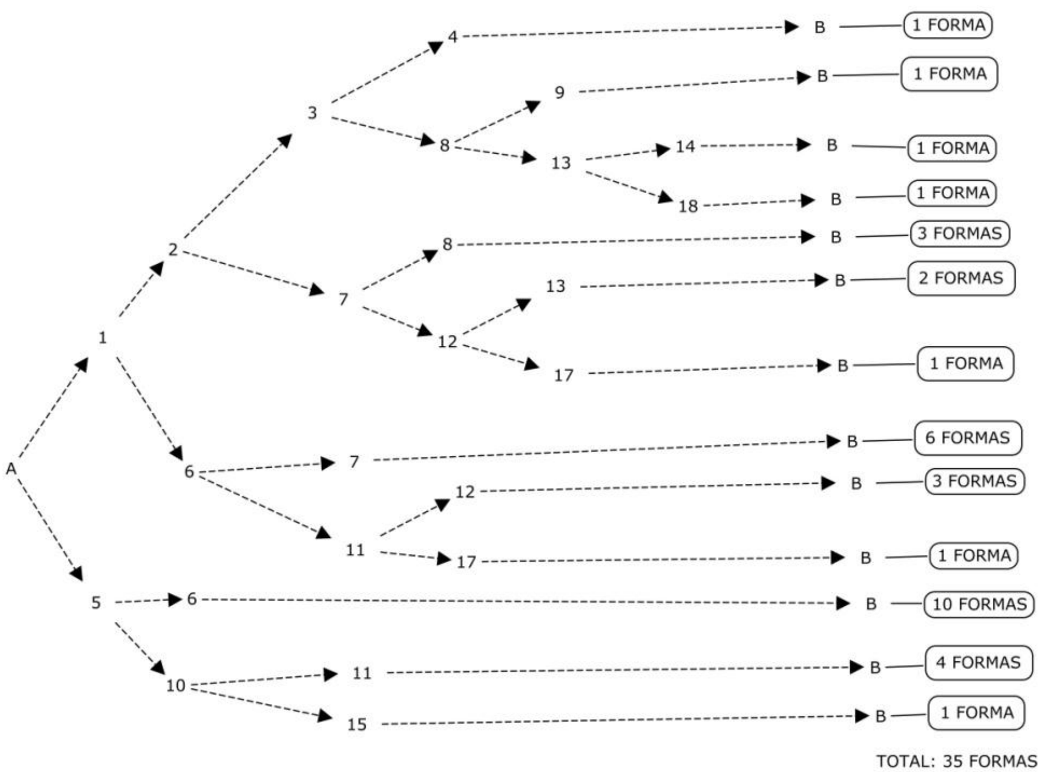


Figura 4. Diagrama de árbol de las posibles rutas del taxista

Respuesta: El taxista puede realizar el trayecto del punto A al punto B de 35 Formas diferentes.

El esquema que se utilizo para encontrar el total de las diferentes formas que tiene el taxista para llegar del punto A al punto B se conoce con el nombre de: **Diagrama de Árbol**.

DIAGRAMA DE ARBOL

Los diagramas de árbol son ordenaciones empleadas para enumerar todas las posibilidades lógicas de una secuencia de eventos, donde cada evento puede ocurrir en un número finito; Y proporcionan un método sistemático de enumeración objetiva de los resultados.

Cada una de las n_1 maneras de completar el primer paso puede representarse como una rama del árbol, cada una de las maneras de completar el segundo paso puede representarse con n_2 ramas que comienzan donde terminan las ramas originales, y así sucesivamente.

¿Para qué sirve?

- Un diagrama de árbol es un método gráfico para identificar todas las partes necesarias para alcanzar algún objetivo final. En mejora de la calidad, los diagramas de árbol se utilizan generalmente para identificar todas las tareas necesarias para implantar una solución.
- Se emplea para descomponer una meta u objetivo en una serie de actividades que deban o puedan hacerse. A través de la representación gráfica de actividades se facilita el entendimiento de las acciones que intervendrán.
- Permite a los miembros de un equipo de trabajo expandir su pensamiento al crear soluciones sin perder de vista el objetivo principal o los objetivos secundarios.
- Ubica a un equipo para que se dirija a situaciones reales versus teóricas. Asimismo, se dimensiona el nivel real de complejidad de algún proyecto y se puede prever el encontrarse con soluciones inviables antes del arranque.

Un diagrama de árbol es una especie de mapa de acontecimientos en donde se describen los eventos básicos que ocurren en un evento. Este gráfico está formado por segmentos de rectas y puntos. Los eventos que ocurren se denotan por puntos. Este diagrama puede ser dibujado de izquierda a derecha o de arriba hacia abajo, no hay restricciones para ello. La estructura se muestra en las siguientes figuras:

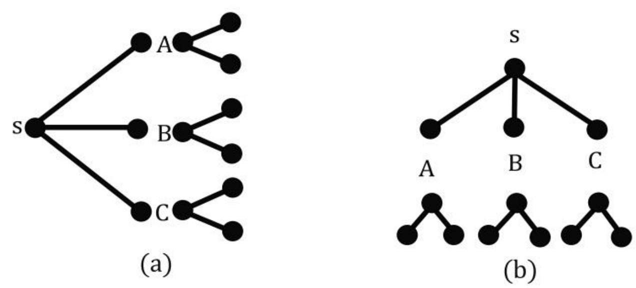


Figura 5. Estructuras de Diagrama de árbol

Una vez elaborado el diagrama de árbol, se puede contar directamente en él el número de posibilidades que se dan. O si no, se puede utilizar el árbol para deducir ese número. Al final, el árbol nos permitirá deducir una fórmula general para cada caso.

¿Por qué se le llama diagrama de árbol?. La respuesta a esta pregunta es sencilla, pues está formado por las siguientes partes: "raíz", "rama", "nodo" o "nudo", y "hoja", "nivel", como puede observarse están representadas en el las partes que corresponden a un árbol.

La raíz representa el nivel 0. Los nodos y las hojas están en un nivel u otro según las ramas que les separa de la raíz. Cada rama añade un nivel "camino": cualquier recorrido por las ramas del árbol desde raíz hasta alguna de las hojas.

Ya que en algunos casos se tienen demasiadas posibilidades donde el diagrama de árbol resulta demasiado tedioso utilizar; es por esta razón que se necesitan utilizar las técnicas de conteo.

Las técnicas de conteo, se fundamentan en dos principios importantes como son: el **"principio de la multiplicación"** y el **"principio de la adición"**, que se analizarán a continuación.

Estos principios se sustenta su utilización en lo siguiente: "**Se observa que una operación o actividad aparece en forma repetitiva y es necesario conocer las formas o maneras que se puede realizar dicha operación".**

PRINCIPIO DE LA MULTIPLICACIÓN

Situación Problemática: Suponga que una persona desea preparar un almuerzo para sus amigos y tiene dos recetas para la sopa, tres para el plato principal y dos para el postre. ¿De cuántas maneras puede esta persona hacer su menú?.

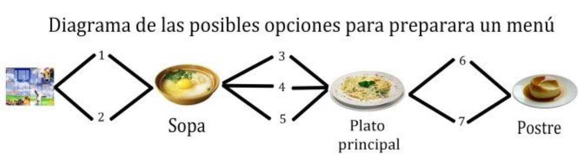


Figura 6. **Diagrama de posibles opciones para elegir en un menú.**

Siguiendo la trayectoria de los números las alternativas que tendrá son: 1-3-6 1--4-6 1-5-6 2-3-6 2-4-6 2-5-6 1-3- 7 1-4-7 1-5-7 2-3-7 2-4-7 2-5-7

Existen en total 12 maneras diferentes de preparar un delicioso almuerzo. Lo que significa que se tiene:

$$\underset{\text{opciones}}{\underline{2}} \quad X \quad \underset{\text{opciones}}{\underline{3}} \quad X \quad \underset{\text{opciones}}{\underline{6}} \quad = \quad 2 \times 3 \times 6 = 12$$

Sopa plato postres

A esta forma de obtener el resultado se conoce como **principio de la multiplicación.**

Enunciado del principio

Si se desea realizar una actividad que consta de r pasos, en donde el primer paso de la actividad a realizar puede ser llevado a cabo de N_1 maneras o formas, el segundo paso de N_2 maneras o formas y el último paso o r-ésimo paso de N_r maneras o formas, entonces esta actividad puede ser llevada a efecto de:

$N_1 \times N_2 \times ....\times\ N_r$ maneras o formas.

El principio multiplicativo implica que cada uno de los pasos de la actividad debe ser llevado a efecto, uno tras otro. La siguiente figura representa este enunciado.

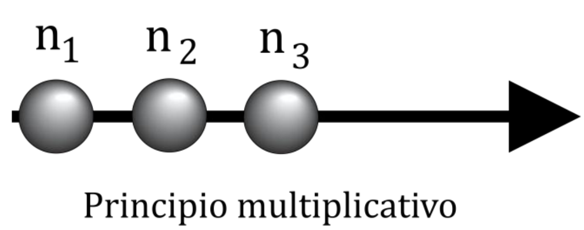


Figura 7. **Secuencia del principio multiplicativo**

Ejemplos de Aplicación

Ejemplo 1: Juan el alumno más inteligente del salón se saca un premio al final del año, el premio consiste en vacaciones todo pagado a cualquiera de 3 posibles lugares que le gustaría ir, usando cualquiera de los 2 medios de transporte disponibles, y acompañado de uno de los 3 familiares que lo pueden acompañar.

a) ¿Cuántas posibilidades diferentes se le presentan a Juan?.

b) Mencione dos de las posibilidades o alternativas que tiene Juan.

Solución

Actividad: Irse de vacaciones Juan

Opciones posibles para Juan		
Lugares	**Medios de transporte**	**Acompañar**
Playa	Bus	Mamá
Parque	Microbús	Papá
Piscinas		Hermano

$$\underset{N_1}{\underline{\;3\;}} \;X\; \underset{N_2}{\underline{\;2\;}} \;X\; \underset{N_3}{\underline{\;3\;}}$$

Las posibilidades que tiene Juan son:

$N_1 X N_2 X N_3 = 3 x 2 x 3 = 12.$

b) Dos posibilidades.

Posibilidad 1: Ir a la playa en bus y que lo acompañe su mamá.

Posibilidad 2: Ir a un parque en microbús y que lo acompañe la mamá.

Ejemplo 2: Carmen alumna del tercer año de bachillerato quiere ir al baile de su graduación, para dicha fiesta ella puede usar uno de cualquiera de sus 4 vestidos, uno de cualquiera de sus 3 pares de zapatos y una de sus 2 bolsas.

a) ¿De cuántas maneras diferentes puede asistir al baile?

b) Ya en el baile Carmen se junta con sus amigas María, Ana y Josefina cada una de ellas puede bailar con cualquier de los 5 jóvenes que están disponibles en la fiesta, ¿Cuántas parejas diferentes es posible formar?

Solución

Actividad: Ir Carmen al baile de graduación

a) ¿De cuántas maneras diferentes puede asistir al baile?

Opciones posibles de Carmen:

Vestidos $N_1 = 4$ Zapatos $N_2 = 3$ Bolsos $N_3 = 2$

Maneras que puede ir vestida Carmen al baile:

$4 X 3 X 2 = 24$

b) ¿Cuántas parejas diferentes es posible formar?

Amigas: María, Ana, Josefina, Carmen $N_1 = 4$

Jóvenes: J1, J2, J3, J4, J5 $N_2 = 5$

El número de parejas diferentes que se forman son: $N_1 X N_2 = 4 X 5 = 20$

PRINCIPIO DE LA SUMA O ADICIÓN

Situación Problemática: Suponga, ahora, que la persona que prepara el menú para sus amigos preparará pescado como plato principal. Para preparar el pescado, él encuentra cinco maneras diferentes de hacerlo al horno, dos para hacerlo frito y tres para prepararlo cocido. ¿De cuántas maneras diferentes puede cocinar su pescado?

Solución

Actividad: Preparar el almuerzo para sus amigos.

Alternativa de plato principal: cocinar el pescado.

Opciones de cocinar el pescado: **5** maneras al horno, **2** maneras de frito y **3** maneras de cocido.

Cada una de las maneras de preparar el pescado es excluyente de las otras dos; ya si el cocinero decide preparar el pescado cocido, ya no podrá prepararlo ni frito ni al horno; de igual manera sucede si decide hacerlo al horno o frito. Por lo tanto:

$$\underset{\text{Al horno}}{\underline{5}} + \underset{\text{Frito}}{\underline{2}} + \underset{\text{Cocido}}{\underline{3}} = 5 + 2 + 3 = 10$$

Son 10 maneras diferentes de cocinar el pescado.

A esta forma de obtener el resultado se conoce como: **Principio de la suma.**

Enunciado del principio

Si se desea llevar a efecto una actividad, la cual tiene formas alternativas para ser realizada, donde la primera de esas alternativas puede ser realizada de N_1 maneras o formas, la segunda alternativa puede realizarse de N_2 maneras o formas .... y la última de las alternativas puede ser realizada de N_r maneras o formas, entonces esta actividad puede ser llevada a cabo de, $N_1 + N_2 + ...+ N_r$ maneras o formas .

En este principio las alternativas tienen la característica de ser mutuamente excluyentes; es decir, sólo una alternativa puede llevarse a cabo y no la dos a la vez. La siguiente figura representa este enunciado.

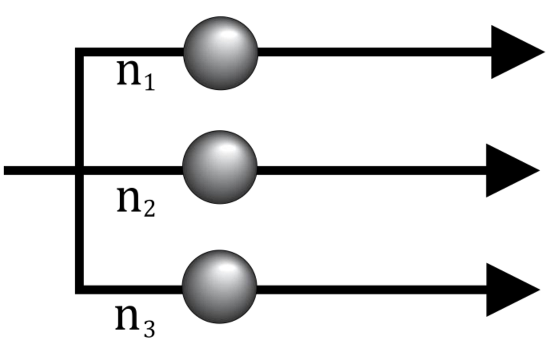


Figura 8. Secuencia del principio aditivo.

Ejemplos de Aplicación

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

Figura 9. Rio Lempa, considerado el más caudaloso de El Salvador.

Solución

Actividad: Cruzar el rio

Alternativas: Hacerlo en bote, lancha o deslizador.

Si cruza el rio en bote no puede cruzarlo en lancha a la vez o en deslizador, por eso estas alternativas son mutuamente excluyentes. Y una vez cruza el rio no necesita el otro medio para cruzarlo; porque la actividad ya ha sido realizada.

Las posibilidades que hay son: 3 botes $= N_1$ ó 2 lanchas $= N_2$ ó 1 deslizador $= N_3$

Lo que indica que para cruzar el rio se puede hacer de: $N_1 + N_2 + N_3 = 3 + 2 + 1 = 6$ formas

Como se puede observar en el siguiente diagrama de árbol.

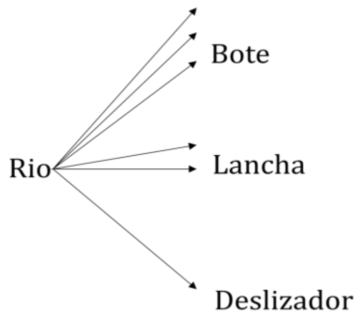


Figura 10. Diagrama de árbol acerca de las diferentes formas de cruzar el rio.

¿Cómo se puedes distinguir cuando hacer uso del principio multiplicativo y cuando del aditivo?

Opción 1: "Es muy simple, cuando se trata de una sola actividad, la cual requiere para ser llevada a efecto de una serie de pasos, entonces haremos uso del **principio multiplicativo** y si la actividad a desarrollar o a ser efectuada tiene alternativas para ser llevada a cabo, haremos uso del **principio aditivo."**

Opción 2: Si se trata de una secuencia de acciones, deberemos usar el **principio multiplicativo**. Si se trata de una sola acción que presenta distintas alternativas de realización, deberemos usar el **principio aditivo**.

Es esencial considerar la formulación del problema, en términos del principio fundamental del conteo, dibujando un diagrama de árbol, identificando cuándo es aplicable el principio multiplicativo y cuándo el aditivo, cuando importa el orden de los resultados y cuándo no, y cuándo es permisible repetir resultados y cuándo no.

FACTORIAL DE UN NÚMERO

En el análisis combinatorio interviene con mucha frecuencia el concepto de factorial de un entero no negativo *n*. Para todo entero positivo *n*, el factorial de *n* ó *n* factorial ó factorial de n, se define como el producto de todos los números enteros positivos desde 1 hasta *n*. Es decir

$$n! = 1 \times 2 \times 3 \times 4 \times \ldots \times (n-1) \times n$$

La multiplicación anterior se puede simbolizar casi siempre como ***n!.***

Si n=1 entonces $n! = 1! = 1$ y

Si n=0 $n! = 0! = 1$

La excepción es el caso de *0!*. El cual conviene definirlo como igual a 1 con objeto de preservar la validez de las fórmulas en casos extremos. Muchas calculadoras traen una tecla factorial.

PERMUTACIONES

El termino permutar significa "cambiar el orden de un grupo de elementos" o "variar la disposición u orden en que estaban dos o más cosas".

Se denomina permutación de n elementos, a los diferentes grupos o maneras en que se pueden ordenar esos n elementos. La permutación implica orden en la colocación de los elementos, y se debe tener en cuenta lo siguiente:

- Los grupos están formados por los mismos n elementos.
- Dos grupos son distintos si el orden en que aparecen los elementos es diferente.

En las permutaciones como interesa el orden, un grupo que contiene (AB) es diferente al grupo (BA).

PERMUTACIONES SIN REPETICION (ORDINARIAS)

Se llaman sin repetición porque los n elementos del grupo son todos distinguibles (distintos) entre sí; es decir, en el grupo los n elementos que lo forman no hay elementos repetidos. Y lo que se cuenta son todas las formas posibles y distintas de ordenarlos.

Se llaman permutaciones de n elementos a las diferentes maneras en que se pueden ordenar esos n elementos; todas las permutaciones constan de los mismos n elementos, pero se consideran diferentes, por el orden en que se colocan éstos. Su notación es: P_n

Las permutaciones sin repetición de m elementos se definen como las distintas formas de ordenar todos esos elementos distintos, por lo que la única diferencia entre ellas es el orden de colocación de sus elementos.

Actividad introductoria: Cinco chicos, entre los cuales están Santiago y Pedro, se ordenan en fila, al azar.

Supongamos que los cinco chicos tienen los nombres siguientes: Santiago, Pedro, Oscar, Facundo y Juan.

Primer lugar	Segundo lugar	Tercer lugar	Cuarto lugar	Quinto lugar

En el primer lugar de la fila podría haberse ubicado cualquiera de los 5 chicos. Entonces para el primer lugar: **5 posibilidades**.

Si en el primer lugar ya se ubico uno de los cinco (cualquiera) para el segundo lugar solo quedan por ubicar cuatro. Entonces para segundo lugar: **4 posibilidades**.

Si para el segundo lugar ya se ubico uno de los cuatro (cualquiera) para el tercer lugar solo quedan por ubicar tres. Entonces para el tercer lugar: **3 posibilidades.**

Si para el tercer lugar ya se ubico uno de los tres (cualquiera) para el cuarto lugar solo quedan por ubicar dos. Entonces para el cuarto lugar: **2 posibilidades.**

Si para el cuarto lugar ya se ubico uno de los dos (cualquiera) para el quinto lugar solo quedan por ubicar uno. Entonces para el quinto lugar: **1 posibilidad.**

Se observa que en cada uno de los lugares va disminuyendo en una unidad, es por el hecho que un chico por naturaleza no puede estar dos veces en la misma fila (elementos sin repetición), cada fila van a tener el mismo número de chicos (cinco) pero por la ubicación del lugar que ocupe cada chico las agrupaciones van hacer diferentes.

Resumiendo, el número total de casos posibles es:

Primer lugar		Segundo lugar		Tercer lugar		Cuarto lugar		Quinto lugar		
5	X	4	X	3	X	2	X	1	=	120

Este número 120, obtenido al multiplicar todos los números naturales menores o iguales a 5 en forma descendente, se conoce como factorial de 5 y se simboliza $5!$.

Generalizando: Para calcular el número de permutaciones que se pueden formar con los n elementos, se hacen las siguientes consideraciones: la elección del primer elemento se puede hacer de *n* maneras diferentes; la elección del segundo elemento se puede hacer de (n-1) maneras diferentes,..., y la elección del n-ésimo elementos sólo se puede hacer de una manera.

Ahora, invocando el principio fundamental del conteo se tiene: $P_n = n(n-1)(n-2)\ldots..x3x2x1$, que conduce a la definición de factorial de n. Por lo tanto: $P_n = n!$

Definición: Se llaman permutaciones ordinarias o sin repetición de n elementos, y se denota por P_n, a los distintos grupos que se pueden formar, de tal manera que en cada grupo estén los n elementos y que un grupo se diferencie de los demás en el orden de colocación de los elementos. Además se tiene que:

$$P_n = n! \qquad \textbf{Ecuación 1}$$

Ejemplo 3: De las 120 posibilidades anteriores, ¿Cuántas hay en las cuales Santiago queda fijo en el segundo lugar y Pedro en el quinto lugar?.

Solución

Con estas restricciones los lugares quedan de la siguiente forma:

Primer lugar	Santiago	Tercer lugar	Cuarto lugar	Pedro

Lo que significa que los otros tres chicos (Oscar, Juan y Facundo) pueden cambiar de ubicación ocupando los otros tres lugares (el primero, el tercero y el cuarto). Mientras Santiago y Pedro conserven los lugares establecidos no importa en qué orden se ubiquen los otros tres chicos.

El primer lugar puede ser ocupado por Juan, Oscar o Facundo (3 posibilidades). Pero una vez que se ubicó Juan en el primer lugar, el tercero sólo puede ser ocupado por Oscar o Facundo (2 posibilidades). Y si Oscar ocupa el tercer lugar, sólo Facundo queda disponible para ocupar el cuarto. Es decir.

Primer lugar	Santiago	Tercer lugar	Cuarto lugar	Pedro	
3	X	2	X 1		= 120

Al ubicar fijos a Santiago y a Pedro en segundo y quinto lugar respectivamente, lo que hizo fue permutar a los otros tres chicos en los otros tres lugares. Es decir: $P_3 = 3! = 3 \times 2 \times 1 = 6$.

Por lo tanto, el número de posibilidades de las 120 en donde Santiago se encuentra en el segundo lugar y Pedro en el quinto son: 6 posibilidades.

PERMUTACIONES CON REPETICION

Algunas veces no todos los elementos son distintos, sino que parte de ellos se repiten. En este caso se tienen n elementos de los cuales n_1 son de un tipo, n_2 son de otro tipo distinto y n_k son del k-ésimo tipo, en donde $n_1 + n_2 + \ldots\ldots + n_k = n$.

Su notación es: $PR_n^{n_1,\ n_2,\ldots\ldots,\ n_k}$

Actividad introductoria: Tengo 3 caramelos de piña, dos de menta y uno de fresa. Los voy a repartir entre mis 6 amigos. ¿De cuántas maneras puedo hacerlo?

Se Llamarán los caramelos: P, P, P, M, M, F.

Si a mi amiga Ana y a mi amigo Benito les doy un caramelo de piña, no importa qué caramelo de piña le dé a cada uno. A ellos les parecerá igual de bien. Los 3 caramelos de piña son **indistinguibles** entre sí y también 2 caramelos de menta son indistinguibles entre sí.

Supongamos que mis amigos están en el siguiente orden: Ana, Benito, Carlos, Delia, Elena, Fran.

Si empiezo a repartir por Ana: Tengo 6 posibilidades: P, P, P, M, M o F.

Una vez he dado el caramelo a Ana, sigo con Benito: Ahora tengo 5 posibilidades: Si le di a Ana un caramelo de piña (P), me quedan para Benito: P, P, M, M o F.

Pero si le di a Ana otro de los caramelos de piña (otra vez P), me quedan P, P, M, M o F.

También pude darle un caramelo de menta (M), entonces quedan para mi amigo P, P, P, M o F.

... Lo que represente en el siguiente árbol.

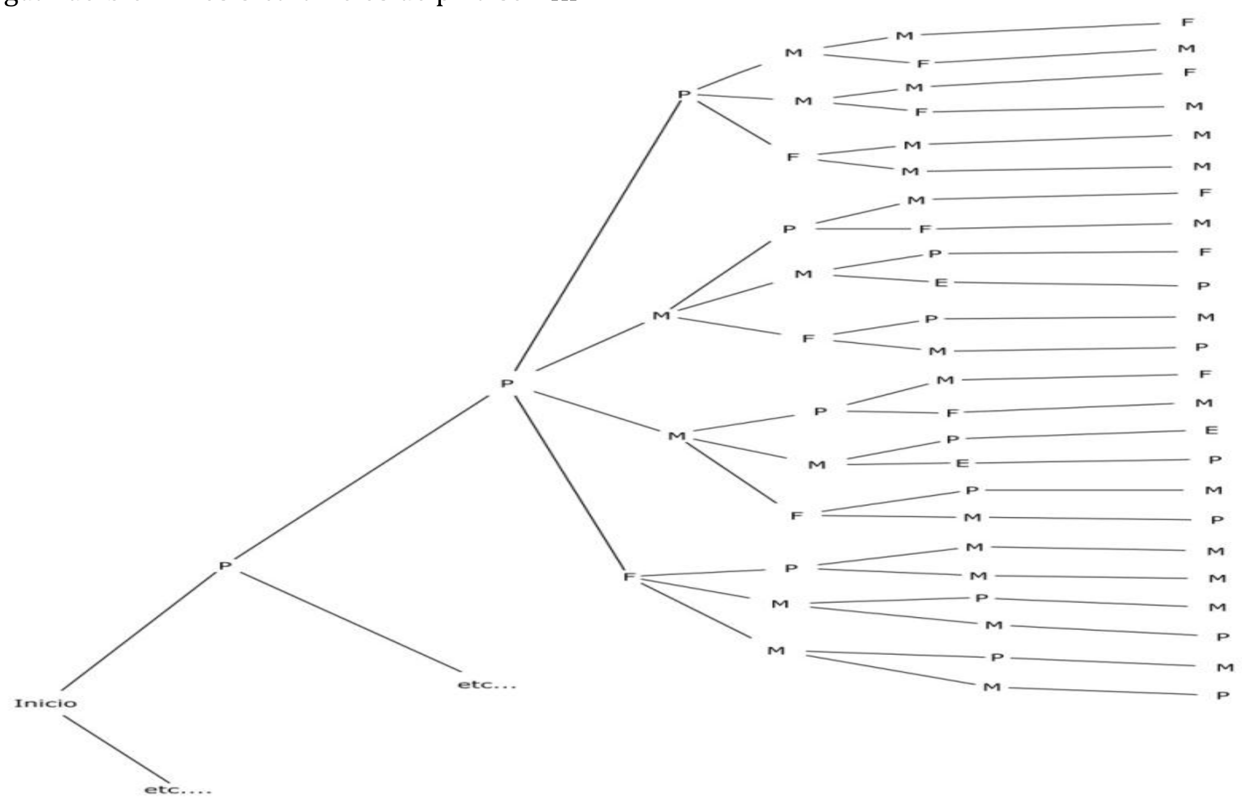


Figura 11. Diagrama de árbol que muestra los posibles resultados de la repartición de caramelos

Pero en este árbol hay caminos repetidos!

Si se cuenta todos los caminos, es como el árbol de las permutaciones normales .Por tanto, la fórmula sería: $P_n = n!$. Los grupos repetidos se deben descartar.

Por cada 3 amigos a quienes haya dado los 3 caramelos de piña, las 3! maneras de dar 3 caramelos de piña a 3 amigos. Eso se hace dividiendo entre 3!. (Si les doy piña a: A, B y C; tengo $P_3 = 3! = 6$ formas de hacerlo y sólo queda una ya que son iguales: ABC, ACB, BAC, BCA, CAB, CBA)

También, por cada 2 amigos a quienes haya dado los 2 caramelos de menta, tengo que descartar las 2! maneras de repartir 2 caramelos de menta a 2 amigos. Eso se hace dividiendo entre 2!.

Y finalmente, por cada posible amigo a quien se lo dé, descartaré las 1! maneras de repartir 1 caramelos de fresa a un amigo. Eso se hace dividiendo entre 1!.

Por lo tanto queda la siguiente cantidad de posibles repartos: $\frac{6!}{3!21!} = 60$ maneras.

Generalización: Existen n_1 permutaciones lineales que conducen a una sola permutación distinguible, porque las permutaciones de los n_1 elementos iguales no son distinguibles entre sí; existen n_2 permutaciones lineales que conducen a una sola permutación distinguible, porque las permutaciones de los n_2 elementos iguales no son distinguibles entre sí; ... y existen n_r permutaciones lineales que conducen a una sola permutación distinguible, porque las permutaciones de los n_r elementos iguales no son distinguibles entre sí. De manera que por cada permutación distinguible hay n_1 permutaciones lineales equivalentes, por cada permutación distinguible hay n_2 permutaciones lineales equivalentes, ..., y por cada permutación distinguible hay n_r permutaciones lineales equivalentes.

Entonces, para calcular el número de permutaciones distinguibles de n elementos, se divide el número de permutaciones lineales de n objetos entre las $n_1!$ permutaciones equivalentes, entre las $n_2!$ permutaciones equivalentes,..., y entre las $n_r!$ permutaciones equivalentes; es decir:

$$PR_n^{n_1,\ n_2,\dots\dots,\ n_k} = \frac{n!}{n_1!n_2!\dots\dots n_r!}$$

Definición: Se llaman permutaciones con repetición de n elementos, distribuidos en k grupos de $n_1, n_2, \dots, n_k$ elementos indistinguibles, respectivamente, de tal forma que $n_1 + n_2 + \dots + n_k = n$, a los distintos grupos que se pueden formar con los n elementos, de tal forma que cada una de ellas se diferencie de las demás en el orden de colocación de sus elementos, excluyendo las reordenaciones de elementos indistinguibles (esto es, que pertenecen a un mismo grupo). Se calcula por:

$$PR_n^{n_1,\ n_2,\dots\dots,\ n_k} = \frac{n!}{n_1!n_2!\dots\dots n_r!}$$ **Ecuación 2**

Ejemplo 4: Si para fijar una placa se cuenta con 7 tornillos: 2 son de acero al carbón, 3 son de acero inoxidable y *2* son de bronce. ¿De cuántas maneras diferentes se pueden colocar tales tornillos, si se distingue el material del que están hechos?

Figura 12. **Tornillos de acero y de bronce**

Solución

n=7 Total de tornillos
n_1=2: Grupo de tornillos de acero al carbón
n_2=3: Grupo de tornillos de acero inoxidable
n_3=2: Grupo de tornillos de bronce
Por lo tanto: $n = n_1 + n_2 + n_3 = 2 + 3 + 2 = 7$

Sustituyendo

$$PR_7^{2\ 3\ 2} = \frac{7!}{2!3!2!} = \frac{7x6x5x4x3!}{2!3!2!}$$

$$= \frac{7x6x5x4}{2!2!} = \frac{840}{4} = 210$$ Maneras diferentes.

Por lo tanto, hay 210 maneras diferentes de colocar los 7 tornillos, si se distingue el material de que están hechos.

PERMUTACIONES CIRCULARES

A una disposición de elementos en cadena cerrada o anillo se le llama ***permutación circular o cíclica.*** Esto es, cuando se ordenan elementos en una curva cerrada. Por ejemplo, en una mesa redonda, en un llavero, la rueda de la fortuna, etc.

Actividad introductoria: Se quiere confeccionar un collar con n cuentas de colores, todas de distinto color.

De cuántas formas se puede formar el collar si se utilizan todas ellas?

El número de ordenaciones distintas de n objetos distintos es $P_n = n!$, sin embargo, las cuentas de un collar quedan uniformemente distribuidas en una circunferencia y cualquier giro que se efectué no cambia el collar como se muestra en la siguiente figura:

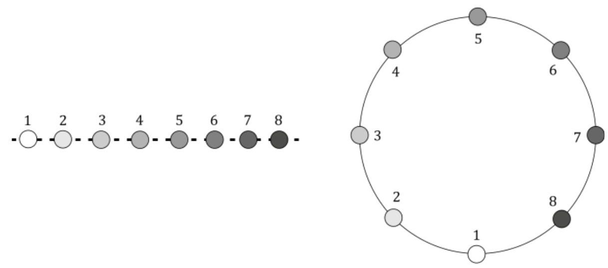


Figura 13. **Permutación circular**

Pero sí la configuración en línea que lo género: hay más ordenaciones en línea que circulares; el problema es cuántas.

En la figura anterior, los 8 giros que se representan, no modifican el collar; de hecho, para confeccionar el collar importa la posición relativa de unas cuentas respecto a otras, más no el orden en que estas han sido colocadas: esto es, se pueden formar $\frac{8!}{8} = \frac{8x7!}{8} = 7!$ collares distintos con 8 cuencas diferentes.

En general, si el collar está formado por n cuentas se podrán formar $(n-1)!$ collares.

Generalización: Existen n permutaciones lineales que, al ser colocadas en círculo, conducen a una misma permutación circular, porque cada elemento queda en la misma posición relativa respecto a los (n-1) elementos restantes; de manera que por cada permutación circular hay *n* permutaciones lineales equivalentes. Entonces, para calcular el número de permutaciones circulares de *n* objetos, se divide el número de permutaciones lineales de *n* elementos entre las n permutaciones equivalentes:

$$PC_n = \frac{P_n}{n} = \frac{n!}{n} = (n-1)!$$ **Ecuación 3**

Definición: Se llaman permutaciones circulares (sin repetición) de n elementos, se denota por PC_n, a los distintos grupos que se pueden formar, de tal manera que en cada grupo entren los n elementos y que un grupo se diferencie de los demás en la posición relativa de los elementos unos respecto a los otros. Además se tiene que: $PC_n = (n-1)!$

Esta fórmula se obtiene siempre que se fije cualquiera de los n elementos en el grupo circular, los restantes **n-1** elementos se consideran como una permutación lineal, la cual es posible hacer de **(n-1)!** maneras.

Ejemplo 5: Junta de comité. ¿De cuántas maneras diferentes se pueden colocar 6 personas, para una junta de comité?

a) Las 6 personas en fila.
b) Las 6 personas en fila, si dos personas deben quedar juntas.
c) Alrededor de una mesa.
d) Alrededor de una mesa, si dos personas deben quedar siempre juntas.

Solución a) Permutaciones sin repetición: n= 6

Las 6 personas solo se deben cambiar de lugar.

Entonces:

$$P_6 = 6! = 6x5x4x3x2x1 = 720 \text{ Maneras.}$$

Solución b)

Como hay la restricción que dos personas tienen que quedar juntas entonces lo que se va a permutar se reduce a 5 grupos; pero además hay que permutar el grupo de las dos personas que van a quedar juntas. Entonces:

$$P_5P_2 = 5!x2! = (5x4x3x2x1)\ (2x1)$$

$$= 120x2 = 240 \text{ Maneras}$$

Solución c) Permutaciones circulares: n=6

$$PC_6 = (6-1)! = 5! = 120 \quad \text{Maneras}$$

Solución d) Similar al caso b)

$$P_2PC_5 = 2!(5-1)! = 2!4! = (2x1)\ (4x3x2x1)$$

$$= 2 \text{ x } 24 = 48 \text{ Maneras}$$

VARIACIONES

Dado un conjunto de n elementos, se denominan variaciones de tamaño m a todos los conjuntos de m elementos escogidos de entre los n, tales que un conjunto difiere de otro en al menos un elemento o en el orden en que se consideran los elementos. Las variaciones implican orden en la colocación de los elementos, y se debe tener en cuenta lo siguiente:

- De los n elementos, sólo m intervienen en las agrupaciones.
- Las agrupaciones de m elementos son distintas si difieren en algún elemento o en su orden de colocación.

VARIACIONES SIN REPETICIÓN (ORDINARIAS)

Se llaman ordenaciones de n elementos de orden m a las diferentes maneras de escoger secuencialmente m elementos de entre n posibles, de modo cada una de las ordenaciones es distinta de las demás, si difiere en alguno de sus objetos o en el orden de ellos. Y denota por:

$$V_{n,m} \text{ o } V_n^m$$

Actividad introductoria: Se desea formar un comité de aula para la organización de un evento cultural en un colegio. Dicho comité está formado por tres alumnos que harían las veces de delegado, vocal y secretario. La clase está formada por 40 alumnos. Nos planteamos resolver la siguiente cuestión: ¿de cuántas formas puede constituirse el comité si una persona no puede ocupar más que un cargo?

Como un estudiante no puede tener más que un cargo, el delegado podrá ser elegido entre los 40 alumnos de la clase; una vez que esté ha sido elegido, el cargo de vocal podrá ser tomado por uno de los 39 alumnos restantes; por último, el cargo de secretario puede ser tomado por uno de los 38 alumnos restantes. Es decir, existen 40 x 39 x 38 formas de constituir el comité.

Generalización: Para calcular el número de ordenaciones de m elementos que se pueden formar con los *n elementos* disponibles, se hacen las siguientes consideraciones: la elección del primer elemento se puede hacer de *n* maneras diferentes; la elección del segundo elemento se puede hacer de (n – 1) maneras diferentes,..., y la elección del m-ésimo objeto se puede hacer de (n – m + 1) maneras diferentes. Ahora, invocando el principio fundamental del conteo se tiene:

$$V_n^m = n(n-1)(n-2)......(n-m+2)(n-m+1)$$

Expresión que al multiplicar y dividir por (n-m) conduce a:

$$V_n^m = \frac{n(n-1)(n-2)......(n-m+2)(n-m+1)(n-m)!}{(n-m)!}$$

Y utilizando la fórmula fundamental de factorial, se tiene:

$$V_n^m = \frac{n!}{(n-m)!}$$

Definición: Se llaman variaciones ordinarias o sin repetición de n elementos, tomados de m en m, se denota V_n^m, a los distintos grupos que se pueden formar con los n elementos, de tal forma que en cada grupo entren m elementos distintos y que un grupo se diferencie de los demás, bien en alguno de sus elementos, bien en su orden de colocación. Se tiene:

$$V_n^m = \frac{n!}{(n-m)!} \qquad \textbf{Ecuación 4}$$

Relación de Variaciones con Permutaciones

Para la deducción de esta fórmula, se ha considerado implícitamente que el número m de elementos a elegir es menor o igual que el número de objetos disponibles: m ≤ n, lo que equivale a no permitir la repetición de elementos en una misma ordenación. El caso particular en el que m = *n*, conduce a la obtención de las ordenaciones de *n* objetos tomados todos a la vez, es decir, a la obtención de la permutaciones de los *n* objetos:

$$V_n^n = \frac{n!}{(n-n)!} = \frac{n!}{0!} = n! = P_n$$

Ejemplo 6: Salon de Clase .¿De cuántas maneras diferentes se pueden sentar los *52* alumnos del grupo de la asignatura de Probabilidad en un salón que dispone de 60 pupitres?

Solución

El primer alumno que entra al salón puede escoger su lugar de entre 60 posibles, el segundo puede escoger lugar de entre 59 posibles,... y así, sucesivamente, hasta el alumno número 52, que puede escoger lugar de entre 9 posibles.

Evidentemente, 8 de los 60 lugares quedarán vacíos; se trata de calcular las ordenaciones de *60* objetos de orden 52:

$$V_{60}^{52}=\frac{60!}{(60\text{-}52)!}=\frac{60!}{8!}=\frac{60x59x58x.....x9x8!}{8!}$$

$$= 60x59x58x.....x9 = 2.06374x10^{77} \text{ Maneras}$$

VARIACIONES CON REPETICION

Se llama ordenaciones con repetición de n objetos, de orden m a las diferentes maneras de efectuar secuencialmente m acciones, cada una de las cuales se puede presentar de n distintas maneras. El hecho de permitir la repetición de elementos, hace que el valor de m no esté restringido, pues el número m de acciones a efectuar puede ser mayor al número n de maneras en que puede presentarse cada una de ellas. Y se denota por: $VR_{n,m}$ o VR_n^m.

Actividad introductoria: Se desea formar un comité de aula para la organización de un evento cultural en un colegio. Dicho comité está formado por tres alumnos que harían las veces de delegado, vocal y secretario. La clase está formada por 40 alumnos. Supongamos ahora que una misma persona puede ocupar más de un cargo, esto es, una persona puede ser a la vez vocal y delegado, por ejemplo. Esta situación es real: muchas veces una misma persona ocupa más de un cargo dentro de una institución. Por ejemplo, profesor y coordinador de ciencias, alumno y miembro de la banda de música del colegio, etc. Entonces se pretende resolver la siguiente cuestión: si en un aula hay n estudiantes, de cuántas formas puede constituirse un comité de m estudiantes si una persona puede ocupar más que un cargo?.

Los 3 cargos deben ser ocupados por alguno de los 40 estudiantes que conforman un aula. Como un estudiante sí puede tener más que un cargo, el delegado podrá ser elegido entre los 40 alumnos de la clase; una vez que este ha sido elegido, el cargo de vocal podrá ser tomado por uno cualquiera de los estudiantes, incluido el delegado electo; por último, el cargo de secretario puede ser tomado igualmente por cualquiera de los 40 estudiantes. Es decir, existen 40 x 40 x 40 formas de constituir el comité.

Al igual que en la anterior situación, el método descrito puede ser extendido para determinar el número de comités de m estudiantes que se pueden formar en un aula de n estudiantes (n $\geq$ m), pudiendo un alumno tener más de un cargo: $n^{(m \text{ veces})}.........n=n^m$

Definición: Se llaman variaciones con repetición de n elementos, tomados de m en m, denotaremos, $VR_{n,m}$, a los distintos grupos que se pueden formar con los n elementos, de tal manera que en cada grupo entren m elementos iguales o distintos y que un grupo se diferencie de los demás, bien en algún elemento, bien en su orden de colocación. Se tiene:

$$VR_n^m = n^m \qquad \textbf{Ecuación 5}$$

Ejemplo 7: Monedas. Considere el experimento consistente en lanzar tres monedas simultáneamente y observar las caras que quedan hacia arriba. Determine el número de maneras en que puede ocurrir tal experimento.

Solución

Nótese que el experimento consistente en lanzar tres monedas simultáneamente es equivalente al experimento de lanzar una moneda tres veces consecutivamente.

Por lo tanto, n=2 y m=3 (m > n). Entonces el número de maneras que se puede ocurrir el experimento es: $VR_2^3 = 2^3 = 8$

COMBINACIONES

En lenguaje común, combinar es: "unir cosas diversas, de manera que formen un compuesto". Al igual que las variaciones y las permutaciones, el concepto de combinación tiene un significado muy concreto en matemáticas: brevemente, es el número de conjuntos de un determinado número de elementos que se pueden formar con n elementos, sin importar el orden de selección, sino qué elementos se toman.

El orden en que se seleccionan los objetos era importante para calcular el número de permutaciones. Sin embargo, ahora se quiere saber el número de grupos posibles de n elementos tomados m a la vez sin importar el orden de los elementos seleccionados. Estos grupos donde no interesa el orden se conocen como combinaciones.

Dado un conjunto de n elementos, se denomina combinaciones de tamaño m a todos los conjuntos que se pueden formar con m elementos tomados de entre los n elementos, de modo que cada conjunto difiera de los demás en por lo menos un elemento. En el caso de las combinaciones se tiene en cuenta los elementos que tiene el subconjunto independientemente de la ordenación que éstos tenga.

En las combinaciones como no interesa el orden, por ejemplo un grupo que contiene (**AB**) es igual al grupo (**BA**).

COMBINACIONES SIN REPETICIÓN

Se llaman combinaciones de *n* objetos de orden ***m*** a los distintos grupos que se pueden formar al escoger secuencialmente ***m*** *elementos* de entre *n* posibles, de modo que cada una de las combinaciones es distinta de las demás, si difiere en uno de sus elementos por lo menos, sin importar el orden. Y se denota por: C_n^m

Actividad introductoria: En el problema de la formación de los comités de aula, el orden de elección de los estudiantes es relevante, puesto que los cargos de delegado, vocal y secretario no son equiparables. Sin embargo, si el comité está formado por tres personas que desempeñaran cargos similares, entonces no es relevante que un estudiante sea elegido en primer, segundo o tercer lugar, sino el hecho mismo de haber sido elegido. Como se ha visto, si el orden de elección es importante (y un alumno no puede tener sino un cargo), existen 40 x 39 x 38 formas de constituir los comités, pero si el orden no importa, hay que dividir esta cantidad por 6, puesto que dados 3 estudiantes, podemos organizarlos de 6 formas distintas (P_3). Así, existen $\frac{40x39x38}{6}$ formas de organizar los comités si los tres integrantes van a desempeñar labores similares.

En general, el razonamiento es válido si es preciso escoger, sin importar el orden, ***m*** estudiantes de entre n (n ≥ m), el número de comités que se pueden formar son: $\frac{V_n^m}{P_m} = \frac{n!}{(n-m)!m!}$

Definición: Se llaman combinaciones ordinarias o sin repetición de n elementos, tomados de ***m*** en ***m***, denotaremos C_n^m, a los diferentes conjuntos de ***m*** elementos distintos. Un conjunto se diferencie de los demás en, al menos, un elemento (no importa el orden de colocación o selección). Se tiene:

$$C_n^m = \binom{n}{m} = \frac{V_n^m}{P_m} = \frac{n!}{(n\text{-}m)!m!} \quad \textbf{Ecuación 6}$$

Ejemplo 8: En una asamblea de socios de una importante empresa del país, compuesta de 7 hombres y 5 mujeres, se acuerda conformar una comisión de verificación de actividades comerciales en la región. Esta comisión debe estar compuesta por 3 hombres y 2 mujeres. ¿De cuántas maneras puede escogerse dicha comisión?

Solución

De los 7 hombres pueden seleccionarse 3. Esto es:

$$C_7^3 = \binom{7}{3} = \frac{7!}{(7\text{-}3)!3!} = \frac{7!}{4!3!} = \frac{7x6x5x4!}{4!3!}$$

$= \frac{7x6x5}{3!} = \frac{7x6x5}{3x2x1} = \frac{210}{6} = 35$ Posibles maneras de seleccionar 3 hombres de un conjunto de 7.

De las 5 mujeres pueden seleccionarse 2. Esto es: $C_5^2 = \binom{5}{2} = \frac{5!}{(5\text{-}2)!2!} = \frac{5!}{3!2!} = \frac{5x4x3!}{3!2!}$

$= \frac{5x4}{2!} = \frac{5x4}{2x1} = \frac{20}{2} = 10$ Posibles maneras de seleccionar 2 mujeres de un conjunto de 5.

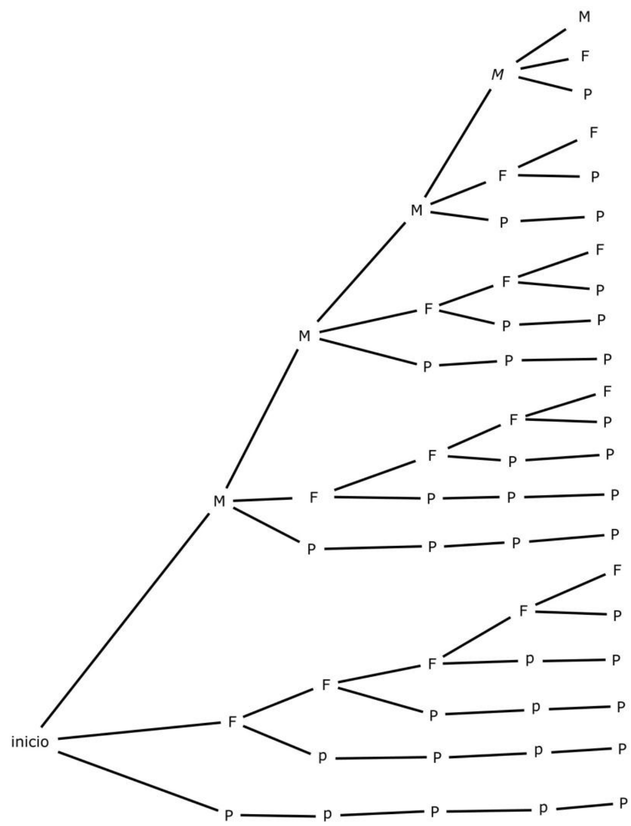
Por el principio de la multiplicación, la comisión puede escogerse de: 35x10 =350 maneras diferentes.

COMBINACIONES CON REPETICIÓN

Se llama combinaciones con repetición de n elementos tomados de m en m, o de orden m, a los distintos grupos de n elementos iguales o distintos que se pueden hacer con los n elementos que tenemos, de forma que dos grupos se diferencian en algún elemento pero no en el orden de colocación. Se representa por $CR_{n,m}$ ó CR_n^m.

Actividad introductoria: Supongamos que tengo un montón de caramelos de menta, de fresa y de piña, y que quiero confeccionar bolsitas de 5 caramelos cada una para repartirlas entre mis amigos el día de mi cumpleaños. ¿Cuántas bolsitas diferentes puedo hacer?

Para resolver este problema, empezaremos directamente por montar el árbol:



Observar que el árbol tiene la misma estructura que un árbol de combinaciones sin repetición. Parece que sea el árbol de las combinaciones de 7 elementos tomados de 5 en 5.

Por lo tanto, para calcular las combinaciones con repetición de 3 elementos tomados de 5 en 5, tengo que calcular $C_7^5 = \frac{7!}{(7-5)!5!} = \frac{7!}{2!5!} = \frac{7x6x5!}{2!5!} = \frac{7x6}{2} = \frac{42}{2} = 21$

Tengo 21 maneras diferentes de llenar una bolsa de 5 caramelos de entre un montón de caramelos de fresa, de menta y de piña.

Definición: Se llama combinaciones con repetición de n elementos tomados de m en m, a los diferentes grupos que pueden formarse con los n elementos dados, tomados de m en m, en los que pueden aparecer elementos repetidos, de modo que dos grupos difieren entre sí cuando, al menos, un elemento es distinto.

$$CR_n^m = \binom{n+m-1}{m} = \frac{(n+m-1)!}{m!(n-1)!} \qquad \textbf{Ecuación 7}$$

Se aplica este concepto cuando nos interese contar de cuántos modos podemos hacer una determinada selección en la que pueden aparecer elementos repetidos, pero no es significativo el orden en que hayan ido saliendo

Ejemplo 9: En una pastelería hay 6 tipos distintos de pasteles. ¿De cuántas formas se pueden elegir 4 pasteles?.

Solución

Si nos gusta un pastel lo podemos pedir hasta cuatro veces.

En este caso no importa el orden en que se eligen los pasteles y pueden repetir.

n=6 y m=4

Sustituyendo en la formula se tiene:

$$CR_6^4=\frac{(6+4-1)!}{4!(6-1)!}=\frac{9!}{4!5!}=\frac{9x8x7x6x5!}{4!5!}$$

$$=\frac{9x8x7x6}{4!}=\frac{9x8x7x6}{4x3x2x1}=\frac{3024}{24}=126$$

Por lo tanto hay 126 formas en que se pueden elegir los 4 pasteles.

<u>¿Cómo diferenciar estas técnicas?</u>

CUADRO RESUMEN

TECNICAS	CARACTERISTICAS	FÓRMULA
PERMUTACIONES SIN REPETICIÓN	▪ Importa el orden. ▪ Intervienen todos los elementos. ▪ No pueden repetirse los elementos.	$P_n=n!$
PERMUTACIONES CON REPETICIÓN	▪ Importa el orden. ▪ Intervienen todos los elementos. ▪ Existen elementos iguales entre sí.	$PR_n^{n_1, n_2,......, n_k}=\frac{n!}{n_1!n_2!.......n_r!}$
VARIACIONES SIN REPETICIÓN	▪ Importa el orden. ▪ No pueden repetirse los elementos.	$V_n^m=\frac{n!}{(n-m)!}$
VARIACIONES CON REPETICIÓN	▪ Importa el orden. ▪ Se pueden repetir los elementos.	$VR_n^m=n^m$
COMBINACIONES SIN REPETICIÓN	▪ No importa el orden. ▪ No pueden repetirse los elementos.	$C_n^m=\binom{n}{m}=\frac{V_n^m}{P_m}=\frac{n!}{(n-m)!m!}$
COMBINACIONES CON REPETICIÓN	▪ No importa el orden. ▪ Sí pueden repetirse los elementos.	$CR_n^m=\binom{n+m-1}{m}=\frac{(n+m-1)!}{m!(n-1)!}$

RECUERDE QUE:

Si los grupos se obtienen de dos o más conjuntos, se aplica la **REGLA DE LA MULTIPLICACIÓN.**

Si los grupos se obtienen de un conjunto e interesa el orden, se aplicas la **REGLA DE LA PERMUTACION o VARIACIÓN.**

Si los grupos se obtienen de un conjunto y no interesa el orden, se aplica la **REGLA DE LA COMBINACIÓN.**

Hay ocasiones que se combina la regla de multiplicar con la permutación o con la combinación.

SITUACIONES PROBLEMÁTICAS RESUELTAS

1) Se tienen 3 semáforos, 2 en esquinas contiguas y otro en medio de ellos. Este último tiene rota la luz amarilla. ¿Cuántas señales de luces diferentes pueden realizarse, entre los 3 semáforos? Y ¿Cuáles son?

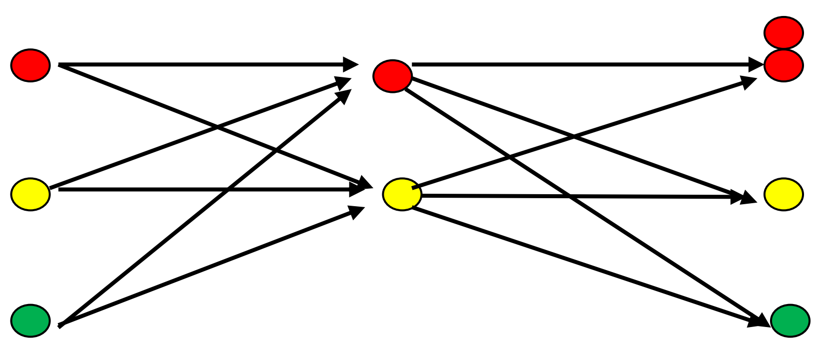
Solución

Con el primer semáforo 3 señales, pero por cada una de estas, en el segundo semáforo, dos señales, finalmente, por cada una de estas 2, con el 3º, podremos realizar 3 señales adicionales, por tanto, por el principio multiplicativo en total se tendrán: $3x2x3 = 18$ señales.

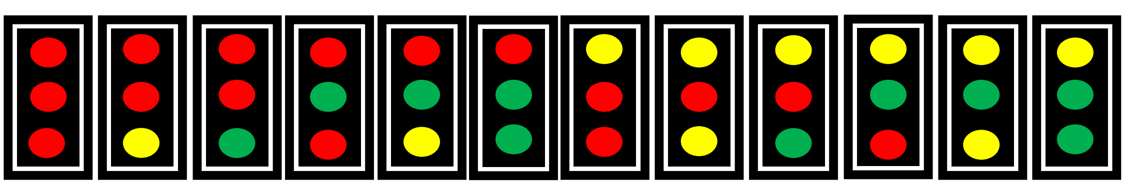
Sean: : Luz roja : Luz amarilla : Luz verde

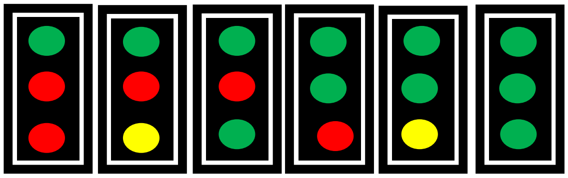
Esquema para obtener señales de luces diferentes que se pueden realizar.

Semáforo 1 Semáforo 2 Semáforo 3



Las señales de luces diferentes que se pueden generar con los tres semáforos son:





2) Se desean crear banderas tricolor de tres bandas horizontales y se dispone de seis rollos de tela con los siguientes colores:

Rojo	Amarillo	Marrón	Negro	Blanco	Verde

Solución

La bandera "RAV" es válida.

La bandera "RAN" es válida.

La bandera "RNA" es válida.

La bandera "NBA" es válida.

La bandera "NBN" no es válida, puesto que no es tricolor.

Las banderas presentadas han sido denominadas: **RAN, RAV, RNA y NBN**. Esta denominación indica cómo se situación de los colores, puesto que están en fila dentro de cada banderas. Se ha determinado que el color superior se corresponde con la primera letra y el más abajo la última.

Con la simple escritura de la inicial de cada color, se puede analizar el problema, comprobando que el cambio de posición de una letra, modifica la bandera, y que no se puede repetir una misma letra.

En este caso, resulta que en la agrupación de colores, es importante la posición de cada color en la bandera, se dispone de seis colores, y sólo se eligen tres para cada agrupación; entonces podemos decir, que cada bandera es una variación sin repetición, de seis colores tomados de tres en tres. Por lo tanto, se utilizará la fórmula siguiente: $V_n^m = \frac{n!}{(n-m)!}$, con n=6 total de colores y m=3 número de colores que se requiere para cada bandera.

$V_6^2 = \frac{6!}{(6-3)!} = \frac{6!}{3!} = \frac{6x5x4x3!}{3!} = 6x5x4 = 120$. Así pues, hay 120 banderas tricolores distintas.

3) Se tiene una familia con cinco miembros: padre, madre, hijo mayor, hijo mediano e hijo menor.

a) Los miembros de la familia se sitúan en una misma fila de un cine. Si cada fila, tiene 25 butacas: ¿De cuántas formas distintas, puede estar la familia sentada en el cine?
b) Al salir del cine, la familia se sienta en una mesa redonda, para cenar: ¿De cuántas formas distintas, puede estar la familia sentada en la cena?
c) Al volver a casa, cada uno se va a una de las cinco camas de la casa para dormir: ¿De cuántas formas distintas, pueden ocupar las camas para dormir?.
d) Al despertar, y desayunar, cada miembro de la familia ingiere un único líquido de los disponibles: té, café, leche y zumo: ¿Cuántos desayunos distintos se pueden producir?

Solución a)

Cómo cada fila tiene 25 butacas, y hay que elegir una, para cada miembro de la familia; entonces hay tantas colocaciones como variaciones de 25 elementos, tomados de cinco en cinco: Entonces n=25 el total de butacas y m=5 los miembros del grupo familiar.

$V_{25}^{5} = \dfrac{25!}{(25\text{-}5)!} = \dfrac{25!}{20!} = \dfrac{25x24x23x22x21x20!}{20!} = 25x24x23x22x21 = 6_{1}375{,}600$ Formas distintas.

Solución b)

Cómo cada mesa redonda tiene 5 asientos, y hay que elegir uno, para cada componente de la familia; entonces hay tantas colocaciones como permutaciones circulares de 5 elementos: entonces n=5 el total de miembros de la familia y cinco también son los asientos que tiene la mesa redonda.

$PC_5 = (5\text{-}1)! = 4! = 4x3x2x1 = 24$ Formas distintas.

Solución c)

Cómo la casa tiene 5 camas, y hay que elegir una, para cada componente de la familia; entonces hay tantas colocaciones como permutaciones circulares de 5 elementos.

$P_5 = 5! = 5x4x3x2x1 = 120$ Formas distintas.

Solución d)

Cómo hay cuatro bebidas disponibles, y hay que elegir una, para cada componente de la familia; entonces hay tantos desayunos como variaciones con repetición de 4 elementos, tomados de cinco en cinco. Entonces n=4 cantidad de bebidas disponibles y m=5 total de miembros de la familia.

$VR_4^5 = 4^5 = 1020$ Desayunos distintos.

4) En una empresa hay 5 plazas vacantes, de las que 3 corresponden a hombres y 2 a mujeres. Se han presentado 10 hombres y 8 mujeres.

   a) ¿De cuántas formas distintas se pueden cubrir las vacantes?

   b) ¿Cuántas posibilidades habrá si las plazas de los hombres tienen todas distintas remuneraciones?

c) ¿Cuántas posibilidades habrá si tanto las plazas de los hombres como las de las mujeres tienen distinta remuneración?

d) ¿De cuántas maneras se pueden ordenar en fila los 10 hombres y las 8 mujeres?

e) ¿De cuántas maneras se pueden ordenar en fila los 10 hombres y las 8 mujeres si los hombres deben estar juntos y las mujeres también?

Solución

Número de plazas vacantes: 5 de las cuales 3 son para hombres y 2 para mujeres

Número de persona que se han presentado: 18 de las cuales 10 son hombres y 8 son mujeres.

Solución a) ¿De cuántas formas distintas se pueden cubrir las vacantes?

Como no existe restricción alguna, es decir el orden en que se seleccionen los participantes para asignarse a cada plaza no interesa y en este caso por naturaleza la misma persona no puede pertenecer ala mismo grupo de seleccionados. Por lo tanto, son combinaciones sin repetición.

El número de formas distintas de elegir a los hombres es:

$$C_{10}^{3}=\binom{10}{3}=\frac{10!}{(10\text{-}3)!3!}=\frac{10!}{7!3!}=\frac{10x9x8x7!}{7!3!}=\frac{10x9x8}{3!}=\frac{10x9x8}{3x2x1}=\frac{720}{6}=120$$

El número de formas distintas de elegir a las mujeres es:

$$C_{8}^{2}=\binom{8}{2}=\frac{8!}{(8\text{-}2)!2!}=\frac{8!}{6!2!}=\frac{8x7x6!}{6!2!}=\frac{8x7}{2!}=\frac{8x7}{2x1}=\frac{56}{2}=28$$

Utilizando el principio de la multiplicación el número de formas distintas de cubrir las plazas es:

$C_{10}^{3}xC_{8}^{2}$ *=120x28= 3360* . Por lo tanto, se tienen 3360 formas diferentes de cubrir las plazas vacantes.

Solución b) ¿Cuántas posibilidades habrá si las plazas de los hombres tienen todas distinta remuneración?

Como existen el criterio que las plazas de los hombres tienen distinta remuneración, significa que el orden interesa en la selección de los hombres ya que cada plaza tiene diferente salario, y para la selección de las mujeres no hay ningún criterio; por lo tanto, para seleccionar las plazas para los hombres se utiliza variaciones sin repetición y para seleccionar las plazas de las mujeres se utilizan combinaciones sin repetición.

El número de formas distintas de elegir a los hombres es:

$$V_{10}^{3}=\frac{10!}{(10\text{-}3)!}=\frac{10!}{7!}=\frac{10x9x8x7!}{7!}=10x9x8=720$$

El número de formas distintas de elegir a las mujeres es:

$$C_8^2 = \binom{8}{2} = \frac{8!}{(8\text{-}2)!2!} = \frac{8!}{6!2!} = \frac{8x7x6!}{6!2!} = \frac{8x7}{2!} = \frac{8x7}{2x1} = \frac{56}{2} = 28$$

Utilizando el principio de la multiplicación el número de formas distintas de cubrir las plazas es:

$V_{10}^3 x C_8^2$ *=720x28= 20160*

Por lo tanto, se tienen 20160 formas diferentes de cubrir las plazas vacantes considerando que las plazas de los hombres tienen diferente remuneración.

Solución c) ¿Cuántas posibilidades habrá si tanto las plazas de los hombres como las de las mujeres tienen distinta remuneración?

Aplicando el análisis del literal anterior del criterio para los hombres. Es decir, utilizando variaciones sin repetición para ambos casos.

El número de formas distintas de elegir a los hombres es:

$$V_{10}^3 = \frac{10!}{(10\text{-}3)!} = \frac{10!}{7!} = \frac{10x9x8x7!}{7!} = 10x9x8 = 720$$

El número de formas distintas de elegir a las mujeres es:

$$V_8^2 = \frac{8!}{(8\text{-}2)!} = \frac{8!}{6!} = \frac{8x7x6!}{6!} = 8x7 = 56$$

Utilizando el principio de la multiplicación el número de formas distintas de cubrir las plazas es:

$V_{10}^3 x V_8^2$ *=720x56= 40,320*

Por lo tanto, se tienen 40,320 formas diferentes de cubrir las plazas vacantes considerando que las plazas de los hombres y de las mujeres tienen diferente remuneración.

Solución d) ¿De cuántas maneras se pueden ordenar en fila los 10 hombres y las 8 mujeres?

Como en esta ordenación se incluyen tanto los hombres como las mujeres son permutaciones sin repetición.

P_{18}=18!= 18x17x16x……….x1=6402373705000000

El número de maneras en que se pueden ordenar los 10 hombres y las 8 mujeres en filas es 6402373705000000

Solución e) ¿De cuántas maneras se pueden ordenar en fila los 10 hombres y las 8 mujeres si los hombres deben estar juntos y las mujeres también?

Ya que dan el criterio que el grupo de hombres y el grupo de mujeres deben están juntos, entonces se deben ordenar por separado cada grupo. Y para ello se utiliza permutaciones sin repetición.

Los hombres se pueden ordenar de: P_{10}= 10 x 9 x 8 x 7 x 6 x 5x 4x 3x 2 x 1=3628800

Las mujeres se pueden ordenar de P_8= 8 x 7 x 6 x 5 x 4 x 3 x 2 x 1=40320

Utilizando el principio de la multiplicación el número de ordenaciones posibles si los hombres deben estar juntos y las mujeres también es: 2 x P_{10} x P_8=2(3628800)(40320)= 14631321

Se considera el doble del producto de las formas en que se pueden ordenar los hombres y las mujeres porque los hombres pueden estar situados delante de las mujeres o detrás.

APLICANDO LO APRENDIDO

1. Sofía y Camila participan en un torneo de tenis. La primera persona que gane dos juegos seguidos o que complete tres, gana el torneo. Use un diagrama de árbol para determinar los posibles resultados del torneo.

2. Una familia desea adquirir una vivienda en un balneario y se le presentan las siguientes posibilidades: casa o apartamento. A su vez, cada una puede ser de 1, 2 o 3 dormitorios. ¿Cuántos tipos posibles de vivienda tiene a disposición?

3. Un profesor de Matemática Discreta cuenta cinco chistes cada mes. ¿Cuántos chistes diferentes debe conocer el profesor para que en un período de 4 años no repita el mismo conjunto de 5 chistes? R/ Debe conocer al menos 8 chistes diferentes.

4. Con las letras de la palabra PATATA: ¿Cuántas palabras distintas pueden escribirse? Palabras que pudieran ser legibles, o no, con significado, o no. R/ 60

5. Se desea crear distintas pinturas, mezclando cantidades iguales de tres de los colores disponibles, en seis botes con las siguientes pinturas:

Rojo	Amarillo	Marrón	Naranja	Blanco	Verde

¿Cuántas pinturas diferentes se pueden formar?

6. Una ferretería, dispone de cinco cajones, y cada cajón, contiene llaves de un mismo color; así pues, se tienen llaves de cinco colores distintos.

Rojo	Amarillo	Marrón	Naranja	Blanco	Verde

Al ser solicitadas cuatro copias de la llave un cliente, estas se fabrican con las seleccionadas de esos cajones. Se trata de estudiar los distintos juegos de llaves que podrían ser fabricadas.

R/ 70 juegos de llaves distintas.

7. ¿Cuántos matrimonios diferentes se pueden efectuar entre 3 hombres y 7 mujeres?.R/21 matrimonios. Y cuáles son?.

8. El testigo de un accidente reporta que la placa del automóvil que huyó era un número de 6 cifras. Recuerda las tres primeras cifras pero ha olvidado las otras 3. ¿Cuántas placas tiene que investigar la policía? R/ 1000 placas.

9. Doce ingenieros del departamento de instrumentación de una fábrica tienen que distribuirse en grupos de 4 hombres para el estudio de proyectos. ¿Cuántos grupos distintos es posible formar? R/495 grupos.

10. La Sra. Pérez tiene cinco sombreros, nueve vestidos, tres bolsos y seis pares de zapatos. ¿De cuántas maneras diferentes puede salir vestida de su casa?.R/810 maneras que puede vestirse.

11. El jefe de cocina de un restaurante quiere usar algunas carnes y vegetales que sobraron el día anterior para preparar un platillo de tres clases de carne y cuatro vegetales. Si hay 5 clases de carne y siete vegetales disponibles, ¿Cuántos platillos puede preparar el cocinero? R/ 350 platos distintos.

12. Una señora tiene 11 amigas de confianza.

    a) ¿De cuántas maneras puede invitar a 5 de ellas a comer? R/462 formas distintas.
    b) ¿De cuántas maneras si 2 de ellas no se llevan y no asisten juntas, es decir si una va la otra no va. R/ 126+256=378

13. Un grupo de investigadores está compuesto por 4 Biólogos 5 Químicos y 3 Médicos. Un experimento que llevarán a cabo requiere de 2 Biólogos, 1 Químico y 2 Médicos.
    a) Calcule de cuántas maneras distintas puede encargarse el experimento al personal disponible. R/ 90 maneras diferentes.
    b) Calcule cuántos, si entre los biólogos hay un jefe que necesariamente participa en el experimento. R/45 maneras diferentes.

14. De los 8 hombres de la tripulación de una barca, dos de ellos solo pueden remar por el lado izquierdo y tres sólo por el lado derecho. ¿De cuántas maneras diferentes se puede colocar la tripulación? R/1728 maneras diferentes.

15. Con los dígitos 1, 2, 3 y 4 forma todos los números de tres cifras que puedas sin que se repita ninguna. ¿Cuáles son? y ¿Cuántos son? R/24 números.

PRUEBA OBJETIVA

1. Un alumno tiene que elegir 7 de las 10 preguntas de un examen.
   a) ¿De cuántas maneras puede elegirlas? R/ 120 maneras.
   b) ¿Y si las 4 primeras son obligatorias? R/ 20 maneras.

2. Cuatro libros de matemática, seis de físicas y dos de química han de ser colocados en una estantería. ¿Cuántas colocaciones distintas admiten si:
   a) Los libro de cada materia han de estar juntos? R/207360 colocaciones distintas.
   b) Solo los de matemática tienen que estar juntos? R/ 8709120 formas de colocar los libros.

3. En una clase de 10 alumnos van a distribuirse 3 premiso. Averiguar de cuantos modos puede hacerse si:
   a) Los premios son diferentes. R/
   b) Los premios son iguales.

Nota: Para los literales anteriores considere los siguientes casos: que no se pueden recibir más de un premio y que se puede recibir más de un premio. **Respuestas**

No se puede repetir

a) 720 maneras de distribuir los premios.
b) 120 maneras de distribuir los premios

Se pueden repetir los premios

a) 1000 maneras de distribuir los premios
b) 220 maneras de distribuir los premios

4. Una pastelería elabora galletas de tres sabores: sencillas, cubiertas de chocolate y rellenas de mermelada, y las envasa en cajas de 100, 200 y 400 gramos. Forma un diagrama en árbol. ¿Cuántos productos diferentes se pueden escoger? R/ 9 tipos de paquete de galletas diferentes.

5. Un niño tiene nueve canicas; tres rojas, tres verdes y tres amarillas, en un bolsillo. Al meter la mano en el bolsillo, saca tres canicas: ¿De cuántas formas distintas pueden distribuirse los colores de las tres canicas? R10

6. En la nevería se pueden escoger entre 4 ingredientes diferentes para elaborar un helado preferido. Los 4 ingredientes se colocan en una bandeja giratoria. ¿En cuántas maneras se pueden disponer? R/ 6 maneras.

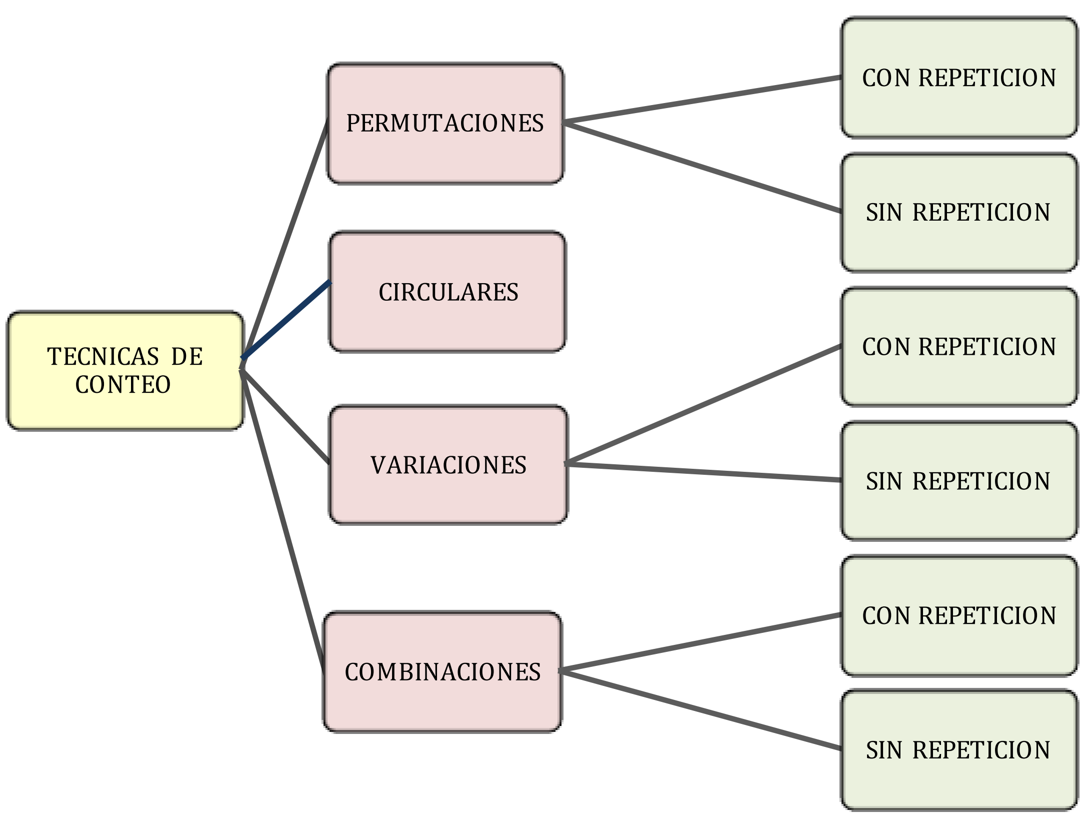
7. Indica si son verdaderas o falsas las siguientes afirmaciones.

   a) En las variaciones con repetición no importa el orden. R/ Falso
   b) En las variaciones sin repetición sí importa el orden. R/ Verdadero
   c) En las permutaciones importa el orden. R/Verdadero
   d) En las combinaciones importa el orden. R/Falso

BIBLIOGRAFIA

1. Johnson R., Cuby P.(1999), *Estadística Elemental*. México:Internacional Thomson Editores, S.A de S.V.
2. Pérez C., (2003).*Estadística. Problemas Resueltos y Aplicaciones*. Madrid: Pearson Educación, S.A.
3. Sarabia J.M. (2000), *Curso Práctico de Estadística*.(2da ed.) . España: Impreso por Gráficas Rogar, S.A Navalcarnero (Madrid).
4. Triola, M.,(2009). *Estadística.* (10a ed.). México: Pearson Educación.
5. Pérez-T. H.E((2007), Estadística para las Ciencias Sociales, del Comportamiento y de la Salud. (3ª. Ed).México: Impreso Edamsa Impresiones, S.A. de C.V.

MAPA CONCEPTUAL



Lección 8 y 9 **Segundo año de Bachillerato** **Unidad IV** **Tiempo: 16 horas clase**

Conceptos de probabilidad

Figura 1. Experimentos aleatorios.

Introducción

El descubrimiento de leyes que rigen y explican los fenómenos que dependen del azar, como los juegos de dados o de cartas, son características que hacen de la estadística y la probabilidad áreas sumamente especiales de la ciencia matemática.

La probabilidad es la rama de la matemática que estudia los fenómenos con incertidumbre. Cuando en un experimento no se pueden realizar las simplificaciones necesarias para conocer con precisión el resultado, se debe recurrir a los modelos aleatorios. Por ejemplo, el número de llamadas telefónicas que debe enlazar una central en cierta hora del día, el número de personas que llegan a una fila bancaria para ser atendidas, la cantidad de lluvia en un lugar determinado en un mes, etc. Y para poder estudiar estos y otros problemas que se presentan es necesario estudiar y comprender la probabilidad.

La probabilidad mide la incertidumbre de un suceso, un número que expresa nuestra creencia en la ocurrencia de un evento incierto.

Objetivos

- Relacionar los conceptos de probabilidad teórica y frecuencia relativa.
- Aplicar distintas técnicas de conteo, distinguiendo las adecuadas para la resolución de cada problema, utilizando la definición clásica de probabilidad.
- Conocer, comprender y manejar los conceptos relacionados con eventos y probabilidad.
- Identifique los procedimientos analíticos como herramientas que le permitan comprender, interpretar y explicar las reglas de la suma y el producto de la probabilidad, así como la probabilidad condicionada.

Importancia

Una disputa entre apostadores en año de 1654 llevó a la creación de la teoría matemática de probabilidad por dos matemáticos franceses: Blaise Pascal y Pierre de Fermat. Así que la probabilidad se estudió, inicialmente, para resolver problemas relacionados con juegos de azar. Actualmente, esta se utiliza en una gran variedad de campos y por tal razón se sigue investigando la misma a través de todo el mundo.

Uno de los usos más frecuentes de probabilidad es en el pronóstico del tiempo. En muchos países, gracias a la probabilidad se determina cuando comienza y termina la temporada de huracanes, y para crear trayectorias posibles de huracanes, tormentas, ondas tropicales, etc

Otro de los usos de la probabilidad es en los juegos de azar, especialmente en los casinos. Las personas que son exitosas en estos tipos de juegos no lo son porque tengan "suerte", sino porque tienen conocimiento de probabilidad.

La economía mundial se rige en gran parte por la probabilidad. La crisis económica actual en los Estados Unidos, se debe en cierto modo a la interpretación que se le ha dado a los modelos probabilístico por parte de los especuladores del sector de los bienes raíces.

En términos generales, la probabilidad nos ayuda en la vida diaria a ser más cautelosos al momento de tomar decisiones. Si se maneja a una velocidad moderada es menos probable que se tenga un accidente y si se tiene un accidente es menos probable que se sufra heridas graves si se utiliza el cinturón de seguridad.

Competencias a reforzar.

Calcula, analiza e interpreta el conocimiento de la probabilidad en la solución de problemas cotidianos, y la utiliza para tomar decisiones.

Presaberes

- Operaciones aritméticas.
- Teoría de conjuntos.
- Cardinalidad de conjuntos.
- Operaciones de conjuntos.
- Relaciones de conjuntos.
- Representación de conjuntos en diagrama de Venn.
- Diagrama de árbol.

LA PROBABILIDAD Y SUS INICIOS

En el siglo XVII, Blaise Pascal y Pierre Fermat, concibieron los principios del Cálculo de Probabilidades a partir de los problemas propuestos por el caballero De Meré, noble francés del siglo XVII aficionado a apostar dinero en juegos de cartas; es así, como se puede afirmar que, desde sus inicios, la teoría de la probabilidad ha estado ligada a los juegos de azar.

A raíz de los trabajos de Pascal y Fermat, la teoría de la probabilidad se popularizó entre la comunidad matemática, por esta razón se promovió su desarrollo. Antes de Pascal esta teoría solo estudiaba la aleatoriedad de los fenómenos, pero fue precisamente Pascal quien comenzó a introducirla en áreas del conocimiento, como la genética, la sicología y la economía, entre otras.

Apuestas ventajosas

Antoine Gombaud, Caballero de Meré fue un filósofo francés, aficionado a las matemáticas y experto jugador, que se interesó particularmente en el análisis riguroso del juego de dados, movido por sus inesperadas pérdidas. Gombaud recurrió a Pascal para que le explicara la razón, pues él sabía que era ventajoso apostar por obtener al menos un seis, en una serie de *4* lanzamientos de un dado, donde efectivamente ganaba; él supuso, que debía ser igualmente ventajoso apostar por obtener al menos un doble seis en una serie de *24* lanzamientos de un par de dados, pero con ello normalmente perdía. Supuso proporcionalidad y utilizó una regla de tres simple: *4* es a *6* igual que *24* a *36*.

No se conoce la solución que dio Pascal al problema; se sabe que lo resolvió porque así se lo hizo saber a Fermat en una carta, invitándolo a descubrirla fácilmente, dados los principios que tenía.

En cada lanzamiento de un dado hay *6* posibles resultados; en una serie de *4* lanzamientos, los resultados posibles son:

$$VR_6^4 = 6^4 = 1296$$

Para calcular el número de resultados que contienen al menos un *6*, conviene hacerlo por complemento, es decir, calculando primero los resultados con valores del *1* al *5,* cuatro veces seguidas: $VR_5^4 = 5^4 = 625$

Por lo tanto, el número de resultados que contienen al menos un seis es: 1296 – 625 = 671. Expresando esto en forma de proporción: *671 : 625,* se distingue claramente la pequeña ventaja que tiene el que apuesta por al menos un seis en *4* lanzamientos del dado. Expresado como probabilidad: *0.5177*

Del mismo modo, en cada lanzamiento de un par de dados hay *36* posibles resultados; en una serie de *24* lanzamientos, los resultados posibles son: $VR_{36}^{24} = 36^{24}$

Para calcular el número de resultados que contienen al menos un doble seis, conviene hacerlo por complemento, es decir, calculando primero los resultados que no son un doble seis, veinticuatro veces seguidas: $VR_{35}^{24} = 35^{24}$

Por lo tanto, el número de resultados que contienen al menos un seis es: $36^{24} - 35^{24}$. Expresando esto en forma de proporción: $\left(36^{24} - 35^{24}\right) : 35^{24}$*,* se tienen que hacer operaciones para darse cuenta de la pequeña ventaja que tiene el que apuesta por al menos un doble seis en *24* lanzamientos de dos dados. Expresado como probabilidad: *0.4945.*

Al observar que la desventaja era tan pequeña, se hace difícil de creer que, efectivamente Gombaud la haya podido percibir empíricamente. Se sabe que el problema ya llevaba bastante tiempo circulando entre los estudiosos de la época; otra posibilidad es que, habiendo llegado Gombaud a ese resultado, por si mismo, le surgieron dudas que quiso disipar con Pascal.

Es fácil darse cuenta que con *25* lanzamientos de un par de dados, en vez de 24, la desventaja se convierte en ventaja:

$$\left(36^{25} - 35^{25}\right) : 35^{25}$$

con una probabilidad equivalente de *0.5055.*

En condiciones de equidad, el problema podría ser planteado como la determinación del número de lanzamientos que garantizan el equilibrio, que ocurre cuando las probabilidades de ganar y perder coinciden.

En la vida cotidiana aparecen muchas situaciones en las que los resultados observados son diferentes aunque las condiciones iniciales en las que se produce la experiencia sean las mismas. Estos fenómenos, denominados aleatorios, se ven afectados por la incertidumbre. En el lenguaje habitual, frases como "probablemente...", "es poco probable que...", "hay muchas posibilidades de que..." hacen referencia a esta incertidumbre.

La probabilidad es la parte de la matemática encargada de estudiar fenómenos relacionados con el Azar; y ha resultado ser una aplicación de las matemáticas que ha tenido un uso más potente en la sociedad. Se podría llegar a definir como "la forma rigurosa de adivinar el futuro". Es por ello que la mayoría de toma de decisiones en empresas y estados haga uso de estadística y probabilidad.

La probabilidad es la mayor o menor posibilidad de que ocurra un determinado suceso. En otras palabras, su noción viene de la necesidad de medir o determinar cuantitativamente la certeza o duda de que un suceso dado ocurra o no.

La **teoría de la probabilidad** pretende ser una herramienta para modelizar y tratar con situaciones de este tipo. Por otra parte, cuando aplicamos las **técnicas** estadísticas a la recogida, análisis e interpretación de los datos, la teoría de la probabilidad proporciona una base para evaluar la fiabilidad de las conclusiones alcanzadas y las inferencias realizadas.

DEFINICIONES BASICAS

Según la naturaleza de los experimentos se pueden distinguir entre Experimentos Determinísticos y Experimentos Aleatorios.

Los **Experimentos Determinísticos** son un conjunto de pruebas realizadas bajo las mismas condiciones que producen los mismos resultados; es decir, son previsibles, no hay incertidumbre acerca del resultado que ocurrirá cuando éstos son repetidos varias veces.

En este tipo de experimentos cuando se conoce las condiciones iniciales del experimento el resultado final está fijado, y desde el principio se puede conocer. Y son experimentos de índole físico, si se tira una piedra se sabe que esta caerá y se podrá incluso conocer el tiempo que está en el aire, o químicos si se ponen dos sustancias a reaccionar se sabe si estas interactúan y en qué medida.

Ejemplo 1: En una molécula de agua (H_2O), la proporción de las masas del hidrógeno y oxígeno están en una razón 2:1, como se puede comprobar al disociar las moléculas de un volumen conocido de agua. Del mismo modo, se pueden mezclar estos dos gases con volúmenes en la razón 2:1, y la mezcla de agua resultante tendrá la masa que es exactamente la suma de 2(masa del H) + 1(masa de O). Así, estamos en presencia de un fenómeno determinista de la química: Sabiendo las masas de cada uno de los elementos es absolutamente predecible la masa del agua a obtener al mezclarlos.

Ejemplo 2: Si dejamos caer una piedra desde una ventana sabemos, sin lugar a dudas, que la pelota bajará. Si la arrojamos hacia arriba, sabemos que subirá durante un determinado intervalo de tiempo; pero después bajará.

Los **Experimentos aleatorios** son el conjunto de pruebas realizadas bajo las mismas condiciones y cuyos resultados son impredecibles; es decir, no se puede anticipar o predecir exactamente qué resultado ocurrirá en el siguiente intento o cuando se realiza.

Los experimentos aleatorios deben cumplir o verifican las siguientes condiciones:

- No es posible predecir con certeza el resultado que se obtendrá, pero si describir completamente los resultados que se pueden presentarse.
- El experimento puede realizarse en las mismas condiciones todas las veces que nos sea posible y siempre se obtendrá idéntico conjunto de resultados.
- Al repetir el experimento, los resultados aparecen en forma predecible, pero formando un patrón (regularidad estadística).

En un Experimento aleatorio si bien es cierto que se conocen todos los resultados posibles, pero no se puede decir con seguridad cuál de ellos ocurrirá en un caso particular.

La estadística y más concretamente la probabilidad se encarga de este tipo de fenómenos, intentado dar una medida de la incertidumbre respecto a la ocurrencia o no, de estos.

Ejemplo 3: Considérese los siguientes experimentos aleatorios.

1) **Experimento 1:** Lanzar al aire una moneda legal (moneda elaborada con material cuya masa se encuentra uniformemente distribuida en el volumen de la moneda) para que caiga sobre una superficie lisa.

Observe que el experimento puede realizarse indefinidamente, en las mismas condiciones, y el conjunto de resultados que obtiene siempre es el mismo {cara, águila}.

2) **Experimento 2:** Lanzar dos veces, sobre una mesa, un dado legal (o sea perfectamente cúbico y con densidad constante) cuyas caras se encuentran numeradas del 1 al 6.

El experimento puede llevarse a cabo las veces que se quiera, en las mismas condiciones, y los resultados serán: {(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6), (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6), (4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6), (5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6), (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)}, donde los números que aparecen como primera componente son el posible resultado que puede obtenerse con el primer lanzamiento y la segunda componente son los resultados posibles para el segundo lanzamiento.

3) **Experimento 3:** Determinar el número de llamadas internacionales que se hacen desde EE. UU, por la empresa TELECOM, entre las 21 horas y 24 horas, los días viernes del mes marzo del año 2012.

El experimento puede realizarse en las mismas condiciones, esperando encontrar como conjunto de resultados posibles el siguiente: {0, 1, 2, 3,….., T}. Esto quiere decir que puede que no llamen, que hagan una llamada, dos llamadas, etc., con el tope de llamadas indicado por la letra T, que depende de la capacidad de los canales de transmisión de la empresa TELECOM.

4) **Experimento 4:** Predecir el estado del tiempo dentro de un año, en El Salvador.

Resulta muy difícil determinar cuál será el estado del tiempo dentro de un año en El Salvador, pero lo que si podemos hacer es describir los posibles estados del tiempo: {seco, húmedo, lluvioso, soleado}

5) **Experimento 5:** Determinar cuántos años va a vivir un recién nacido. Aun cuando el evento morir sea seguro, es imposible determinar cuándo ocurrirá, por lo que no resulta fácil predecir con certeza el tiempo de vida de un recién nacido. No obstante, pueden hacerse pronostico utilizando tablas de esperanza de vida de la población a la cual pertenece este nuevo ser. En este caso el conjunto de todos los resultados posibles puede escribirse en la forma:

   $\{x \in R / 0 \leq x \leq l\}$, donde l es la edad máxima contemplada en la tabla, y x indica las diferentes edades que un individuo de esta población puede vivir.

6) **Otros ejemplos:**
   - El equipo que ganará el partido de fútbol en el recreo.
   - Número premiado de la lotería.
   - Número de piezas defectuosas en la producción de una fábrica durante un día.
   - El número de accidentes de tráfico durante un fin de semana
   - El valor de la temperatura que hará el 10 de Mayo de 2012.
   - El número de la cara del dado que caerá hacia arriba al lanzarlo al aire.
   - Tiempo que hay que esperar para ser atendidos en un Banco.

ACTIVIDAD INTRODUCTORIA: CARRERA DE CONEJOS

Cada jugador pone una ficha en la salida en el número que más le guste, sin poder elegir un número que ya está elegido. Por turno se van tirando dos dados y se suman los valores que resultan en la cara superior; y si el que ha tirado tiene ese número moverá su ficha, de lo contrario no moverá y tirará el jugador siguiente. El conejo que llegue primero a la meta ganará. (No hay número del grupo definido)

 7											
 8											
 9											
 10											
 11											
 12											

En grupo Reflexionar sobre:

a) ¿Qué conejo ha ganado?

b) ¿Todos los conejos tienen las mismas posibilidades de llegar a la meta? ¿Por qué?

c) ¿Había algún número con más posibilidades de salir? ¿Por qué?

d) Si tuvieras que jugar de nuevo, ¿Qué número o números elegirías? ¿Por qué?

e) ¿Se puede considerar aleatorio este juego? ¿Por qué?

ESPACIO MUESTRAL Y TIPOS DE SUCESO

Dado un experimento aleatorio cualquiera, se define el **espacio muestral** de éste; como el conjunto de todos los posibles resultados que se pueden obtener del experimento. Se representa por E y cada elemento de él es llamado **punto muestral**.

Un **Evento o suceso** es un resultado particular de un experimento aleatorio. En términos de conjuntos, un evento es un subconjunto del espacio muestral. Por lo general se le representa por las primeras letras del alfabeto: A, B, C, …..

Ejemplo 4: Determinar los espacios muéstrales de los experimentos dados anteriormente.

Experimento 1: Lanzamiento de una moneda.

E = {cara, águila}.

Experimento 2: Lanzamiento de dos dados legales.

E = {(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6), (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6), (4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6), (5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6), (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)}

Experimento 3: Número de llamadas internacionales que suministra TELECOM.

E = {0, 1, 2, 3, ….., T}.

Experimento 4: Estado del tiempo en El Salvador Espacio Muestral: E = {seco, húmedo, lluvioso, soleado}

Experimento 5: Tiempo de vida de un recién nacido. Espacio muestral: $E = \{x \in R / 0 \leq x \leq l\}$

Ejemplo 5: De los experimentos del 1 al 3 planteados anteriormente, obtener dos sucesos de cada uno de ellos.

Sucesos del Experimento 1

A: Obtener cara en el lanzamiento de la moneda entonces A = {cara}

B: Obtener cara o águila en el lanzamiento de la moneda entonces B = {cara, águila}

Sucesos del Experimento 2

A: En el primer lanzamiento se obtuvo el número 3 entonces

A = {(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6)}

B: En el segundo lanzamiento se obtuvo un número par entonces B = {(1, 2),(1, 4), (1, 6), (2, 2), (2, 4), (2, 6), (3, 2), (3, 4), (3, 6), (4, 2), (4, 4), (4, 6), (5 2), (5, 4), (5, 6), (6, 2), (6, 4), (6, 6)}

Sucesos del Experimento 3

F: El número de llamadas internacionales fue a lo sumo 3 F = {1, 2, 3}

G: El número de llamadas internacionales fue de 50 a 100 G = {50, 51, 52,….., 100}

Los eventos formados por un único elemento (conjuntos unitarios) se llaman **eventos simples**; y si están formados por más de un elemento, se les llama **eventos compuestos.**

Si E es el espacio muestral asociado a un experimento, entonces E y el conjunto vacío (ϕ) son eventos del experimento. Al espacio muestral E se le llama **evento seguro** y ϕ se llama **evento imposible.**

- **Evento Seguro (E):** Es aquel que se verifica siempre, sea cual sea el resultado del experimento, por lo tanto estará formado por todos los resultados posibles. El suceso seguro coincide con el espacio muestral.
- **Evento Imposible (ϕ):** Es el que no se puede obtener como resultado del experimento aleatorio, el que no ocurre nunca y no tiene ningún elemento.

Ejemplo 6: El experimento consiste en el lanzamiento de un dado. Escriba un evento imposible y el evento seguro.

Evento seguro: E = {1, 2, 3, 4, 5, 6}

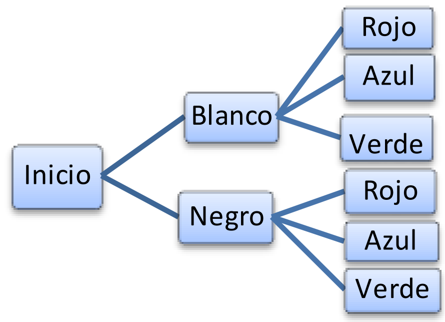
Evento imposible: B: Obtener un número mayor que 6 en el experimento, entonces B={ϕ}. Lo que significa que el evento B nunca ocurrirá pues el dado solo tiene números del 1 al 6.

Eventos igualmente probables: Todos tienen la misma probabilidad de ocurrir (equiprobables).

Eventos dependientes: Aquellos en que la ocurrencia de uno afecta la probabilidad de ocurrencia de los demás.

Eventos independientes: La ocurrencia de uno no afecta la probabilidad de ocurrencia o no de los demás.

Ejemplo 7: En un equipo de fútbol-sala disponen para jugar de pantalones blancos o negros, y de camisetas rojas, azules o verdes. ¿De cuántas maneras se pueden vestir para un partido?



E={blanco-Rojo, blanco-Azul, blanco-Verde, Negro-Rojo, Negro-Azul, Negro-Verde}

OPERACIONES Y RELACIONES DE SUCESOS

Una *operación* entre sucesos de un experimento aleatorio es una regla o criterio que nos permite obtener otro suceso del mismo experimento aleatorio. Las dos operaciones más importantes son la ***unión*** **y la** ***intersección***.

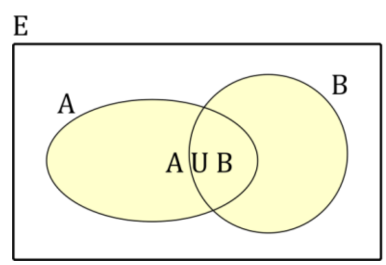
Con los sucesos se opera de manera similar a como se hace en los conjuntos y sus operaciones se definen de manera análoga. Los sucesos a considerar serán los correspondientes a un experimento aleatorio y por tanto serán subconjuntos del espacio muestral *E*.

Por ser los sucesos subconjuntos, sobre ellos se pueden definir las siguientes operaciones. Sean A, B y C sucesos de un Espacio Muestral E, obtenido de un experimento aleatorio.

- Unión de Sucesos

Se define la unión de los sucesos A y B, como el suceso formado por todos los puntos muestrales que pertenecen, al menos, a uno de los sucesos; es decir, contiene todos los elementos que están en A o B. Y se denota por: A U B. De forma matemática se expresa como: A U B={X∈E/X∈A ó X∈B o ambos} y se lee: el suceso A ó B ; y es el suceso de las X tal que X pertenece a A o a B o a ambos.

Representación mediante el diagrama de Venn-Euler.



Es fácil demostrar las siguientes propiedades:

1) Asociativa: (A U B) U C = A U (B U C).

2) Conmutativa: A U B = B U A.

3) Idempotencia: A U A = A

4) Absorbente $A \cup E = E$

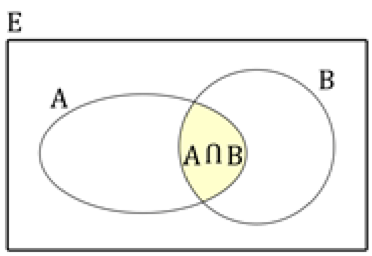
5) Elemento neutro $A \cup \phi = A$

En términos generales, dados *n* sucesos A_1, A_2, A_3,..., A_n, su unión denotada por $\bigcup_{i=1}^{n} A_i$ es otro suceso formado por los resultados o puntos muéstrales que pertenecen al menos a uno de los sucesos A_i.

- Intersección de sucesos

Se define la intersección de los sucesos A y B, como el suceso formado por todos los puntos muestrales que pertenecen a ambos sucesos; es decir, contiene a la vez todos los elementos que están en A y en B. Y se denota por: A ∩ B. De forma matemática se expresa como: $A \cap B = \{X \in E / X \in A \text{ y } X \in B\}$ y se lee como "A y B".

Representación mediante el diagrama de Venn-Euler.



Es fácil demostrar las siguientes propiedades:

1) Asociativa: $(A \cap B) \cap C = A \cap (B \cap C)$.

2) Conmutativa: $A \cap B = B \cap A$

3) Idempotente: $A \cap A = A$.

4) Absorbente $A \cap \phi = \phi$

5) Elemento Neutro $A \cap E = A$

En términos generales, dados n sucesos A_1, A_2, A_3,..., A_n, su intersección denotada por $\bigcap_{i=1}^{n} A_i$; es otro suceso formado por los resultados o puntos muéstrales que pertenecen a todos los sucesos A_i.

Encontrar la intersección de dos conjuntos A y B, significa encontrar los puntos muéstrales comunes de A y B.

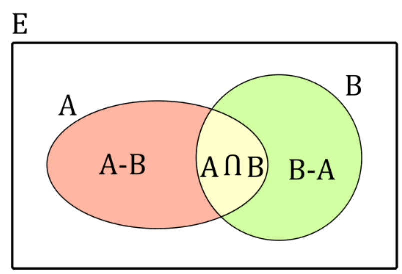
- Diferencia de sucesos

Se define la diferencia de los suceso A y B, como el suceso formado por los puntos muéstrales que pertenecen A y no pertenecen a B. Es decir, es el suceso con puntos muéstrales que están en A pero que no están en B. Y se denota por: A - B. De forma matemática se expresa como: $A - B = \{X \in E / X \in A \text{ y } X \notin B\}$.

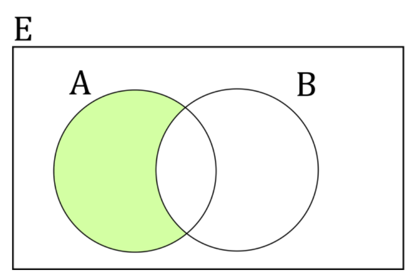
De manera análoga se define la diferencia de los sucesos B y A, como el suceso formado por los puntos muéstrales que pertenecen B y no pertenecen a A. Es decir, es el suceso con puntos muéstrales que están en B pero que no están en A. Y se denota por: B-A. De forma matemática se expresa como:

$B - A = \{X \in E / X \in B \text{ y } X \notin A\}$

Representación mediante el diagrama de Venn-Euler.



$$A - B = A \cap \bar{B}$$

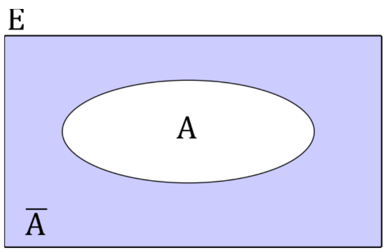


Es fácil de demostrar lo siguiente: $A - B \neq B - A$, lo que indica que la resta aritmética no cumple la propiedad conmutativa.

- Suceso contrario o Complementario.

Se define el complementario del suceso A, con $A \subset E$; como el suceso formado por los puntos muéstrales que no pertenecen a A, es decir contiene todos los elementos de E, que no están en A. Y se denota por: A^c ó $\overline{A}$. De forma matemática se expresa como: $A^c = \{X \in E \,/\, X \notin A\}$.

Representación mediante el diagrama de Venn-Euler.



Es fácil de demostrar las siguientes propiedades:

1) $E - A = A^c$
2) $A \cup A^c = E$
3) $A \cap A^c = \phi$
4) $(A^c)^c = A$

Sean los sucesos A y B del espacio muestral E: Si $A \cap B = \phi$ y $A \cup B = E$, entonces A y B se llaman complementarios. Y se denota por $B = A^c$ (que se lee B es el complemento de A) o $A = B^c$

Propiedades Mixtas

1) Distributiva: $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ y $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

2) Simplificativa: $A \cup (A \cap B) = A$ y $A \cap (B \cup A) = A$.

3) Leyes de De Morgan:

El complemento de la unión de dos sucesos es la intersección de los complementos de dichos sucesos: $\overline{A \cup B} = \overline{A} \cap \overline{B}$

El complemento de la intersección de dos sucesos es la unión de los complementos de dichos sucesos: $\overline{A} \cup \overline{B} = \overline{A \cap B}$

Nota: Todas estas propiedades se pueden generalizarse a más de dos eventos.

A partir de estas operaciones podemos distinguir entre los siguientes tipos de sucesos:

- Suceso Incluido

Se dice que el suceso *A* está incluido en suceso *B* si todos los puntos muestral de *A* pertenecen también a *B*. Es decir, siempre que ocurre el suceso *A*, también ocurre el suceso *B*. Y se denota por $A \subseteq B$, y se lee A implica B. Si $A \subset B$ entonces $X \in A \Rightarrow X \in B$.

- Igualdad de Sucesos

Se dice que el suceso *A* y el suceso *B* son iguales si siempre que ocurre el suceso *A* también ocurre *B* y al revés. Es decir que siempre que se verifica uno de ellos se verifica también el otro. Y denota por A=B. Lo que significa: $A=B \Leftrightarrow A \subset B \wedge B \subset A$.

- Sucesos incompatibles o excluyentes

Se dice que dos sucesos A y B son incompatibles cuando no pueden ocurrir nunca a la vez. Es decir, no tiene ningún punto muestra en común y por tanto $A \cap B = \phi$. Y de lo contrario se denominan sucesos compatibles.

Ejemplo 8: El experimento consiste en el lanzamiento de un dado legal. Sean los eventos:

A: "sale un número par",

B: "sale un número impar",

C: "sale el número 2",

D: "sale un número primo",

E: "sale el número 7".

Determine:

a) El espacio muestral.

b) Los puntos muéstrales de los eventos antes mencionados.

c) Que relaciones se pueden observar de los eventos descritos de este experimento.

Solución a) Como cuando se lanza un dado solo existen seis posibilidades que son: 1, 2, 3, 4, 5 y 6. Por lo tanto el espacio muestral es: E= {1, 2, 3, 4, 5, 6}

Solución b) Los puntos muéstrales de los sucesos son:

$A = \{2, 4, 6\}$ $B = \{1, 3, 5\}$ $C = \{2\}$
$D = \{2, 3, 5\}$ $E = \{\phi\}$

Solución c)

i) $C \subset A$

La relación $C \subset A$: significa que todo elemento de C es elemento de A, por consiguiente, si el evento C ocurre, entonces el evento A ocurre.

ii) $C \subset D$

La relación $C \subset D$: significa que todo elemento de C es elemento de D, por consiguiente, si el evento C ocurre, entonces el evento D ocurre.

iii) $A \cap B = \phi$

La relación $A \cap B = \phi$: significa que los eventos A y B no pueden ocurrir simultáneamente, es decir, que si sale un numero par, entonces no puede salir un número impar.

iv) $A \cup B = E$

La relación $A \cup B = E$: significa que en cualquier ensayo que se realice del experimento, ocurre con seguridad que saldrá un número par o un número impar.

Ejemplo 9: En el primer año de bachillerato con 45 alumnos, seleccionado al azar, se realizo una encuesta sobre los deportes que practican; los resultados son los siguientes:

Deportes	Número de alumnos
Baloncesto	26
Futbol	29
Voleibol	26
Baloncesto y futbol	17
Baloncesto y Voleibol	15
Fútbol y Voleibol	16
Todos	10

a) Represente mediante un diagrama de Venn la información dada.

b) Sea el evento B_1: "Estudiantes que solo practican Baloncesto". Determine los puntos muéstrales.

c) Sea el evento F_1: "Estudiantes que solo practican Fútbol". Determine cuantos puntos muestrales tiene.

d) Sea el evento V_1: "Estudiantes que solo practican Voleibol". Determine cuantos puntos muéstrales tiene.

e) Sea el evento "Estudiantes que practican Baloncesto y Fútbol". Escríbalo de forma simbólica como operaciones de eventos. ¿Cuántos puntos muéstrales tiene?.

f) Sea el evento "Estudiantes que practican Baloncesto y Voleibol". Escríbalo de forma simbólica como operaciones de eventos. ¿Cuántos puntos muéstrales tiene?.

g) Sea el evento $F \cap V$. Escríbalo de forma verbal y el número de puntos muéstrales.

h) Sea el evento $F \cup V$. Escríbalo de forma verbal y el número de puntos muéstrales.

i) Sea el evento N: "Estudiantes que no practican ningún deporte". Determine cuantos puntos muéstrales tiene.

Solución a) Diagrama de Venn

Sean los eventos:

V: Los alumnos que practican Voleibol

F: los alumnos que practican Fútbol

B: Los alumnos que practica Baloncesto

Observe que un alumno que practica tanto Baloncesto como Fútbol se encuentra en dos eventos, de manera que 26+29 cuenta a los alumnos dos veces. Para corregir este error se debe restar 17 (los que practican ambos deportes simultáneamente). Así 26+29-17= 38 alumnos. Es decir $B \cup F = B + F - B \cap F$.

Igualmente un alumno que practica tanto Baloncesto como Voleibol se encuentra en dos eventos, de manera que 26+26 cuenta a los alumnos dos veces. Para corregir este error se debe restar 15(los que practican ambos deportes simultáneamente). Así 26+26-15= 37 alumnos. Es decir $B \cup V = B + V - B \cap V$.

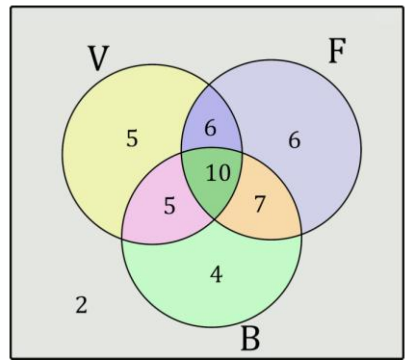
De la misma forma un alumno que practica tanto Fútbol como Voleibol se encuentra en dos eventos, de manera que 26+29 cuenta a los alumnos dos veces. Para corregir este error se debe restar 16 (los que practican ambos deportes simultáneamente). Así 29+26-16= 39 alumnos. Es decir $F \cup V = F + V - F \cap V$.

Como $V \cap F \cap B = 10$ alumnos entonces $B \cap F = 10+7$, $B \cap V = 10+5$ y $F \cap V = 10+6$.

Ahora bien, los alumnos que practican Baloncesto son 26 entonces ya se tiene 10+5+7 = 22 faltan 4 que son los alumnos que solo practican Baloncesto.

En el otro caso, los alumnos que practican Fútbol son 29 entonces ya se tiene 10+6+7= 23 faltan 6 que son los alumnos que solo practican Fútbol.

Por último, los alumnos que practican Voleibol son 26 entonces ya se tiene 10+6+5= 21 faltan 5 que son los alumnos que solo practican Voleibol. Quedando el diagrama de Ven de la siguiente forma:



Teniendo el diagrama de Venn, ya es fácil determinar los puntos muéstrales de cada situación pedida.

Solución b) Evento B_1: "Estudiantes que solo practican Baloncesto". Este evento es la parte del evento B, la cual no está relacionada con ninguno de los otros eventos (color verde musco): 4 alumnos.

Solución c) Evento F_1: "Estudiantes que solo practican Fútbol". Este evento es la parte del evento F, la cual no está relacionada con ninguno de los otros eventos (color celeste): 6 alumnos.

Solución d) Evento V_1: "Estudiantes que solo practican Voleibol". Este evento es la parte del evento V, la cual no está relacionada con ninguno de los otros eventos (color amarillo): 5 alumnos.

Solución e) Sea el evento: "Estudiantes que practican Baloncesto y Fútbol". La forma de representar este evento es: $B \cap F$ y tiene 17 puntos muéstrales, que en la cantidad de estudiantes que practican ambos deportes.

Solución f) Sea el evento "Estudiantes que practican Baloncesto y Voleibol". La forma de representar este evento es: B∩V y tiene 15 puntos muéstrales, que es la cantidad de estudiantes que practican ambos deportes.

Solución g) Evento F∪V, la forma verbal de expresarlo es: "Estudiantes que practican Fútbol y Voleibol" y tiene 16 puntos muéstrales, que es la cantidad de estudiantes que practican ambos deportes.

Solución h) evento F∪V, la forma verbal de expresarlo es: "Estudiantes que practican Fútbol o Voleibol" y tiene 39 puntos muéstrales, que es la cantidad de estudiantes que practican uno u otro deporte.

Solución i) Evento N: "Estudiantes que no practican ningún deporte". Se refiere a los alumnos que no practican Baloncesto ni Fútbol ni Voleibol, en este caso es el evento que no está relacionado con B, F y V o sea el complemento de ellos, y se observa que tiene 2 puntos muéstrales. Es decir 2 alumnos de los 45 no practican ningún deporte de los considerados.

NOTA: Cuando se desea analizar más de tres eventos es recomendable utilizar las tablas de doble entrada o tablas de contingencia.

PROBABILIDAD DE EVENTOS

En nuestro lenguaje cotidiano es común utilizar las expresiones probablemente, probabilidad de, es posible que suceda, es probable que, etc. Estas palabras o frases se usan para marcar la ocurrencia de un fenómeno, evento o experimento.

Por ejemplo, se suele decir "es probable que llueva esta tarde", dando a entender que se tiene mucha confianza o seguridad de que el evento "llueva esta tarde" sí suceda.

Diariamente se escuchan afirmaciones que llevan implícito el concepto de probabilidad como por ejemplo: los pronósticos del tiempo que indican las probabilidades de lluvia; los doctores indican la probabilidad que tiene un enfermo de curarse si realiza al pie de la letra sus tratamientos farmacológicos, los docentes especulan sobre las posibilidades de éxito del estudiantado si dedican más tiempo al estudio, las compañías encuestadoras predicen las oportunidades que tienen los políticos de ganar una elección determinada, etc. La Teoría de la Probabilidad es una rama de las matemáticas que se encarga de los eventos que se realizan al azar o fenómenos aleatorios.

Al llevar a cabo una realización de un experimento aleatorio, se es consientes de que no se puede predecir el resultado, sin embargo se tiene a menudo información sobre las "Posibilidades" que tiene un determinado suceso de ocurrir. Por lo tanto se quiere cuantificar de alguna manera esta información que se llamaría probabilidad del suceso.

Obtener la probabilidad de un suceso, consiste en encontrar el número que nos indica qué tan "probable" es que ese evento ocurra; es decir, mide la creencia que se tiene de que va a ocurrir un suceso especifico.

Se define la probabilidad como un número comprendido entre 0 y 1, que se le asigna a un evento para señalar su posibilidad de ocurrencia. Por lo general las probabilidades se expresan en porcentajes (oscilando 0% - 100%), también se pueden expresar con números decimales (oscilando 0-1).

Se obtiene la probabilidad de 1 si los eventos ocurren con certeza y se obtiene probabilidad de 0 si los eventos no pueden ocurrir; se obtiene una probabilidad de 0.5 si los eventos tienen la misma posibilidad de suceder o de no suceder. Se obtiene una probabilidad ente 0 y 0.5, en los eventos que tenga más posibilidades de no suceder que de suceder; y se obtiene una probabilidad entre 0.5 y 1, en los eventos que tengan más posibilidades de suceder que de no suceder.

La probabilidad es, en realidad, un valor numérico que debe cumplir con ciertas condiciones o propiedades matemáticas y que se asocia a un evento o suceso determinado para expresar el grado de confianza en la verificación futura de dicho evento.

Sea E el espacio muestral asociado a un experimento aleatorio. La probabilidad de cualquier evento A de E, es el número real P(A) que satisface los siguientes axiomas de probabilidad:

1) Para cada suceso A, su probabilidad es un número entre 0 y 1, es decir, $0 \leq P(A) \leq 1$.
2) La probabilidad total es 1, $P(E) = 1$, donde E es el suceso seguro.
3) Si A y B son dos sucesos incompatibles ($A \cap B = \phi$), la probabilidad de su unión es igual a la suma de sus probabilidades; es decir que $P(A \cup B) = P(A) + P(B)$.

CONSECUENCIAS DE LOS AXIOMAS DE PROBABILIDAD

1) La probabilidad del suceso imposible ϕ es cero: $P(\phi)=0$

   Justificación: $E=E\cup\phi$ y $E\cap\phi=\phi$ ya que $P(E)= P(E)+P(\phi)$ entonces $P(\phi)=P(E)- P(E)=0$

2) La probabilidad de el complemento de un suceso es igual a uno menos la probabilidad del suceso: $P(A^c)=1 - P(A)$

   Justificación: $E = A\cup A^c$ y $A\cap A^c=\phi$ ya que $P(E) = P(A)+P(A^c)$ entonces

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

3) Si A y B son sucesos cualquiera no necesariamente excluyentes entonces:

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

   Justificación: $A\cup B=A\cup(B\cap A^c)$ tal que A y $B\cap A^c$ son sucesos excluyentes

   También $B = (A\cap B)\cup(B\cap A^c)$, $A\cap B$ y $B\cap A^c$ son sucesos excluyentes

   Por lo tanto:

$P(A\cup B) = P(A) + P(B\cap A^c)$ y $P(B) = P(A\cap B) + P(B\cap A^c)$ entonces $P(B\cap A^c) = P(B) - P(A\cap B)$ y en consecuencia $P(A\cup B) = P(A)+ P(B)- P(A\cap B)$

4) Si A, B y C son sucesos cualquiera no necesariamente excluyentes entonces:

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

5) Si $A \subseteq B$ entonces $P(A) \leq P(B)$

RESULTADOS IGUALMENTE PROBABLES

Se llaman sucesos equiprobables a los sucesos elementales de un espacio muestral que tienen la misma probabilidad de ocurrir. Es decir, si n es el numero de suceso elementales de E, la probabilidad de cada suceso elemental es la misma, e igual a $\frac{1}{n}$.

Justificación

En un experimento aleatorio existen n resultados diferentes y todos los resultados son igualmente probables. Entonces se tiene que E, se puede representar como:

$E=\{A_1, A_2, A_3, \dots A_i, \dots A_n\}$ Como se sabe que todos los resultados son igualmente probables se tiene que: $P(A_1) = P(A_2) = P(A_3) = \dots = P(A_i) = \dots = P(A_n)$ entonces $E = A_1 \cup A_2 \cup A_3 \cup \dots \cup A_i \cup \dots \cup A_n$

Se tiene que: $P(E) = P(A_1) \cup P(A_2) \cup P(A_3) \cup \dots \cup P(A_i) \cup \dots \cup P(A_n)$

$P(E) = P(A_1) + P(A_2) + P(A_3) + \dots + P(A_i) + \dots + P(A_n)$ por ser mutuamente excluyentes

$1= P(A_1) + P(A_2) + P(A_3) + \dots + P(A_i) + \dots + P(A_n)$

$1= n\,P(A_i)$ porque todas las probabilidades son iguales.

Por lo tanto: $\frac{1}{n}= P(A_i)$, donde $1 \le i \le n$

Lo que indica que la probabilidad que ocurra un resultado cualquiera es igual a uno entre el número total de posibles resultados.

Ejemplo10: Experimentos equiprobables.

1. Sea el experimento el "lanzamiento de un dado simétrico" y $E = \{1,2,3,4,5,6\}$, entonces E es un espacio equiprobable. Es decir: $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = \frac{1}{6}$
2. Sea el experimento "lanzamiento de una moneda equilibrada" y $S = \{c,x\}$ entonces E es un espacio equiprobable.

   $P(c) = P(x)= \frac{1}{2}$
3. Sea el experimento "dos lanzamientos de una moneda" y $E = \{(c,c), (c,x), (x,c), (x,x)\}$, entonces E es un espacio equiprobable. $P(c,c) = P(c,x) = P(x,c) = P(x,x) = \frac{1}{4}$
4. Sea el experimento "dos lanzamientos de una moneda" y $E = \{0, 1, 2\}$ donde 0, 1 y 2 indican el número de caras obtenidas, entonces E no es un espacio equiprobable, puesto que:

{0}: es equivalente a {(x,x)} entonces

$P(0) = \frac{1}{4}$

{1} :es equivalente a { (c,x), (x,c) } entonces

$P(1) = \frac{2}{4}$

{2} :es equivalente a { (c,c) } entonces $P(2)= \frac{1}{4}$

Conectivos "o" u "y"

Suele ser importante el uso correcto de algunos términos en español que se usan cotidianamente en el cálculo de probabilidades.

Conectivo "y"

Este conectivo y significa que se está interesado en la ocurrencia simultánea o conjunta de dos resultados en una situación aleatoria.

Ejemplo 11: Calcular la probabilidad de obtener tres cincos en el lanzamiento de 3 dados.

Como los eventos no están relacionados (son independientes), entonces:

$$P(5 \text{ y } 5 \text{ y } 5) = P(5 \cap 5 \cap 5) = P(5) \times P(5) \times P(5)$$
$$= \frac{1}{6} x \frac{1}{6} x \frac{1}{6} = \frac{1}{216}$$

La respuesta se obtiene relacionando el conectivo "y" con la intersección de sucesos y con la operación aritmética de la multiplicación.

El razonamiento anterior se puede aplicar siempre y cuando la probabilidad de ocurrencia de cualquiera de los dos eventos no afecte la probabilidad de ocurrencia del otro, es decir, cuando ambos eventos no estén relacionados.

Conectivo "o"

Supóngase que se desea calcular la probabilidad de obtener un número par en el lanzamiento de un dado, es decir, calcular la probabilidad de obtener 2 o 4 o 6.

$$P(2ó4ó6) = P(2 \cup 4 \cup 6)$$
$$= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2} = 0.5$$

La respuesta se obtiene relacionando el conectivo "o" con la unión de sucesos y con la operación aritmética adición.

Observación: La condición para poder sumar probabilidades en esta forma es que los eventos sean mutuamente excluyentes, es decir, que no puedan ocurrir conjuntamente. Este procedimiento puede conducir a errores si los eventos no son mutuamente excluyentes.

Ejemplo 12: Sean los eventos:

A: Un tirador acierta en el blanco

B: Otro tirador acierta en el mismo blanco.

Si se sabe que: P(A)=0.8 y P(B)= 0.7 Determinar la probabilidad de que los dos tiradores apuntando al mismo blanco, acierten uno u otro.

Si se quiere utilizar el procedimiento anterior en esta situación, se tiene que: $P(A \text{ o } B) = P(A \cup B) = P(A) + P(B) = 0.8 + 0.7 = 1.5$ resultado evidentemente absurdo, porque como ya se ha señalado, la probabilidad de un evento no puede ser mayor a 1. El error proviene del hecho de no considerar que ambos eventos no son mutuamente excluyentes, porque es muy posible que ambos tiradores hagan blanco simultáneamente. La forma correcta de calcularla es $P(A \text{ ó } B) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$, es decir, se debe restar de 1.5, la probabilidad de que ambos tiradores hagan blanco simultáneamente.

ENFOQUES DE PROBABILIDAD

Aun no existe una interpretación única para definir probabilidad. Los estadísticos, filósofos y científicos no han podido homogenizar el concepto, por lo que existen tres enfoques que son:

1) Frecuencia Relativa
2) Enfoque Clásico
3) Probabilidad Subjetiva

1) Enfoque de frecuencia relativa

Una sencilla manera de obtener la probabilidad de un suceso aleatorio es a través de la tabla de frecuencias relativas de ese experimento. A esa probabilidad la llamamos probabilidad empírica o concepto frecuentista de probabilidad, porque se obtiene una vez realizado el experimento.

Actividad introductoria: Se divide la clase en grupos de 3 o 4 alumnos. Cada grupo realiza 50 lanzamientos de un dado y anota en una tabla siguiente los resultados obtenidos:

Resultados dado	Recuento	Frecuencia Absoluta (n_i)	Frecuencia Relativa (f_i)
1			
2			
3			
4			
5			
6			
Total			

Frecuencia Relativa: $f_i = \dfrac{n_i}{n}$

Donde: ni: frecuencia absoluta

n: número de veces que se lanzo el dado

Después en otra tabla se recogen los datos de toda la clase.

Resultados	G1 (n_i)	G2 (n_i)	G3 (n_i)	G4 (n_i)	G5 (n_i)	G6 (n_i)	G7 (n_i)	…….	Frecuencia Absoluta(n_i)	Frecuencia Relativa(f_i)
1										
2										
3										
4										
5										
6										
Total	50	50	50	50	50	50	50			

¿A qué valor tiende la frecuencia relativa de cada resultado?

Conclusiones

1. Cuando el experimento se repite pocas veces el azar se muestra algo caprichoso, sin embargo al juntar los resultados de toda la clase se aprecia una mayor regularidad.
2. Se observa que en el resultado de cada lanzamiento no influye el anterior, es decir no podemos predecir el resultado antes de lanzar el dado.
3. Es evidente que al lanzar un dado cúbico de seis caras numeradas del 1 al 6 aparecen seis resultados posibles:

   $\{1, 2, 3, 4, 5, 6\}$
4. "A medida que se realizan más lanzamientos la diferencia entre las frecuencias de los resultados tiende a ser más pequeña, concretamente se aproxima al valor $1/6 = 0.1\overline{6}$, y concluye señalando que "en este caso decimos que la probabilidad de que salga cualquiera de los números, es 1/6"

Supongamos que repetimos n veces un experimento aleatorio, y que el suceso A ha ocurrido en n_A ocasiones. Entonces la frecuencia relativa de ocurrencia del suceso A, tal como se ha definido será:

$$fr(A) = \frac{n_A}{n}$$

La frecuencia relativa tiene las siguientes propiedades:

1. $0 \leq f_r(A) \leq 1$ cualquiera que sea el suceso *A*.
2. $fr(A \cup B) = fr(A) + fr(B)$ en el caso que $A \cap B = \phi$
3. $f_r(E) = 1$ (Suceso Seguro) y $f_r(\emptyset) = 0$ (Suceso Imposible)

Además, se puede demostrar que:

- La suma de las frecuencias relativas correspondientes a todos los resultados (sucesos elementales) es 1.
- La frecuencia relativa de un suceso es un número comprendido entre 0 y 1.
- La frecuencia relativa de un suceso es igual a la suma de las frecuencias relativas de los sucesos elementales que lo componen.
- La frecuencia relativa del suceso seguro es 1.
- El suceso seguro lo forman todos los sucesos elementales y la suma de sus frecuencias relativas es 1.
- El suceso imposible no lo forma ningún resultado, es el vacío, y su frecuencia relativa es 0.

La frecuencia relativa de un suceso A, tiende a estabilizarse en torno a un número a medida que el número de pruebas del experimento crece indefinidamente. A este número se le llama **Probabilidad del suceso A.**

Esta propiedad es conocida como **"Ley de los grandes números"** establecida por *Jakob* Bernouilli*:* la cual dice "Que al repetir un experimento aleatorio un número muy grande de veces, la frecuencia relativa de cada uno de los sucesos elementales tiende a estabilizarse aproximándose a un número fijo que es la *probabilidad* de que ese suceso ocurra".

Definición: Dado un experimento aleatorio, la ***probabilidad de un suceso* A*, P*(A)**, es el valor hacia el cual se aproximan las frecuencias relativas de dicho suceso conforme aumenta el número de realizaciones del experimento. Es decir, la **probabilidad** de un suceso es la proporción de veces que el suceso ocurriría en un número muy grande de pruebas. Y se estima de la siguiente forma:

$$P(A) = \frac{\text{Número de veces que ocurrió } A}{\text{Número de veces que se repite el experimento}}$$

$$P(A) = \frac{N_A}{n}$$

La frecuencia relativa y, por tanto, la probabilidad, se aproximan más y más cuanto mayor es el número de repeticiones de un mismo experimento aleatorio. Es decir:

$$\mathrm{P(A)} = \lim_{n\to\infty} fr(A) = \lim_{n\to\infty} \frac{\mathrm{n_A}}{n}$$

La existencia de dicho límite, ya era claramente intuida desde la antigüedad en los juegos de azar, sin embargo; este límite se basa en resultados experimentales o empíricos, y por tanto no da un procedimiento de cálculo para la probabilidad, sino sólo un valor aproximado a ésta.

La probabilidad de frecuencia relativa, es llamada también **probabilidad empírica o a posteriori**, debido a que se obtiene el resultado después de llevar a cabo el experimento un gran número de veces.

La probabilidad frecuencial de un evento es el valor fijo al que tienden las frecuencias relativas de ocurrencia del evento de acuerdo a la regularidad estadística. Esta definición sería la más real, pero proporciona probabilidades aproximadas, es decir, proporciona estimaciones y no valores reales. Además, los resultados son a posteriori, pues se necesita realizar el experimento para poder obtenerlo.

Inconveniente: Esta definición presenta el inconveniente de tener que realizar el experimento; un gran número de veces y además siempre obtendremos un valor aproximado de la probabilidad.

2) Enfoque de probabilidad clásica

Como se ha podido comprobar, resulta bastante tedioso asignar probabilidades a partir de las frecuencias relativas, pues es necesario realizar el experimento una gran cantidad de veces para conseguir una buena aproximación de la verdadera probabilidad de un suceso y, aun así, nunca se está seguro de conseguirla. Por esa razón es necesario introducir un método alternativo para el cálculo de probabilidades que sea más manejable.

Este enfoque se basa en el supuesto de que todos los resultados posibles de un experimento aleatorio son igualmente probables; es decir, cada uno de los elementos del espacio muestral tienen la misma probabilidad de salir.

Una probabilidad tiene que ser definida de forma que a cada suceso le corresponda un número, y que se cumplan las propiedades de la frecuencia relativa o axiomas de probabilidad. Estos requisitos se cumplen con la siguiente definición que se conoce como **Regla de Laplace**.

Definición: La probabilidad de un suceso aleatorio A, que se representa por P(A), es igual al cociente del número de casos favorables a dicho suceso, entre el número de casos posibles del experimento. Es decir:

$$P(A) = \frac{\text{Número de casos favorables al suceso A}}{\text{Número de casos posibles}}$$

De donde los **casos posibles** son todos los resultados del experimento, es decir, el número de puntos muéstrales del espacio muestral E y los **casos favorables** son los puntos muéstrales que componen el suceso A.

Esta forma de obtener la probabilidad se denomina a priori debido a que es posible conocer el resultado con anterioridad, es decir sin llevar a cabo el experimento y sólo basado en un razonamiento lógico.

Se puede observar que:

- P(A) es el cociente de dos números positivos; por lo tanto se cumple que $P(A) \geq 0$.
- El número de casos favorables a A nunca puede ser mayor que el número de casos posibles; por lo tanto $P(A) \leq 1$.

La probabilidad definida de esta forma cumple con las siguientes propiedades:

- $P(A) \geq 0$
- $P(E) = 1$
- Si A y B son sucesos, cumpliendo que $A \cap B = \emptyset$, entonces se cumple: $P(A \cup B) = P(A) + P(B)$.

Inconvenientes

- Al exigir la simetría de los casos se está suponiendo una idea de igualdad de probabilidad.
- Solo es aplicado en el caso que el número total de casos es finito.

3) Enfoque de probabilidad subjetiva

Las probabilidades obtenidas mediante el enfoque de frecuencia relativa reciben el nombre de probabilidades objetivas, ya que se derivan de hechos.

Existen varios sucesos de sumo interés cuyas probabilidades no se pueden calcular tomando en cuenta los métodos de frecuencia relativa ni con la teoría de la probabilidad clásica. Surge entonces, el punto de vistas subjetivo el cual hace hincapié en la probabilidad que resulta de una opinión, creencia, o juicio personal sobre una situación determinada. El enfoque subjetivo denominado también probabilidad personal, asigna a los eventos probabilidades, aun cuando los datos experimentales sean escasos o imposibles de obtener.

Los que toman decisiones utilizando este tipo de probabilidad se fundamentan en sus propias experiencias personales y en muchos casos en presentimientos.

La probabilidad personal se ha vuelto sistemáticamente popular entre los teóricos de la toma de decisiones. Los defensores de esta corriente tratan de buscar soluciones a la asignación de probabilidades de aquellos eventos que solo ocurren una vez o que no pueden estar sometidos a experimentos repetidos. La asignación de probabilidades a un evento en estas condiciones, más que un juicio arbitrario, es un juicio de valor.

La probabilidad subjetiva de un evento se la asigna la persona que hace el estudio, y depende del conocimiento que esta persona tenga sobre el tema. Precisamente por su carácter de subjetividad no se considera con validez científica, aunque en la vida diaria es de las más comunes que se utilizan al no apoyarse más que en el sentido común y los conocimientos previos, y no en resultados estadísticos.

Por ejemplo, cuando alguien opina que hay un 80% de posibilidades de que un grupo de alpinistas conquiste el monte Everest, la persona está asignando subjetivamente una probabilidad de 0.8 al suceso. Es claro que esta opinión puede o no ser compartida por otras personas, pues es posible que los demás señalen otras probabilidades de éxito a los alpinistas. Aquí no existe un espacio muestral finito, ni los elementos son igualmente posibles, ni se puede calcular la frecuencia relativa, ni hay forma de hacer intervenir los enunciados de la probabilidad axiomática; lo cual manifiesta claramente que la probabilidad subjetiva sólo refleja el grado de seguridad o credibilidad que una persona tiene o asigne sobre la ocurrencia de un suceso.

Inconvenientes

- Las estimaciones subjetivas suelen ser difíciles de comprobar si son cuestionadas.
- Los prejuicios pueden influir. Las ideas preconcebidas respecto a lo que debería suceder pueden afectar la objetividad, así como los sentimientos acerca de lo que uno quiere que suceda. Algunas veces es difícil eliminar estos prejuicios, ya que por lo regular son inconscientes. La capacidad, experiencia y actitud profesional pueden ayudar a superar tales dificultades.

EJEMPLOS

Ejemplo13: En una caja hay diez bolas numeradas del 0 al 9. Se repite 100 veces el experimento de extraer una bola y remplazarla. Los resultados obtenidos fueron:

No. Bola	0	1	2	3	4	5	6	7	8	9	Suma
n_i	7	13	11	12	8	10	12	6	10	11	100

Sean los sucesos: A: «múltiplo de 3», B: «número impar» y C: «divisor de 6». Encontrar la frecuencia relativa de los sucesos: A, B, C, $A \cup B$, $A \cap B$, $A \cup C$ y $A \cap C$.

Solución

Primeramente se obtendrá la frecuencia relativa de cada uno de los sucesos elementales, utilizando la formula dada. $f_i = \frac{n_i}{n}$

No. Bola	0	1	2	3	4	5	6	7	8	9	Suma
n_i	$\frac{7}{100}=0.07$	$\frac{13}{100}=0.13$	$\frac{11}{100}=0.11$	$\frac{12}{100}=0.12$	$\frac{8}{100}=0.08$	$\frac{10}{100}=0.10$	$\frac{12}{100}=0.12$	$\frac{6}{100}=0.06$	$\frac{10}{100}=0.10$	$\frac{11}{100}=0.11$	1

Los puntos muéstrales de los sucesos son:

$A = \{3, 6, 9\}$ $B = \{1, 3, 5, 7, 9\}$ $C = \{1, 2, 3, 6\}$ $A \cup B = \{1, 3, 5, 6, 7, 9\}$ $A \cap B = \{3, 9\}$

$A \cup C = \{1, 2, 3, 6, 9\}$ $A \cap C = \{3, 6\}$

La frecuencia relativa de cada suceso viene dada como la suma de las frecuencias relativas de los puntos muéstrales que está compuesto.

Frecuencia Relativa del suceso A

$A = \{3, 6, 9\}$ $fr(A) = fr(3) + fr(6) + fr(9) = 0.12 + 0.12 + 0.11 = 0.35$

Frecuencia Relativa del suceso B

$B = \{1, 3, 5, 7, 9\}$ $fr(B) = fr(1)+fr(3)+fr(5)+fr(7)+fr(9) = 0.13 + 0.12 + 0.10 + 0.06 + 0.11 = 0.52$

Frecuencia Relativa del suceso C

$C = \{1, 2, 3, 6\}$ $fr(C) = fr(1) + fr(2) + fr(3) + fr(6) = 0.13 + 0.11 + 0.12 + 0.12 = 0.48$

Frecuencia Relativa del suceso $A \cup B$

$A \cup B = \{1, 3, 5, 6, 7, 9\}$ $fr(A \cup B) = fr(1) + fr(3) + fr(5) + fr(6) + fr(7) + fr(9)$

$= 0.13 + 0.12 + 0.10 + 0.12 + 0.06 + 0.11 = 0.64$

Frecuencia Relativa del suceso $A \cap B$

$A \cap B = \{3, 9\}$ $fr(A \cap B) = fr(3) + fr(9) = 0.12 + 0.11 = 0.23$

Frecuencia Relativa del suceso $A \cup C$

$A \cup C = \{1, 2, 3, 6, 9\}$ $fr(A \cup C) = fr(1) + fr(2) + fr(3) + fr(6) + fr(9)$

$= 0.13 + 0.11 + 0.12 + 0.12 + 0.11 = 0.59$

Frecuencia Relativa del suceso $A \cap C$

$A \cap C = \{3, 6\}$ $fr(A \cap C) = fr(3) + fr(6) = 0.12 + 0.12 = 0.24$

Ejemplo 14: En un centro escolar hay 1000 alumnos y alumnas distribuidos de la siguiente manera:

	CHICOS	CHICAS
USAN GAFAS	147	135
NO USAN GAGAS	368	350

Si se elige al azar uno de ellos. Calcular las probabilidades de que:

a) Sea chico b) Sea chica c) Use Gafas d) No use gafas e) Sea una chica con gafas

Solución:

Sean los sucesos: C: Chicos $\overline{C}$: Chicas G: Usa Gafas $\overline{G}$: No usa Gafas

Completando la tabla se tiene:

	C	$\overline{C}$	Total
G	147	135	282
$\overline{G}$	368	350	718
Total	515	485	1000

Solución a) La probabilidad de que sea chico es:

$$P(C) = \frac{\text{Total de chicos}}{\text{Total de chichos y chicas}} = \frac{515}{1000} = 0.515$$

Solución b) La probabilidad de que sea chica es:

$$P(\overline{C}) = \frac{\text{Total de chicas}}{\text{Total de chichos y chicas}} = \frac{485}{1000} = 0.485$$

Solución c) La probabilidad de que use gafas es:

$$P(G) = \frac{\text{Total de los usan gafas}}{\text{Total de chichos y chicas}} = \frac{282}{1000} = 0.282$$

Solución d) La probabilidad de que no use gafas es:

$$P(G) = \frac{\text{Total que no usan gafas}}{\text{Total de chichos y chicas}} = \frac{718}{1000} = 0.718$$

Solución e) La probabilidad de que sea una chica con gafas es:

$$P(G) = \frac{\text{Total de chicas con gafas}}{\text{Total de chichos y chicas}} = \frac{135}{1000} = 0.135$$

Ejemplo 15: Manuel trabaja en una estación de ferrocarril vendiendo café y jugo de naranja a los usuarios. El martes pasado vendió 60 cafés grandes, 25 cafés chicos, 45 jugos grandes y 20 jugos pequeños. Si esta distribución refleja con precisión la preferencia de sus clientes.

a) ¿Cuál es la probabilidad de que su primer cliente del siguiente martes compre un jugo de naranja grande?

b) ¿Cuál es la probabilidad de que su primer cliente del siguiente martes compre un café?

c) ¿Cuál es la probabilidad de que su primer cliente del siguiente martes compre té?

Solución

Sean los sucesos:

J: comprar jugo de naranja grande   C: comprar un café   T: comprar un té.

Número de clientes que compraron cafés grandes: 60

Número de clientes que compraron cafés chicos: 25

Número de clientes que compraron jugos grandes: 45

Número de clientes que compraron Jugos pequeños: 20

Solución a) ¿Cuál es la probabilidad de que su primer cliente del siguiente martes compre un jugo de naranja grande?

$P(J) = \dfrac{\text{Número de clientes que compraron un jugo de naranja grande}}{\text{Número total de cientes}} = \dfrac{45}{150} = 0.3$. Lo que significa que el suceso J tiene la posibilidad de no suceder; es decir que el primer cliente del siguiente martes es posible que no le compre un jugo de naranja grande.

Solución b) ¿Cuál es la probabilidad de que su primer cliente del siguiente martes compre un café?

$P(C) = \dfrac{\text{Número de clientes que compraron café}}{\text{Número total de cientes}} = \dfrac{60+25}{150} = \dfrac{85}{150} = 0.57$. Lo que significa que el suceso C tiene la posibilidad de suceder; es decir que el primer cliente del siguiente martes es posible que le compre un café.

Solución c) ¿Cuál es la probabilidad de que su primer cliente del siguiente martes compre té?

Ninguno de los clientes compro té, pues no estaba dentro de las opciones. Por eso, la probabilidad de que el primer cliente compre té es: $P(T) = \dfrac{\text{Número de clientes que compraron té}}{\text{Número total de cientes}} = \dfrac{0}{150} = 0$.

Lo que significa que T es el suceso imposible y por lo tanto tiene probabilidad cero.

Ejemplo 16: En un depósito se encuentran 12 semillas de rosas rojas y 8 semillas de rosas amarillas, se seleccionan dos semillas aleatoriamente. ¿Cuál es la probabilidad de que:

a) ambas semillas resulten de rosas rojas?

b) una semilla resulte de rosas rojas y la otras semilla de rosas amarillas?

Solución:

En total hay 20 semillas de ambos colores en el depósito, de donde se seleccionan las dos semillas.

Como no es importante el orden en que deben aparecer las semillas, entonces al seleccionar dos de las 20 corresponde a un caso de combinación, es decir, 20 semillas tomadas de dos en dos.

Así, el número total de cados posibles es: $\binom{20}{2} = \frac{20!}{(20-2)!2!} = \frac{20!}{18!2!} = \frac{20x19x18!}{18!2!} = \frac{20x19}{2} = \frac{380}{2} = 90$

Para los casos favorables se debe tener en cuenta la condición impuesta por cada pregunta.

Solución a) ¿Cuál es la probabilidad de que ambas semillas resulten de rosas rojas?

Sea el evento R: Ambas semillas seleccionadas resultan de rosas rojas.

Hay 12 semillas de rosas rojas tomadas de dos en dos, entonces los casos favorables son: $\binom{12}{2} = \frac{12!}{(12-2)!2!} = \frac{12!}{10!2!} = \frac{12x11x10!}{10!2!} = \frac{12x11}{2} = \frac{132}{2} = 66$

Por lo tanto, la probabilidad pedida es:

$$P(R) = \frac{\text{Número de casos favorables a R}}{\text{Número de casos posibles}} = \frac{66}{190} = 0.66$$

Solución b) ¿Cuál es la probabilidad de que una semilla resulte de rosas rojas y la otra semilla de rosas amarillas?

D: una semilla de cada color

Los casos favorables:

$$\text{Semillas rojas} = \binom{12}{1} = \frac{12!}{(12-1)!1!} = \frac{12!}{11!1!} = \frac{12x11!}{11!1!} = \frac{12}{1} = 12$$

$$\text{Semillas amarillas} = \binom{8}{1} = \frac{8!}{(8-1)!1!} = \frac{8!}{7!1!} = \frac{8x7!}{7!1!} = \frac{8}{1} = 8$$

Por lo tanto, los casos favorables para el evento D es: 12x8= 96

Así, la probabilidad de D es: $P(D) = = \frac{\text{Número de casos favorables a D}}{\text{Número de casos posibles}} = \frac{96}{190} = 0.5052$

PROBABILIDAD CONDICIONAL

Hasta ahora hemos definido la probabilidad de un suceso *A* referida a todo el espacio muestral *E* del experimento. Supongamos ahora la existencia de otro suceso *B* definido sobre *E* y que no sea incompatible con *A*, es decir que $(A \cap B) \neq \varphi$. Esto significa que los sucesos *A* y *B* tienen puntos muestrales en común. Supongamos adicionalmente que tenemos la certeza de que ha ocurrido el suceso *B*. Ahora estamos interesados en saber cómo cambia la probabilidad de A sabiendo que ha ocurrido B. Sabiendo que ha ocurrido *B*, la probabilidad de que ocurra *A* se representa por $P(A / B)$ y se le conoce como *probabilidad condicional*. En estas circunstancias, para calcular la probabilidad de *A* hay que cambiar el espacio de referencia el cual, ahora ya no es *E* sino *B*, y habrá que exigir que no sea un espacio nulo, es decir, debe cumplirse que $P(B) > 0$. Si sabemos que el suceso *B* ha ocurrido, entonces se sabe que el resultado del experimento es uno de los incluidos en *B*. Por tanto, para evaluar la probabilidad de que ocurra *A*, se debe considerar el conjunto de los resultados incluidos en *B* que también implique la ocurrencia de *A*. Este conjunto viene dado por la intersección de *A* y *B*, es decir $(A \cap B)$.

Actividad introductoria: Sea el experimento que consiste en lanzar dos dados, y sean A y B dos sucesos asociados a él.

El suceso A ocurre cuando se obtienen cifras pares en ambos dados y B cuando las dos cifras son iguales.

Entonces:

A: Las cifras son pares en ambos dados

A = {(2,2),(2,4) ,(4,2) ,(2,6) ,(6,2) ,(4,6) ,(6,4) ,(4,4) ,(6,6)}

B: Los números son iguales

B = {,(1,1) ,(2,2) ,(3,3) ,(4,4) ,(5,5) ,(6,6)}

El espacio muestral está formado por 36 sucesos elementales ($VR_{6,2}=6^2=36$)

El suceso A tiene 9 sucesos elementales ($VR_{3,2} = 3^2=9$)

El suceso B tiene 6 sucesos elementales.

De acuerdo a la definición clásica, sus probabilidades son: $P(A) = \frac{9}{36}$ y $P(B) = \frac{6}{36}$

Pero ahora se quiere calcular la probabilidad de B, cuando previamente se conoce que ha ocurrido A, se considera que el espacio muestral ha cambiado, pasando a ser A, y se debe analizar entre los sucesos elementales de A los que corresponden a B, y se obtiene: Espacio Muestral A={(2,2), (2,4), (4,2), (2,6), (6,2), (4,6), (6,4), (4,4), (6,6)}

La probabilidad de B dado que ha ocurrido A es 3/9 que se simboliza como $P(B/A) = \frac{3}{9} = \frac{1}{3}$.

Obsérvese que la probabilidad de B se ha modificado al ocurrir después de A. En consecuencia, la aparición de A influye en la verificación de B. En la situación descrita B es un sucedo dependiente de A.

DEFINICIONES

Sean A y B dos sucesos de un mismo experimento aleatorio, y $P(B) \neq 0$. Se llama probabilidad del suceso A condicionado por B, y la denotamos por $p(A / B)$ al número definido por la fórmula:

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$ Se lee "probabilidad de A condicionada a B"

De lo anterior se deduce que $P(A \cap B) = P(B)\, P(A/B)$

Sean A y B dos sucesos de un mismo experimento aleatorio, y $p(A) \neq 0$. Llamamos probabilidad del suceso B condicionado por A, y la denotamos por $p(B / A)$ al número definido por la fórmula:

$$P(B/A) = \frac{P(A \cap B)}{P(A)}$$ Se lee "probabilidad de B condicionada a A"

De lo anterior se deduce que $P(B \cap A) = P(A)\, P(B/A)$.

Ejemplo 17: La probabilidad de que un hombre casado vea cierto programa de televisión es 0,4 y la probabilidad de que una mujer casada vea el programa es 0,5. La probabilidad de que un hombre vea el programa, dado que su esposa lo hace, es 0,7. Encuentre la probabilidad de que:

a) Un matrimonio vea el programa.
b) Una esposa vea el programa dado que su esposo lo ve
c) Al menos una persona de un matrimonio vea el programa.
   Solución

Sea:

H: El hombre vea televisión

M: La mujer vea televisión

Entonces: $P(H)=0.4$ $P(M)=0.5$ $P(H/M)=0.7$

Solución a) Un matrimonio vea el programa.

La probabilidad de que un matrimonio vea el programa es la probabilidad de que el hombre y la mujer vean el programa, es decir, la probabilidad de la intersección de H y M.

$$P(H/M) = \frac{P(H \cap M)}{P(M)}$$

Entonces:

$P(H \cap M)=P(H/M)P(M)=0.7*0.5=0.35$

Solución b) Una esposa vea el programa dado que su esposo lo ve.

La probabilidad condicional pedida es: $P(M/H)$

$$P(M/H) = \frac{P(M \cap H)}{P(H)} = \frac{0.35}{0.40} = 0.875$$

Solución c) Al menos una persona de un matrimonio vea el programa.

$P(M \cup H) = P(M) + P(H) - P(M \cap H)$

$P(M \cup H) = 0.5 + 0.4 - 0.35 = 0.55$

Una vez dado el concepto de probabilidad condicional no resulta difícil demostrar que esta definición satisface los siguientes tres resultados de la probabilidad.

REGLA DEL PRODUCTO

Si en ambas definiciones de probabilidad condicional se despeja $P(A\cap B)$, se tiene:

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

A esta forma expresar la probabilidad de la intersección de dos sucesos se le conoce como **regla del producto**.

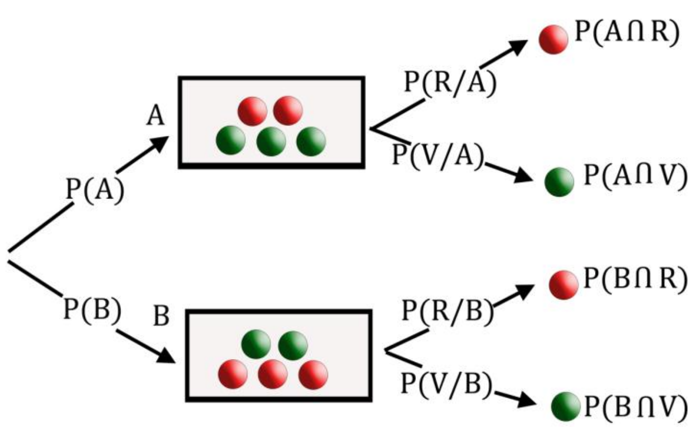
Si en lugar de tener dos sucesos se tuvieran tres, entonces la probabilidad de la intersección de los tres vendrá dada por:

$P(A\cap B\cap C) = P(A)\,P(B/A)\,P(C/A\cap B)$

Y en general para un número k de sucesos viene dada por:

$P(A_1\cap A_2\cap \ldots\cap A_k) = P(A_1)$
$P(A_2/A_1) \ldots P(A_k/A_1\cap A_2\cap \ldots \cap A_{k-1})$

Estas relaciones se pueden observar con mayor facilidad utilizando un diagrama de árbol de probabilidades. Por ejemplo, se tiene el siguiente experimento : Dos urnas, *A* y *B* ,la urna *A*, contiene 3 bolas verdes y 2 bolas rojas, la urna *B* contiene 2 bolas verdes y 3 bolas rojas. Se realiza el experimento en dos tiempos, primero se selecciona urna por un procedimiento aleatorio y posteriormente de la urna elegida se extrae una bola. El diagrama de árbol de probabilidades es el siguiente:



En las primeras ramas se colocan las **probabilidades apriori**, en las segundas las **probabilidades condicionales**. El producto de estas da origen a las **probabilidades conjuntas.**

Por ejemplo, la probabilidad de que sea roja la bola seleccionada sabiendo que pertenece a la urna A se obtiene como:

$$P(R/A) = P(A\cap R)\times\left(\frac{1}{P(A)}\right) = \frac{P(A\cap R)}{P(A)} \iff P(A\cap R) = P(A)\times P(R/A)$$

Ejemplo 18: En un sistema de alarma la probabilidad que se produzca un peligro es de 0,10; si este produce, la probabilidad que la alarma funcione es de 0.95. La probabilidad que funcione la alarma sin haber peligro es 0.03.

Determinar la probabilidad que haya un peligro y la alarma funcione.

Solución:

Sean los eventos:

P: Existe peligro $\overline{P}$: No existe peligro

F: La alarma funcione

Entonces: $P(P)=0.10$ $P(F/P)=0.95$

La probabilidad pedida es: $P(P\cap F) = P(P)P(F/P) = 0.95*0.10 = 0.095$

La probabilidad de que haya un peligro y la alarma funciones es de 0.095

SUCESOS INDEPENDIENTES

En este caso la presencia de B no altera la probabilidad del suceso A. En estas circunstancias se dice que la probabilidad de A no depende de la presencia de B. Esta idea se puede expresar también diciendo que A y B son dos sucesos independientes.

Es decir, los sucesos A y B se dicen que son independientes cuando la presencia de uno de ellos no afecta a la probabilidad del otro.

Sean A y B dos sucesos del espacio muestral. El suceso A se dice independiente del suceso B si el conocimiento de la ocurrencia de B no modifica la probabilidad de aparición de A, es decir, tienen que verificar al menos una de las siguientes **condiciones**:

- $P(B/A) = P(B)$: Es decir, que la probabilidad de que se dé el suceso B, condicionada a que previamente se haya dado el suceso A, es exactamente igual a la probabilidad de B. **Ejemplo:** la probabilidad de que al tirar una moneda salga cara (suceso B), condicionada a que haga buen tiempo (suceso A), es igual a la propia probabilidad del suceso B.
- $P(A/B) = P(A)$: Es decir, que la probabilidad de que se dé el suceso A, condicionada a que previamente se haya dado el suceso B, es exactamente igual a la probabilidad de A. **Ejemplo:** la probabilidad de que haga buen tiempo (suceso A), condicionada a que al tirar una moneda salga cara (suceso B), es igual a la propia probabilidad del suceso A.
- $P(A\cap B)= P(A) \text{ x } P(B)$: Es decir, que la probabilidad de que se dé el suceso conjunto A y B es exactamente igual a la probabilidad del suceso A multiplicada por la probabilidad del suceso B. **Ejemplo:** la probabilidad de que haga buen tiempo (suceso A) y salga cara al tirar una moneda (suceso B), es igual a la probabilidad del suceso A multiplicada por la probabilidad del suceso B.

Pero que dos sucesos sean independientes no significa que sean mutuamente excluyentes. Este segundo caso se da cuando esos sucesos no pueden ocurrir simultáneamente y, por lo tanto, su intersección es el suceso imposible, por lo que su probabilidad será nula.

Si en lugar de tener los sucesos A y B se tuvieran los sucesos A, B y C, entonces se diría que los tres son independientes si lo son dos a dos y los tres a la vez. Es decir si se cumple que:

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

$P(A\cap C) = P(A/C)\ P(C) = P(A)P(C)$

$P(B\cap C) = P(B/C)\ P(C) = P(B)P(C)$

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

Ejemplo 19: Una clase se compone de veinte alumnos y diez alumnas. La mitad de las alumnas y la mitad de los alumnos aprueban las matemáticas. Calcula la probabilidad de que, al elegir una persona al azar, resulte ser:

a) Alumna o que aprueba las matemáticas.

b) Alumno que suspenda las matemáticas.

c) Sabiendo que es alumno, ¿Cuál es la probabilidad de que apruebe las matemáticas?

d) ¿Son independientes los sucesos ALUMNO y APRUEBA MATEMÁTICAS?

Solución:

Sean los sucesos: A: alumnas $\overline{A}$: alumnos AM: Aprueba las matemáticas $\overline{AM}$: Suspenden las matemáticas.

Tabla de contingencia

	Alumnas (A)	Alumnos ($\overline{A}$)	Total
Aprueban Mat(AM)	5	10	15
Suspenden Mat($\overline{AM}$)	5	10	15
Total	10	20	30

Solucion a)

$$P(A \cup AM) = P(A) + P(AM) - P(A \cap AM) = \frac{10}{30} + \frac{15}{30} - \frac{5}{30} = \frac{20}{30} = \frac{2}{3} = 0.666$$

Solución b) $P(\overline{A} \cap AM) = \frac{10}{30} = \frac{1}{3} = 0.333$

Solución c) $P(AM/\overline{A}) = \frac{P(AM \cap \overline{A})}{P(\overline{A})} = \frac{\frac{10}{30}}{\frac{20}{30}} = \frac{300}{600} = \frac{1}{2} = 0.5$

Solución d) Hay que verificar si: $P(\overline{A} \cap AM) = P(\overline{A}) * P(AM)$ calculando los valores.

$P(\overline{A} \cap AM) = \frac{10}{30} = \frac{1}{3} = 0.333$ $\quad$ $P(AM) = \frac{15}{30} = \frac{1}{2} = 0.5$ $\quad$ $P(\overline{A}) = \frac{20}{30} = \frac{2}{3} = 0.666$

$P(\overline{A} \cap AM) = 0.333$ $\quad$ $P(\overline{A}) * P(AM) = (0.5) * (0.666) = 0.333$

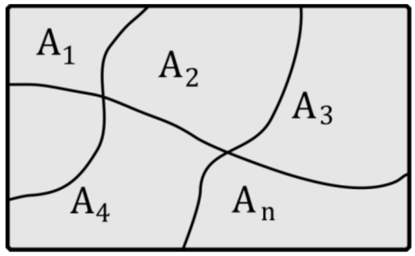
Como la igualdad son independientes; es decir que ser alumno no implica aprobar matemática.

Probado de otra formar se debe verificar si: $P(AM/\overline{A}) = P(AM)$

$P(AM/\overline{A}) = 0.5$ $\quad$ $P(AM) = 0.5$ $\quad$ Se cumple ; por lo tanto son independientes.

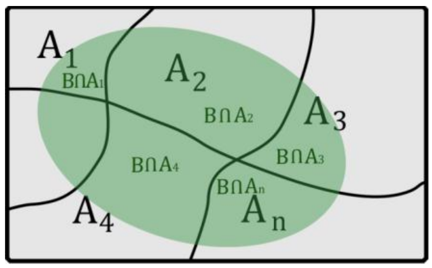
PROBABILIDAD TOTAL

Se llama **partición** al conjunto de eventos $\mathbf{A}_i$ tales que $E = A_1 \cup A_2 \cup \ldots\ldots \cup A_n$ y $A_i \cup A_j = \emptyset$; es decir un conjunto de eventos mutuamente excluyentes y que componen todo el espacio muestral E. La siguiente figura presenta el diagrama de Venn que corresponde a la partición de un espacio muestral E en $\boldsymbol{A_n}$ eventos.



Para cualquier evento B, éste puede definirse como un evento compuesto de varios subconjuntos mutuamente excluyentes, esto es:

$$B=(B\cap A_1)\cup (B\cap A_2)\cup .... \cup(B\cap A_n)$$



La **probabilidad total** de un evento es la suma exhaustiva de las probabilidades de todos los casos mutuamente excluyentes que conducen a dicho evento., Es así como la **regla de probabilidad total** afirma:

$$P(B) = P(B\cap A_1) + P(B\cap A_2) + \cdots + (B\cap A_n)$$
$$P(B) = P(B/A_1)\cdot P(A_1) + P(B/A_2)\cdot P(A_2) + \cdots + P(B/A_n)\cdot P(A_n)$$

En términos de sumatoria, quedar expresada así:

$$P(B) = \sum_{i=1}^{n} P(B/A_n)\cdot P(A_n)$$

Conjuntamente con este resultado se deduce un teorema muy importante en probabilidad conocido como teorema de Bayes.

Según lo anterior si se cumplen las siguientes condiciones:

Sean $A_1, A_2, \ldots\ldots A_n$ sucesos tales que:

i) $A_i \cap A_j = \phi \quad \forall i \neq j$ (disjuntos dos a dos)

ii) $E = \bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \ldots\ldots \cup A_n$

iii) $P(A_i) \neq 0\ \forall i$

y sea B otro suceso de E para el que se conocen las probabilidades $P(B/A_i)$, i=1,2,…,n Entonces:

$$P(B)=\sum_{i=1}^{n} P(B/A_i)P(A_i)$$ Teorema de probabilidad Total

$$P(A_i/B)=\frac{P(B\cap A_i)}{P(B)}=\frac{P(B/A_i)P(A_i)}{\sum_{i=1}^{n} P(B/A_i)P(A_i)}$$ Teorema de Bayes

Las probabilidades $P(A_i)$ se denominan probabilidades a priori.

Las probabilidades $P(A_i/B)$ se denominan probabilidades a posteriori.

Las probabilidades $P(B/A_i)$ se denominan verosimilitudes.

Ejemplo 20: En un sistema de alarma, la probabilidad de que esta funcione habiendo peligro es 0.95 y la de que funcione por error sin haber peligro es 0.03. Si la probabilidad de haber peligro es 0.1:

a) Calcular el porcentaje de veces que habiendo funcionado la alarma no haya peligro.

b) Hallar la probabilidad de que haya peligro y la alarma no funcione.

c) Calcular la probabilidad de que no habiendo funcionado la alarma haya peligro.

d) ¿Cuál es la probabilidad de que la alarma funcione?

Solución:

Definimos los sucesos: A: "Hay situación de peligro" F: "La alarma funciona"

A^c: "No hay situación de peligro" F^c: " La alarma no funciona"

Entonces: $P(F/A)=0.95$, $P(F/A^c)=0.03$ $P(A)=0.1$ $P(A^c)=1-P(A)=1-0.1=0.90$

Solución a) Calcular el porcentaje de veces que habiendo funcionado la alarma no haya peligro.

$$P(A^c/F)=\frac{P(F/A^C)\ P(A\ )}{P(F/A)\ P(A)+P(F/A^C)\ P(A\ )}=\frac{(0.03)(0.9)}{(0.95)(0.1)+(0.03)(0.9)}=\frac{0.027}{0.095+0.027}=\frac{0.027}{0.122}=0.2213$$, luego el porcentaje es: 22.13%

Solución b) Hallar la probabilidad de que haya peligro y la alarma no funcione.

$P(A\cap F^C)=P(F^C/A)P(A)=(0.05)(0.1)=0.005$

Solución c) Calcular la probabilidad de que no habiendo funcionado la alarma haya peligro.

$$P(A/F^c)=\frac{P(F^c/A)\ P(A)}{P(F^c/A)\ P(A)+P(F^c/A^C)\ P(A\ )}=\frac{(0.05)(0.1)}{(0.05)(0.1)+(0.97)(0.9)}=\frac{0.005}{0.005+0.873}=\frac{0.005}{0.878}=0.005694$$

Solución d) ¿Cuál es la probabilidad de que la alarma funcione?

$P(F)=P(F/A)P(A)+P(F/A^c)P(A^c)=(0.95)(0.1)+(0.03)(0.9)=0.095+0.027=0.122$

APLICANDO LO APRENDIDO

1. Un estudiante responde al azar a cuatro preguntas de verdadero o falso.
   a. Escriba el espacio muestral.
   b. Escriba el suceso responder "falso" a una sola pregunta.
   c. Escriba el suceso responder "verdadero" al menos a 3 preguntas.
   d. Escriba la unión de estos dos sucesos, la intersección y la diferencia del 2º y el 1º.

2. Se tiene una urna con nueve bolas numeradas del 1 al 9. Se realiza el experimento, que consiste en sacar una bola de la urna, anotar el número y devolverla a la urna. Considérese los siguientes sucesos: A="salir un número primo" y B="salir un número cuadrado". Responde a las cuestiones siguientes:
   a. Calcula los sucesos $A \cup B$ y $A \cap B$.
   b. Los sucesos A y B, ¿son compatibles o incompatibles?
   c. Encuentra los sucesos contrarios de A y B.

3. Un taller sabe que por término medio acuden: por la mañana tres automóviles con problemas eléctricos, ocho con problemas mecánicos y tres con problemas de chapa, y por la tarde dos con problemas eléctricos, tres con problemas mecánicos y uno con problemas de chapa.
   a. Hacer una tabla ordenando los datos anteriores.
   b. Calcular el porcentaje de los que acuden por la tarde. R/30%
   c. Calcular el porcentaje de los que acuden por problemas mecánicos. R/55%
   d. Calcular la probabilidad de que un automóvil con problemas eléctricos acuda por la mañana. R/0.6

4. Una empresa planea probar un nuevo producto en una zona de mercado elegida aleatoriamente. Las zonas de mercado pueden clasificarse con base en la ubicación y en la densidad de población. El numero de mercados en cada categoría se muestran en la siguiente tabla:

	Densidad de Población		Total
Ubicación	Urbano(U)	Rural(R)	Total
Este(E)	25	50	75
Oeste(O)	20	30	50
Total	45	80	125

   a. ¿Cuál es la probabilidad que el mercado de prueba elegido este en Este, P(E)?
   b. ¿Cuál es la probabilidad que el mercado de prueba elegido este en Oeste, P(O)?
   c. ¿Cuál es la probabilidad que el mercado de prueba elegido este en una zona Urbana, P(U)?
   d. ¿Cuál es la probabilidad que el mercado de prueba elegido este en una zona Rural, P(R)?
   e. ¿Cuál es la probabilidad que el mercado este en una zona Rural del Oeste, P(O y R)?
   f. ¿Cuál es la probabilidad de que este en una zona del Este o en una zona Urbana, P(E o U)?
   g. ¿Cuál es la probabilidad de que si este en el Este , sea una zona Urbana, P(U/E)?
   h. ¿Son la Ubicación y la densidad de la población independientes?
   i. ¿Qué significa independencia o dependencia en esta situación?

5. Sean los sucesos A: "haga buen tiempo" con probabilidad del 0,4 y B: "tener un accidente" con probabilidad de 0,1. Verifique si ambos sucesos son independientes y realice sus respectivas conclusiones.

6. Una compañía dedicada al transporte público tiene tres líneas en una ciudad, de forma que el 60% de los autobuses cubre el servicio de la primera línea, el 30% cubre la segunda y el 10% cubre el servicio de la tercera línea. Se sabe que la probabilidad de que, diariamente, un autobús tenga desperfectos mecánicos es del 2%, 4% y 1%, respectivamente, para cada línea. Determine la probabilidad de que, un día cualquiera, un autobús tenga desperfectos mecánicos. R/0.025

7. Los alumnos de Bachillerato de un Instituto Nacional proceden de 3 localidades A, B y C, siendo un 20 % de A, un 30 % de B y el resto de C. El 80 % de los alumnos de A cursa 1º de Bachillerato y el resto 2º.El 50 % de los alumnos de B cursa 1º de Bachillerato y el resto 2º. El 60 % de los alumnos de C cursa 1º de Bachillerato y el resto 2º.

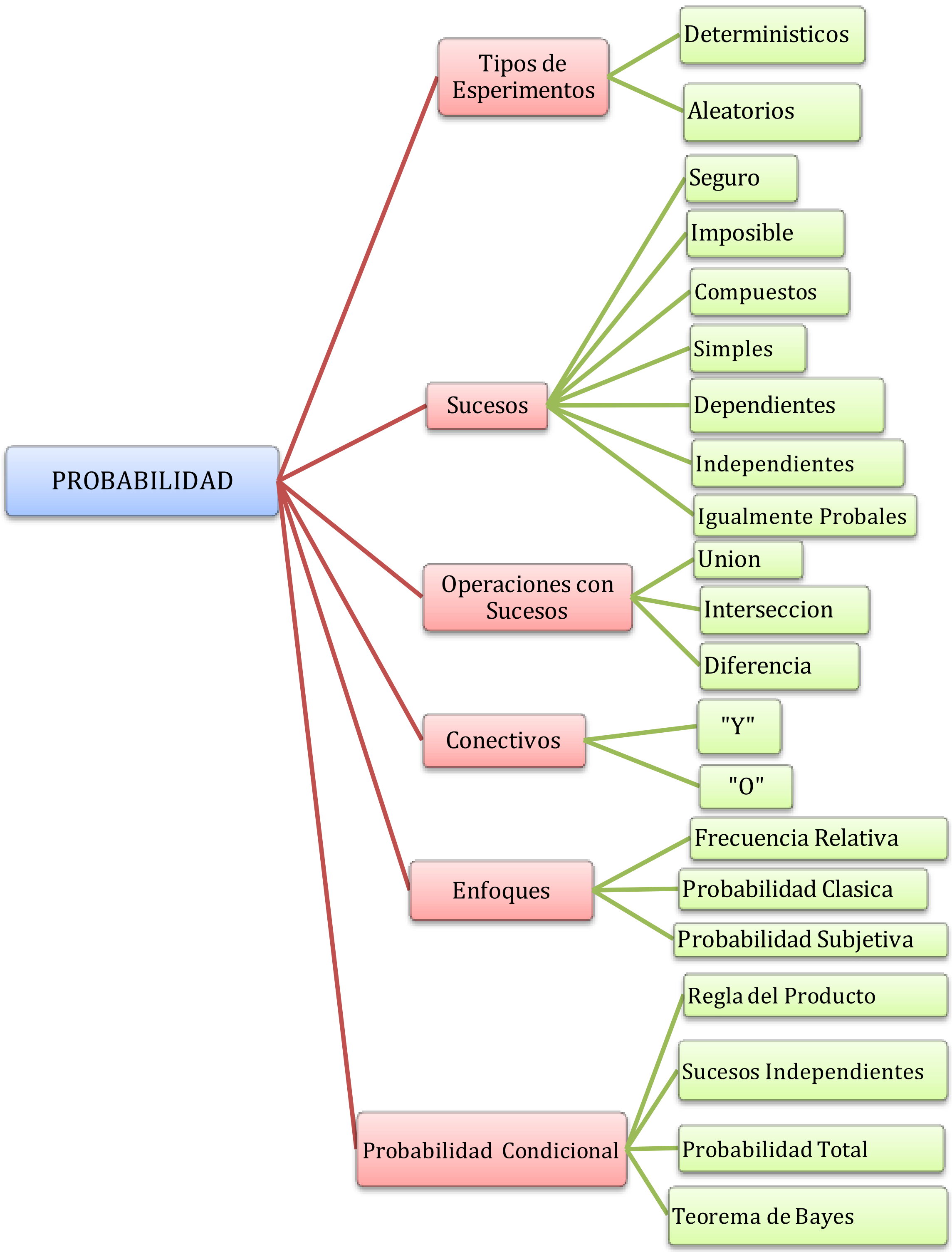
   a. Seleccionado, al azar, un alumno de Bachillerato de ese Instituto Nacional, ¿Cuál es la probabilidad de que sea de 2º ? R/0.39

   b. Si se elige, al azar, un alumno de Bachillerato de ese Instituto Nacional y éste es un alumno de 1º, ¿Cuál es la probabilidad de que proceda de la localidad B? R/0.246

BIBLIOGRAFIA

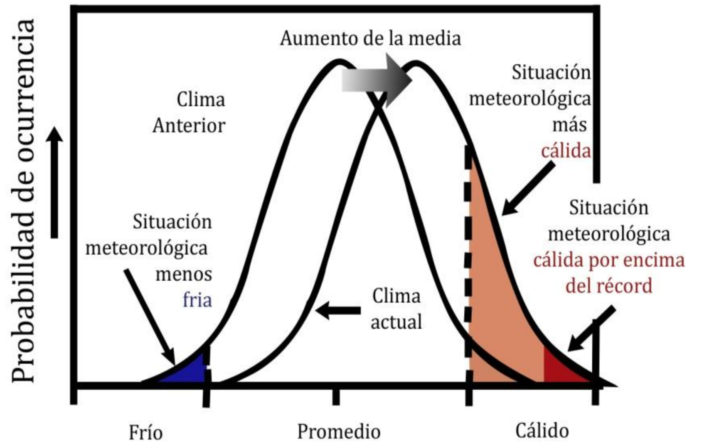
1. Johnson R., Cuby P.(1999), *Estadística Elemental*. México:Internacional Thomson Editores, S.A de S.V .
2. Pérez C., (2003).*Estadística. Problemas Resueltos y Aplicaciones*. Madrid: Pearson Educación, S.A.
3. Sarabia J.M. (2000), *Curso Práctico de Estadística*.(2da ed.) . España: Impreso por Gráficas Rogar, S.A Navalcarnero (Madrid).
4. Triola, M.,(2009). *Estadística.* (10a ed.). México: Pearson Educación.
5. Pérez-T. H.E((2007), Estadística para las Ciencias Sociales, del Comportamiento y de la Salud. (3ª. Ed).México: Impreso Edamsa Impresiones, S.A. de C.V.
6. Vladimir Moreno/Mauricio Restrepo(2001). Nuevo ALFA 10, serie de Matemática con énfasis en competencias (2da ed). Bogotá: Editorial Norma.

DIAGRAMA DE CONTENIDOS



Lección 10 **Segundo año de Bachillerato** **Unidad V** **Tiempo: 8 horas clase**

Distribuciones de Probabilidad



Gráfica 1. Distribución normal de la temperatura.

Introducción

Uno de los conceptos más importantes de la teoría de probabilidades es el de variable aleatoria que, intuitivamente, puede definirse como cualquier característica medible que toma diferentes valores con probabilidades determinadas. Toda variable aleatoria posee una distribución de probabilidad que describe su comportamiento (vale decir, que desagrega el 1 a lo largo de los valores posibles de la variable).

Si la variable es discreta, es decir, si toma valores aislados dentro de un intervalo, su distribución de probabilidad especifica todos los valores posibles de la variable junto con la probabilidad de que cada uno ocurra. En el caso continuo, es decir, cuando la variable puede tomar cualquier valor de un intervalo, la distribución de probabilidad permite determinar las probabilidades correspondientes con subintervalos de valores. Una forma usual de describir la distribución de probabilidad de una variable aleatoria es mediante la denominada función de densidad, y lo que se conoce como función de distribución representa las probabilidades acumuladas.

Objetivos

- Distinguir entre distribuciones de probabilidad discreta y continua.
- Describir las características de la distribución Binomial y aplicarla en casos prácticos.
- Describir las características de la distribución normal y aplicarla en casos prácticos.
- Manejar las tablas de la distribución Normal.

Importancia

La inferencia estadística y la obtención de conclusiones relativas a las características de una población a partir de los datos de una muestra obtenida de la misma.

La interrelación de los modelos probabilísticos y estadísticos queda patente en estos casos, pues vemos que en ellos se realizan análisis estadísticos de los datos de las muestras, y las conclusiones siempre se expresan en términos de probabilidad. Por lo tanto, cuando se trabaja los conceptos de distribuciones de probabilidad, se está proponiendo un modelo que puede ser aplicable o no a una situación aleatoria cualquiera, y no una simple colección de fórmulas y técnicas. Como modelo discreto se trabaja la distribución Binomial, y como modelo continúo la distribución Normal. Las distribuciones normales son sumamente importantes puesto que ocurren con gran frecuencia en las aplicaciones reales y porque desempeñan un papel fundamental en los métodos de estadística inferencial.

Competencias a reforzar.

- Identificación de distribuciones continuas y discretas.
- Destreza en el uso de la fórmula de la distribución binomial.
- Habilidad en el uso de las tablas de la distribución normal.

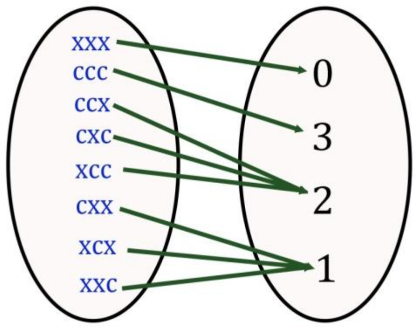
Presaberes

- Diagrama de árbol.
- Concepto de Factorial de un número.
- Diagrama de Venn.
- Teoría de conjuntos.
- Potenciación.
- Elaboración de gráficos.

VARIABLES ALEATORIAS

En gran cantidad de experimentos aleatorios es necesario cuantificar los resultados, es decir, asignar a cada resultado del experimento un número, con el fin de poder realizar un estudio matemático.

Actividad Introductoria: El experimento consiste en lanzar tres monedas al aire, supongamos que a cada elemento de su espacio muestral E= {ccc, ccx, cxc, xcc, cxx, xcx, xxc, xxx} se le asigna un número real, el correspondiente al número de caras.



Gráfica 2. Diagrama de Venn.

Esta correspondencia que se acaba de construir es una función del espacio muestral E en el conjunto de los números reales (R). A esta función se le llama variable aleatoria y se denota por lo general por X.

Ejemplos:

- En el experimento aleatorio que consiste en lanzar dos dados, se puede asignar a cada resultado la suma de los valores obtenidos en cada dado.
- Considere el experimento que consiste en elegir al azar 500 personas y medir su estatura. La ley que asocia a cada persona con su estatura es una variable aleatoria.

Una variable aleatoria es aquella que asume un valor numérico único para cada uno de los resultados que aparecen en el espacio muestral de un experimento aleatorio; es decir, es la transformación del espacio muestral en un conjunto numérico. Una variable es aleatoria si su valor está determinado por el azar. En gran número de experimentos aleatorios es necesario, para su tratamiento matemático, cuantificar los resultados de modo que se asigne un número real a cada uno de los resultados posibles del experimento. Usualmente se representa por las últimas letras del alfabeto: X, Y o Z, y para los valores concretos de cada uno de ellas se designa las respectivas letras minúsculas x,y, z. Es decir, se representan mediante letras mayúsculas y pueden tomar n posibles valores: $\mathbf{X} = \{ x_1, x_2, \dots, x_i, \dots, x_n \}$

Matemáticamente una variable aleatoria X es una función cuyo dominio es la colección de eventos del espacio muestral S y cuyo rango R_x, es un subconjunto de los números reales.

Ejemplo 1:

- Del experimento de lanzar tres monedas el espacio muestral es: E = {ccc, ccx, cxc, xcc, cxx, xcx, xxc, xxx}, si lo que interesa es conocer la cantidad de caras que pueden aparecer, se define entonces la variable aleatoria X: Número de caras que aparecen, siendo su dominio de definición: X={0,1,2,3}.
- De una caja que contiene 5 bolas numeradas del 1 al 5 se extraen 3 bolas una por una y sin reposición. Sea **X**: El mayor de los tres números sacados, es una variable aleatoria.

El espacio muestral es: E = {(1,2,3), (1,2,4), (1,2,5),(1,3,4),(1,3,5),(1,4,5), (2,3,4),(2,3,5), (2,4,5), (3,4,5)} y la variable aleatoria X asume los valores: 3, 4 y 5. Es decir, X={3,4,5}.

- Consideramos el experimento de lanzar un dado equilibrado dos veces. Sea X = Suma de las dos tiradas. El espacio muestral es {(1, 1), (1, 2), . . . , (6, 6)}, y para cada suceso elemental, se puede calcular el valor de X. Por ejemplo, si el resultado del experimento es (3, 4) luego X = 7. Los valores que toma son: X={2,3,4,5,6,7,8,9,10,11,12}

La mayor parte de las variables aleatorias se pueden expresar numéricamente, y por tanto son clasificadas como discretas y continuas.

Una ***variable aleatoria discreta*** tiene un número finito de valores o un número de valores contables, donde "contable" se refiere al hecho que podría haber un número infinito de valores, pero que puede asociarse con un proceso de conteo. Es decir toma valores enteros o un número finito de valores o infinito numerable.

Ejemplo 2:

- X: Número de "caras" al lanzar 3 monedas (puede tomar sólo los valores: 0,1,2,3).
- Y: Número de llamadas diarias que se hacen por teléfono móvil (puede tomar los valores 0,1,2,3,… infinito numerable).
- Z: Número de enfermos que se reciben cada día en un determinado hospital.

Si el rango de valores R_x de la variable aleatoria X es finito o infinito enumerable entonces se dice que es una **variable aleatoria discreta.**

Una ***variable aleatoria continua*** tiene un número infinito de valores, y estos valores pueden asociarse con mediciones en una escala continua, de manera que no existan huecos o interrupciones. Es decir, teóricamente, puede tomar todos los valores de un intervalo de **R**.

Ejemplo 3:

- X: Estatura de una población (en centímetros). Puede tomar cualquier valor en el intervalo [0,250] .
- Y: Tiempo máximo que he estado hablando por teléfono alguna vez (en minutos). Puede tomar cualquier valor en el intervalo [0,+∞).
- Z: Temperatura ambiente en San Salvador un determinado día.

Si su rango de valores R_x es infinito no enumerable entonces se dice que es una **variable aleatoria continua.**

DISTRIBUCION DE PROBABILIDAD DE UNA VARIABLE ALEATORIA DISCRETA

Como para una variable aleatoria es imposible saber con exactitud qué valor tomará en un momento dado, para describir el comportamiento de las mismas se recurre al uso de las probabilidades.

Actividad introductoria: Retomando el experimento del lanzamiento de dos dados. Se tiene la tabla de resultados.

X	Sucesos elementales					
2	(1,1)					
3	(1,2)	(2,1)				
4	(1,3)	(2,2)	(3,1)			
5	(1,4)	(2,3)	(3,2)	(4,1)		
6	(1,5)	(2,4)	(3,3)	(4,2)	(5,1)	
7	(1,6)	(2,5)	(3,4)	(4,3)	(5,2)	(6,1)
8	(2,6)	(3,5)	(4,4)	(5,3)	(6,2)	
9	(3,6)	(4,5)	(5,4)	(6,3)		
10	(4,6)	(5,5)	(6,4)			
11	(5,6)	(6,5)				
12	(6,6)					

Tabla 1: Resultados del lanzamiento de dos dados

Y como X es la variable de la suma de los valores observada de las caras de los dados. Se puede calcular la probabilidad de que la suma sea igual a 8, contando todos los resultados donde la suma es ocho. El evento en que la suma es ocho contiene 5 resultados:(2,6), (3,5), (4,4), (5,3), (6,2); y se tienen 36 resultados; por lo tanto, la probabilidad deseada es

$\dfrac{Casos\ favorables}{Casos\ posibles} = \dfrac{5}{36}$. Se puede repetir este proceso con cada uno de los resultados para obtener la siguiente tabla.

Resultado	2	3	4	5	6	7	8	9	10	11	12
Probabilidad P(X=x)	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
	0.027	0.055	0.083	0.111	0.138	0.166	0.138	0.111	0.083	0.055	0.027

Tabla 2: Distribución de probabilidad del experimento de lanzamiento de dos dados

Con esta tabla se ha encontrado la distribución de probabilidad de los valores posibles de la suma al tirar dos dados.

La distribución de probabilidad puede representarse gráficamente. Sin importar cuál sea la representación gráfica que se utilice, los valores de la variable aleatoria se trazan en el eje horizontal y la probabilidad asociada con cada valor de la variable aleatoria se traza en el eje vertical. La distribución de probabilidad de una variable aleatoria discreta puede representarse por medio de un conjunto de un segmento de recta trazados en los valores de X y cuyas longitudes representan probabilidades de cada valor de X. Sin embargo, para representar las distribuciones de probabilidad se usa con mayor frecuencia un histograma normal; en el cual se usa el área de cada barra para representar la probabilidad asignada.

Gráfica 3: Histograma de la distribución de probabilidad

La probabilidad de observar un valor particular de la variable aleatoria, digamos $x = 3$ esta dado por la altura de la línea sobre el valor de 3, es decir $P(x = 3) = \frac{2}{36} = 0.055$. De igual manera, en vez de asociar la altura de la barra con la probabilidad, se puede ver que el área de la barra sobre el 3 es $\frac{2}{36} \times 1 = \frac{2}{36}$, ya que la altura de la barra es $\frac{2}{36}$ y su ancho es 1. Usar el área de las barras para representar la probabilidad es muy útil para extender la noción de probabilidad a otras variables. Se puede utilizar el histograma de probabilidades para calcular probabilidades tal como: $P(X \leq 4)$.

$P(X \leq 4) = P(X = 2 \text{ ó } X = 3 \text{ ó } X = 4)$
$P(X \leq 4) = P(X = 2) + P(X = 3) + P(X = 4)$

Ya que los eventos X=2, X=3 y X=4 son disjuntos. Entonces:

$$P(X \leq 4) = \frac{1}{36} + \frac{2}{36} + \frac{3}{36} = \frac{6}{36}$$

Sumando las áreas de las barras que están sobre el 4 y a su izquierda. Se debe ser muy cuidadoso con las desigualdades, ya que $P(X \leq 4) = \frac{6}{36}$, mientras que $P(X < 4) = \frac{3}{36}$

En estadística la **distribución de probabilidad** para una variable aleatoria discreta X es una tabla, gráfica o fórmula que da la probabilidad P(X=x) asociada a cada posible valor de X.

$$P:\Re \rightarrow [0,1]$$
$$x \rightarrow P(x) = P(X = x)$$

La distribución de probabilidad debe cumplir:

- $0 \leq$ cada $\mathrm{P(X=x}_i)$ ≤ 1: Cada valor de probabilidad debe ubicarse entre 0 y 1, inclusive.
- $\sum_{\text{toda x}} \mathrm{P(X)=1}$: la suma de las probabilidades asignadas a cada uno de los valores de la variable aleatoria debe ser igual a 1.
- $P(X \leq x) = \sum_{\mathrm{i,\, x_i \leq x}} \mathrm{P(X=x}_i)$
- $\mathrm{P(X>x)=1\text{-}P(X \leq x)}$

Si un experimento con espacio muestral E, tiene asociada la variable aleatoria X, es natural que se planteen preguntas como: ¿Cuál es la probabilidad de que X tome un determinado valor?, esto nos lleva a establecer, por convenio, la siguiente notación:

- (X=*x*): representa el suceso "la variable aleatoria X toma el valor *x*", y p(X=*x*) representa la probabilidad de dicho suceso.
- (X<*x*): representa el suceso "la variable aleatoria X toma un valor menor a *x*", y p(X<*x*) representa la probabilidad de que la variable aleatoria X tome un valor menor a *x*.
- (X ≤*x*): representa el suceso "la variable aleatoria X toma un valor menor o igual a *x*", y p(X ≤*x*) representa la probabilidad de que la variable aleatoria X tome un valor menor o igual a *x*.

Ejemplo 4: Si se está interesado en el experimento del lanzamiento de los dos dados, en las siguientes probabilidades.

a) La suma sea menor o igual que 5

b) La suma esté entre 6 y 8 inclusive.

c) La suma sea mayor de 3.

Solución a)

$$P(X \leq 5) = P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5)$$
$$P(X \leq 5) = \frac{1}{36} + \frac{2}{36} + \frac{3}{36} + \frac{4}{36} = \frac{10}{36}$$

Solución b)

$$P(6 \leq X \leq 8) = P(X = 6) + P(X = 7) + P(X = 8)$$
$$P(6 \leq X \leq 8) = \frac{5}{36} + \frac{6}{36} + \frac{5}{36} = \frac{16}{36}$$

Solución c)

$$P(X > 3) = 1 - P(X \leq 3) = 1 - [P(X = 2) + P(X = 3)]$$
$$P(X > 3) = 1 - \left[\frac{1}{36} + \frac{2}{36}\right] = 1 - \frac{3}{36} = \frac{33}{36}$$

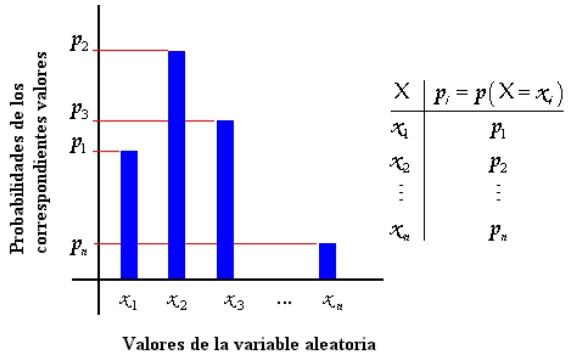
FUNCIÓN DE PROBABILIDAD DE UNA VARIABLE ALEATORIA DISCRETA

Cuando se conocen características o se efectúan estudios sobre el comportamiento de una variable, se puede desarrollar un modelo que brinde una descripción probabilística de la variable, el cual tendrá además implícito un grupo de condiciones que debe cumplir la variable.

Algunas veces es conveniente escribir una regla que exprese algebraicamente la probabilidad de un evento en términos del valor de la variable aleatoria. Esta expresión suele escribirse como una fórmula y se denomina **función de probabilidad**.

Una vez definida una variable aleatoria *X*, se puede definir una función de probabilidad asociada a ella, de la siguiente forma:}

$$f:\Re \rightarrow \Re$$
$$x \rightarrow f(x) = P(X = x)$$



Grafica 4: Representación de la función de probabilidad

La representación gráfica más usual de la función de probabilidad.

En estadística la función de probabilidad para una variable aleatoria discreta X es una tabla, gráfica o fórmula que da la probabilidad $P(X=x)$ asociada a cada posible valor de X. Una función de probabilidad puede ser tan sencilla como una lista, apareando los valores de una variable aleatoria con sus probabilidades. No obstante, una función de probabilidad se expresa más a menudo como una fórmula.

La función de probabilidad debe cumplir:

- $0 \leq$ cada $f(x_i)$ ≤ 1: Cada valor de probabilidad debe ubicarse entre 0 y 1, inclusive.
- $\sum_{\text{toda x}} f(x_i) = 1$: La suma de las probabilidades asignadas a cada uno de los valores de la variable aleatoria debe ser igual a 1.

Ejemplo 5: Considérese un dado que ha sido modificado de modo que tiene una cara con un punto, dos caras con dos puntos y tres caras con tres puntos. Sea X el número de valores que se observan cuando se lanza el dado. La distribución de probabilidad viene dada por:

x	1	2	3
P(X)	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{3}{6}$

Se puede observar que cada una de las probabilidades puede representarse por el valor de x dividido entre 6. Es decir, cada P(X) es igual al valor de x dividido entre 6, donde X=1,2 o 3. Así, $f(X) = \frac{X}{6}$ para X=1,2 o 3; esta expresión como fórmula representa la función de probabilidad de este experimento.

MEDIA, VARIANZA Y DESVIACIÓN ESTANDAR

MEDIA

La media da información acerca de la tendencia central de los datos, y se denota por μ. La media es el valor promedio ponderado en el que los valores posibles de la variable aleatoria se ponderan según las probabilidades correspondientes de ocurrencia, también se denomina **valor esperado** y se simboliza por E(X). Y se obtiene como:

$\mu = E(X) = \sum_{i=1}^{n} xP(X)$, donde P(X) es la probabilidad de valores posibles de la variable aleatoria X. Es decir, se multiplica cada valor de x por la probabilidad de que ocurra, y luego se suman estos valores.

VARIANZA

La varianza describe la dispersión de los datos que componen la distribución, se denota por σ^2. Y se obtiene como:

$\sigma^2 = \sum_{i=1}^{n}\left[(x-\mu)^2 P(X)\right]$ o de forma alternativa como

$$\sigma^2 = \sum_{i=1}^{n}\left[x^2 P(X)\right] - \mu^2$$

DESVIACIÓN ESTANDAR

La desviación estándar que se denota por σ, se obtiene al extraer la raíz cuadrada de la varianza. Es decir,

$$\sigma=\sqrt{\sum_{i=1}^{n}\left[x^2 P(X)\right]-\mu^2}$$

DISTRIBUCIÓN BINOMIAL

Actividad Introductoria:

Considérese el siguiente experimento de probabilidad. Un profesor aplica a sus estudiantes un examen sorpresa con cuatro preguntas de opción múltiple (3 opciones). Uno de los alumnos no ha estudiado y por tanto decide contestar las cuatro preguntas al azar, adivinando las respuestas sin leer las bases ni las opciones.

Antes de observar las respuestas correctas del examen y encontrar como le fue a este estudiante, se considerarán algunos hechos que podrían suceder si un examen se contesta de esta forma.

1) ¿Cuántas de las cuatro preguntas es probable contestar correctamente?

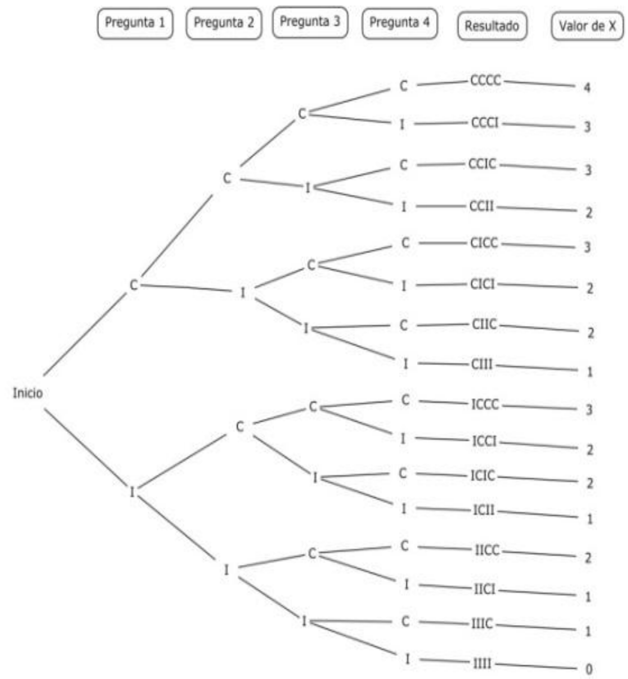
2) ¿Cuán probable es contestar de manera adecuada más de la mitad de las preguntas?

3) ¿Cuál es la probabilidad de elegir las respuestas acertadas de las cuatro preguntas?

4) ¿Cuál es la probabilidad de elegir respuestas incorrectas de las cuatro preguntas?

5) Si todo un grupo consta al azar un examen. ¿Cuál cree usted que será el número "promedio" de respuestas correctas del grupo?

Para contestar a estas interrogantes se comenzara con un diagrama de árbol que representa el espacio muestral, mostrando las 16 formas posibles que hay para contestar el examen de 4 preguntas. Donde cada una de ellas fue contestada C: correcta e I: incorrectamente.



Gráfica 5: Diagrama de árbol

La información del diagrama de árbol se transformara en una distribución de probabilidad. Sea X el "número de respuestas correctas" en el examen cuando este fue contestado al azar por algún estudiante.

La variable aleatoria X pude tomar cualquiera de los valores de 0, 1, 2, 3 o 4 para cada examen. En el diagrama de árbol se presentan las 16 ramas que representan los cinco valores de x. El evento x=4 se tienen "cuatro respuestas correctas", y está representado por la rama superior del árbol (primera), y el evento x=0, "cero respuestas correctas, se muestra en la rama inferior (última). Los otros eventos "una respuesta correcta", "dos respuestas correctas" y "tres respuestas correctas", se representan cada uno, por medio de varias ramas del árbol. Se puede observar que el evento x=1 ocurre en cuatro ramas distintas, el evento x=2 ocurre en seis ramas distintas y el evento x=3 ocurre en cuatro ramas distintas.

Ya que cada pregunta individual tiene una sola respuesta correcta de las tres posibilidades, la probabilidad de elegir la respuesta correcta de una pregunta individual es $\frac{1}{3}$. La probabilidad de elegir una respuesta incorrecta en cada pregunta es $\frac{2}{3}$. Las probabilidades de cada valor de X pueden encontrarse al calcular las probabilidades de todas las ramas y luego combinar las probabilidades de las ramas de valores de X semejantes. A continuación se presentan los cálculos.

- P(X=0): es la probabilidad de cero preguntas hayan sido contestadas correctamente y que cuatro haya sido contestadas incorrectamente (Resultado final de la rama IIII).

$$P(X=0) = \frac{2}{3}x\frac{2}{3}x\frac{2}{3}x\frac{2}{3}=\left(\frac{2}{3}\right)^4 = \frac{16}{81}\approx 0.198$$

- P(X=1): es la probabilidad de una pregunta hayan sido contestada correctamente y las otras tres haya sido contestadas incorrectamente (En el diagrama hay cuatro ramas en que ocurre esto, las cuales son: CIII, ICII, IICI, IIIC, y cada una tiene la misma probabilidad).

$$P(X=1)= (4)\frac{1}{3}x\frac{2}{3}x\frac{2}{3}x\frac{2}{3}= (4)\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)^3 = \frac{32}{81}\approx 0.395$$

- P(X=2): es la probabilidad de que dos preguntas hayan sido contestadas correctamente y las otras dos hayan sido contestadas incorrectamente (En el diagrama hay seis ramas en que ocurre esto, las cuales son: CCII, CICI, CIIC, ICCI, ICIC, IICC, y cada una tiene la misma probabilidad).

$$P(X=2)=(6)\frac{1}{3}x\frac{1}{3}x\frac{2}{3}x\frac{2}{3}=(6)\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^2 = \frac{24}{81}\approx 0.296$$

- P(X=3): es la probabilidad de que tres preguntas hayan sido contestadas correctamente y la otra haya sido contestadas incorrectamente (En el diagrama hay cuatro ramas en que ocurre esto, las cuales son: IIIC, CCIC, CICC, ICCC y cada una tiene la misma probabilidad).

$$P(X=3)=(4)\frac{1}{3}x\frac{1}{3}x\frac{1}{3}x\frac{2}{3}=(4)\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^1 = \frac{8}{81}\approx 0.099$$

- P(X=4): es la probabilidad de que las cuatro preguntas hayan sido contestadas correctamente (En el diagrama hay una sola rama en la que las cuatro son correctas, la cual es: CCCC).

$$P(X=4)=\frac{1}{3}x\frac{1}{3}x\frac{1}{3}x\frac{1}{3}=\left(\frac{1}{3}\right)^4 = \frac{1}{81}\approx 0.012$$

Obteniéndose la siguiente distribución de probabilidad.

X	0	1	2	3	4
P(X)	0.198	0.395	0.296	0.099	0.012

Las respuestas a las preguntas plateadas al inicio de la actividad; ya es posible contestarlas:

1) ¿Cuántas de las cuatro preguntas es probable contestar correctamente? La ocurrencia más probable es obtener una respuesta correcta; ya que su probabilidad es de 0.395
2) ¿Cuán probable es contestar de manera adecuada más de la mitad de las preguntas? Tener mas de la mitad de las respuestas correctas se representa con X=3 ó 4; su probabilidad total es de 0.111. (Es decir, este examen se aprueba solo el 11% de las veces, contestándolo al azar).
3) ¿Cuál es la probabilidad de elegir las respuestas acertadas de las cuatro preguntas?: P(las 4 respuestas son correctas)=

P(X=4)=0.012, es decir, todas las respuestas son correctas sólo el 1% de las veces.

4) ¿Cuál es la probabilidad de elegir respuestas incorrectas de las cuatro preguntas?: P(las 4 respuestas son incorrectas)= P(X=0)=0.198, es decir, casi el 20% de las veces.
5) Si todo un grupo consta al azar un examen. ¿Cuál cree usted que será el número "promedio" de respuestas correctas del grupo?. Puede esperarse que el promedio del grupo sea $\frac{1}{3}$ de 4 o 1.33, respuestas correctas.

En la realidad existen muchos experimentos que están compuestos de ensayos repetidos cuyos resultados pueden clasificarse en algunas de las siguientes categorías: existo y fracaso. Por ejemplo, se pueden mencionar:

- Lanzamiento de monedas, en donde solo hay dos resultados cara ó corona.
- Determinar si un producto cumple su función o no, dos respuestas: defectuoso o no defectuoso.
- Averiguar si las respuestas a una pregunta de un examen son correctas o incorrectas.
- Investigar sobre la calidad de un acusado en un juicio, culpable o inocente (no culpable).
- Comprobar el efecto de un fármaco en un paciente, efectivo o no efectivo.

Hay experimentos en que los ensayos tienen muchos resultados que, bajo las condiciones idóneas, puede satisfacer esta descripción general de ser clasificados en alguna de las categorías antes mencionadas. Por ejemplo, si sólo se esta interesado en saber si se obtiene un "uno" ó no, en el lanzamiento de un dado, entonces en realidad solo hay dos resultados: se obtiene un "uno" ó "cualquier otro número".

Los experimentos descritos anteriormente se denominan experimentos de probabilidad binomial; y tienen las siguientes propiedades:

- **Propiedad 1:** Hay n ensayos independientes repetidos.
- **Propiedad 2:** En cada prueba del experimento sólo son posibles dos resultados: el suceso A (éxito) y su contrario 1-A (fracaso).
- **Propiedad 3:** La probabilidad del suceso A es constante, se representa por p, y no varía de una prueba a otra. La probabilidad de fracaso es 1- p y se representa por q. Es decir, P(A)=p , P(1-A)= q=1-p, por lo tanto p+q=1.
- **Propiedad 4:** La variable aleatoria binomial, X, es el conteo del número de ensayos con éxito que ocurren; X puede asumir cualquier valor entero de cero a n. Dicho de otra manera es el número de éxitos que puedo obtener al realizar n ensayos.

Todo experimento que tenga estas características se dice que sigue el modelo de la ***distribución Binomial***. A la variable X que expresa el número de éxitos obtenidos en "n" ensayos o pruebas, se le llama ***variable aleatoria binomial***.

En el experimento del examen de las cuatro preguntas, que son cuatro ensayos cuando todas las respuestas se obtienen al azar se denomina experimento binomial; ya que cumple:

Propiedad 1: Un ensayo consiste en contestar una pregunta, repetido n=4 veces. Los ensayos son independientes, ya que la respuesta correcta en cualquier pregunta no es afectada por las respuestas en las otras preguntas.

Propiedad 2: Hay dos resultados posibles en cada ensayo: éxito=C, suceso A: "respuesta correcta" y fracaso=I, suceso B: "respuesta incorrecta".

Propiedad 3: En cada ensayo p= P(correcta)= $\frac{1}{3}$ y q=P(fracaso)= $\frac{2}{3}$, entonces p + q=1.

Propiedad 4: Variable X: número de respuestas correctas en el examen (experimento total y pude ser cualquier valor entero de 0 a 4).

En este tipo de experimentos es importante considerar los ensayos independientes; significa que el resultado de un ensayo no afecta la probabilidad de éxito de cualquier otro en el experimento. En otras palabras, la probabilidad de "éxito" permanece constante a lo largo de todo el experimento.

Ejemplo: Considérese el experimento de lanzar 12 veces un dado y obtener un "uno" o "cualquier otro número". Luego de haber realizado los 12 lanzamiento se reporta el número de "unos" obtenidos. La variable aleatoria X seria el número de veces que se observa un "uno' en los n=12 ensayos. Como el resultado que interesa es "uno", se considera como "éxito"; en consecuencia, p=P(uno)= $\frac{1}{6}$ y q=P(2,3,4,5,6)= $\frac{5}{6}$. Este experimento es binomial pues cumple las cuatro propiedades.

La clave para trabajar con cualquier experimento de probabilidad es su distribución de probabilidad. Todos los experimentos binomiales tiene las mismas propiedades y para representarse cualquiera de ellos puede usarse el mismo esquema de organización.

DEDUCCIÓN DE LA FUNCIÓN DE PROBABILIDAD BINOMIAL

En el ejemplo del examen con cuatro preguntas de tres opciones de respuesta cada uno, se ha comprobado que es experimento binomial y se obtuvieron los siguientes resultados para los diferentes valores que pude tomar X.

$$P(X=0)=\left(\frac{2}{3}\right)^4 \qquad P(X=1)=(4)\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)^3$$

$$P(X=2)=(6)\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^2 \qquad P(X=3)=(4)\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^1$$

$$P(X=4)=\left(\frac{1}{3}\right)^4$$

Por lo tanto, se tiene que n=4, q= 1/3, q=2/3 y x=0,1,2,3 o 4, obsérvese la regularidad o relación que tienen los resultados cuando X=1, 2 ó 3 en relación a los valores de n, p, q, x. Los valores enteros que contienen representan el número de combinaciones o formas que en cuatro ensayos puedan ocurrir exactamente x=1, 2 ó 3 éxitos. Que pueden ser representados por $\binom{4}{1}$, $\binom{4}{2}$, $\binom{4}{3}$ respectivamente.

En el caso de x=0 y x=4 el valor entero es 1, sin embargo; se puede representar por: $\binom{4}{0}\binom{4}{4}$ respectivamente.

En términos de p, q y se tiene:

$$P(X=1)=\binom{4}{1}(p)^1(q)^{3=4-1}, \quad P(X=2)=\binom{4}{2}(p)^2(q)^{2=4-2},$$

$$P(X=3)=\binom{4}{3}(p)^3(q)^{3=4-3}.$$

Al generalizar se tiene:

$$P(X=x)=\binom{4}{x}(p)^x(q)^{4-x}$$

que representa la función de probabilidad binomial.

La variable binomial es una variable aleatoria discreta, sólo puede tomar los valores 0, 1, 2, 3, 4,...,n suponiendo que se han realizado n ensayos. Como hay que considerar todas las maneras posibles de obtener k-éxitos y (n-k) fracasos debemos calcular éstas por combinaciones (número combinatorio n sobre k).

La **distribución Binomial** se suele representar por *B(n,p),* siendo *n* y *p* los parámetros de dicha distribución. Es decir, si X: "número de éxitos en las n pruebas" sigue una distribución binomial de parámetros n y p, entonces en este caso se escribe X~B(n,p).

Ejemplo: Al contestar al azar un examen tipo test, formado por 8 preguntas con 5 posibles repuestas cada una, si se estudia el nº de aciertos se tendrá una experiencia binomial de 8 experimentos independientes, con dos posibilidades (acierto o error) con probabilidades respectivas $\frac{1}{5}$ y $\frac{4}{5}$, que son constantes en cada experimento. Si la variable X mide el nº de aciertos, seguirá la distribución B(8, $\frac{1}{5}$).

Definición: Sea X una variable aleatoria discreta, se dice que se distribuye como una distribución binomial de parámetros (n,p). Siempre se debe de verificar que n>1 y que p tome valores entre 0 y 1. La probabilidad de obtener k éxitos en las n repeticiones viene dada por la expresión:

$$P(X=k)=\binom{n}{k}(p)^{k}(q)^{n-k} \quad ,\forall k=0,1,2,\ldots,n$$

Donde:

n: número de ensayos

k: número de éxitos en los n ensayos

p: probabilidad de éxito en cualquier ensayo

q: probabilidad de fracaso en cualquier ensayo (q=1-p)

Esta función de distribución en realidad está compuesta por el producto de tres términos, los que representan:

$\binom{n}{k}$: El número de formas en que en n ensayos pueden ocurrir exactamente k éxitos.

$(p)^k$: La probabilidad de k éxitos.

$(q)^{n-k}$: La probabilidad de que en los (n-k) ensayos restantes ocurrirá fracaso.

El término $\binom{n}{k}$ que siempre será un número entero positivo. Este término se denomina coeficiente binomial y se encuentra aplicando la fórmula: $\binom{n}{k}=\frac{n!}{(n-k)!k!}$

Media y Desviación Estándar de la Distribución Binomial

La media y la desviación estándar de una distribución de probabilidad binomial pueden encontrarse aplicando las dos fórmulas siguientes:

Media: $\mu=n*p$ Varianza: $\sigma=\sqrt{npq}$

Ejemplo 6: Un agente de seguros vende pólizas a 5 individuos, todos de la misma edad. De acuerdo con las tablas actuariales, la probabilidad de que un individuo con esa edad viva 30 años más es de $\frac{3}{5}$. Determinar la probabilidad de que dentro de 30 años vivan:

a) Los cinco individuos. b) Al menos tres.

c) Sólo dos. d) Al menos uno

Solución

La variable aleatoria en estudio X: número de individuos que viven dentro de 30 años.

La variable aleatoria X tiene sigue una distribución binomial de parámetros n=5 y p=0.6; ya que dentro de 30 años se pueden presentar dos situaciones: que la persona viva (p=$\frac{3}{5}$) o que haya muerto ($q=\frac{2}{5}$). Es decir, X~B(5,0.6). Entonces para k=0, 1, 2, 3, 4 y 5, se tiene la siguiente función de probabilidad:

$$P(X=k)=\binom{5}{k}(0.6)^k(0.4)^{5-k}\quad ,\forall k=0,1,2,.3,4,5$$

Solución: a) Probabilidad de que dentro de 30 años vivan los cinco individuos.

Se necesitan k=5 éxitos; por tanto se debe calcular P(X=5).

$$P(X=5)=\binom{5}{5}(0.6)^5(0.4)^{5-5}=\frac{5!}{(5-5)!5!}*(0.6)^5*(0.4)^0$$

$$=\frac{5!}{0!5!}*(0.6)^5*1=1*(0.6)^5=(0.6)^5=0.07776$$

Solucion: b) Probabilidad de que dentro de 30 años vivan al menos tres individuos.

Se necesitan k≥3 éxitos; por tanto se debe calcular P(X≥3).

$$P(X\geq 3)=1-P(X<3)=1-\left[P(X=0)+P(X=1)+P(X=2)\right]$$

$$=1-\left[\binom{5}{0}(0.6)^0(0.4)^{5-0}+\binom{5}{1}(0.6)^1(0.4)^{5-1}+\binom{5}{2}(0.6)^2(0.4)^{5-2}\right]$$

$$=1-\left[1*1*(0.4)^5+5*(0.6)(0.4)^4+\binom{5}{2}(0.6)^2(0.4)^3\right]$$

$$=1-\left[0.01024+0.0768+0.2304\right]$$

$$=1-0.31744=0.68256$$

Solución: c) Probabilidad de que dentro de 30 años vivan solo dos individuos.

Se necesitan k=2 éxitos; por tanto se debe calcular P(X=2).

$$P(X=2)=\binom{5}{2}(0.6)^2(0.4)^{5-2}$$

$$=\frac{5!}{(5-2)!2!}(0.6)^2(0.4)^3$$

$$=\frac{5!}{3!2!}*(0.6)^2(0.4)^3$$

$$=10*(0.6)^2(0.4)^3$$

$$=(10)(0.36)(0.064)=0.2304$$

Solución: d) Probabilidad de que dentro de 30 años vivan al menos un individuo.

Se necesitan k≥1 éxitos; por tanto se debe calcular P(X≥1).

$$P(X\geq 1)=1-P(X<1)=1-P(X=0)+=1-0.01024=0.98976$$

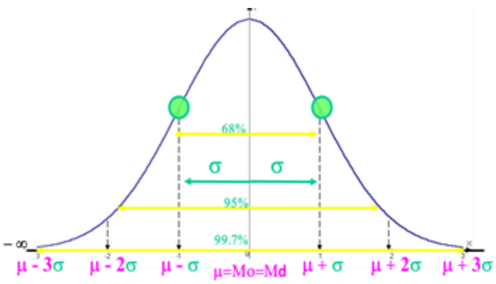
DISTRIBUCIÓN NORMAL

La Normal es, sin duda la distribución de probabilidad más importante del cálculo de probabilidades y de la Estadística; ya que multitud de variables aleatorias continuas siguen una distribución normal o aproximadamente normal; y se le denomina con el nombre de campana de Gauss, pues al representar su función de probabilidad, ésta tiene forma de campana, con campo de variación de]-α, α[.

La distribución de probabilidad normal y la curva normal que la representa, tienen las siguientes características:

- La curva normal tiene forma de campana invertida y un sólo pico en el centro de la distribución. De esta manera, la media aritmética, la mediana y la moda de la distribución son iguales y se localizan en el pico. Así, la mitad del área bajo la curva se encuentra a la derecha de este punto central y la otra mitad está a la izquierda de dicho punto.

- La distribución de probabilidad normal es simétrica alrededor de su media y la desviación típica es la que determina el recorrido de la misma.
- La curva normal desciende suavemente en ambas direcciones a partir del valor central. Es asintótica, lo que quiere decir que la curva se acerca cada vez más al eje X pero jamás llega a tocarlo. Es decir, las "colas" de la curva se extienden de manera indefinida en ambas direcciones.
- Los puntos de inflexión tienen como abscisas los valores $\mu \pm \sigma$.



Gráfica 6: Distribución normal

Para distribuciones de tipo discreto, la suma de todos los valores de la probabilidad debía ser 1. Para el caso de las distribuciones de tipo continuo esta condición se transforma en que el área total bajo la curva ha de ser 1. La clave de este tipo de distribuciones está en que existe una correspondencia entre área y probabilidad, de forma que la probabilidad de que la variable esté entre dos valores a y b es exactamente el área entre a y b.

La distribución normal queda especificada por dos parámetros de los que depende su función de distribución y que resultan ser la media y la desviación típica o estándar de la distribución.

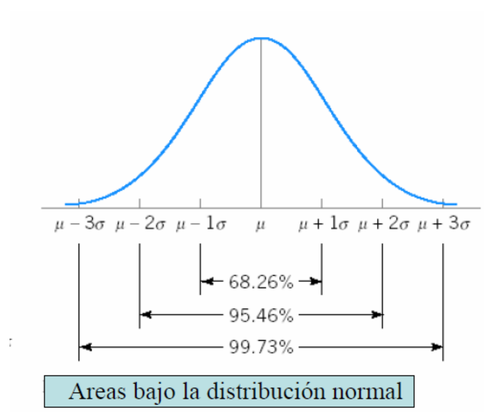
Para indicar que una variable aleatoria X sigue una distribución normal de media μ y desviación estándar σ, se representa mediante la expresión: $X \sim N(\mu,\sigma)$.

REGLA EMPÍRICA

Si una variable está distribuida normalmente, entonces: a menos de una desviación estándar de la media hay aproximadamente 68% de los datos ($\mu \pm \sigma$); a menos de dos desviaciones estándar de la media hay aproximadamente 95% de los datos ($\mu \pm 2\sigma$); y a menos de tres desviaciones estándar de la media hay aproximadamente el 99.7% de los datos ($\mu \pm 3\sigma$).

Por lo tanto, para una distribución normal, la mayor parte de todos los valores yacen a tres desviaciones estándar de la media.

Si calculada la media μ y la desviación típica σ de los datos, se cumple aproximadamente estos porcentajes se puede considerar que el conjunto de datos se ajusta a una distribución normal.



Gráfica 7: Representación de la regla empírica

DEFINICIÓN DE FUNCIÓN DE DENSIDAD DE PROBABILIDAD

Si una variable aleatoria X continua, se dice que se distribuye como una normal $X \sim N(\mu,\sigma)$; $\mu \in R$ y $\sigma > 0$, donde se verifica que $-\alpha < X < +\alpha$, μ, es el valor medio de la distribución y precisamente donde se sitúa el centro de la curva (campana de Gauss), y σ es cualquier valor entre -α y

+ α, si su función de distribución viene dada por:

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

Donde:

f(x): Denota la función de distribución

x : valores observados de la variable en estudio

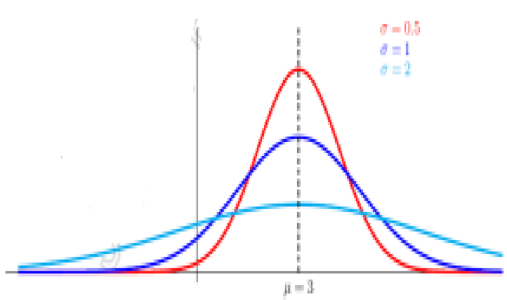
σ^2: varianza (parámetro de la distribución)

μ: media (parámetro de la distribución)

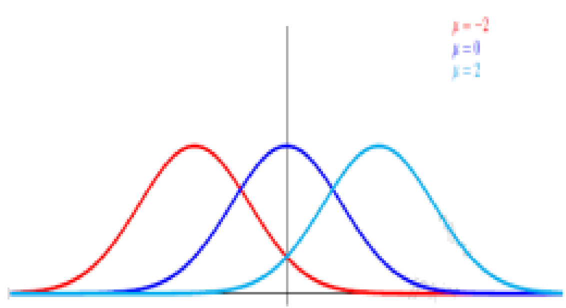
e= 2.71828 (base del Logaritmo natural)

π : 3.1416

Teniendo en cuenta la fórmula presentada, la distribución normal puede adoptar diferentes formas, tantas como distintos valores de μ y σ se consideren (o sea, infinitas). Cada uno de estos posibles modelos integra la familia de la distribución normal. Como se observa en las siguientes graficas:



Gráfica 8: Valores de μ = 3 y σ = 0.5, 1, 2

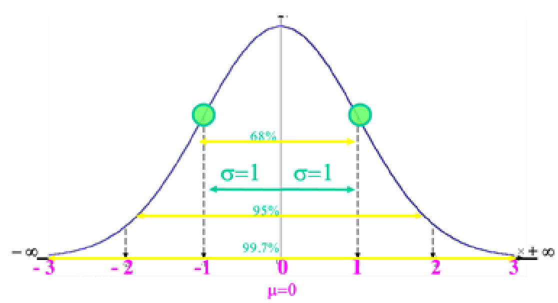


Gráfica 9: Valores de μ = -2, 0, 2 y σ =1

Como puedes observar, la media indica el eje de simetría de la distribución, mientras la desviación típica es la que determina el recorrido de la misma.

DISTRIBUCIÓN NORMAL ESTÁNDAR

Se observó que no existe una sola distribución de probabilidad normal, sino una "familia" de ellas. Por tanto, el número de distribuciones normales es ilimitado y sería imposible proporcionar una tabla de probabilidades para cada combinación de μ y σ. Para resolver este problema, se utiliza un solo "miembro" de la familia de distribuciones normales, aquella cuya media es 0 y desviación estándar 1 que es la que se conoce como **distribución estándar normal**, de forma que todas las distribuciones normales pueden convertirse a la estándar, restando la media de cada observación y dividiendo por la desviación estándar. La grafica en estas condiciones toma la siguiente forma:



Gráfica 10: Distribución normal estándar

Si en la expresión de la función de distribución general de la normal se sustituye μ =0 yσ=1; se obtiene la siguiente función de distribución:

$f(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}$. Esta es la distribución normal de la variable aleatoria Z; denominada "puntaje Z", "puntaje estándar" o "puntaje normal".

Propiedades

- El área total bajo la curva normal es igual a 1.

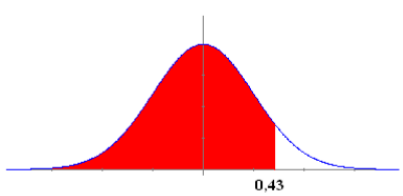
- La distribución tiene forma de montículo y es simétrica; se extiende indefinidamente en ambas direcciones, tendiendo al eje horizontal pero sin tocarlo.
- La distribución tiene una media de 0 y una desviación estándar de 1.
- La media divide el área a la mitad, 0.5 a cada lado.
- Casi toda el área está entre Z=3.00 y z= 3.00, es decir, casi entre estos valores está el 100% del área bajo la curva.

Las fórmulas anteriores no se usaran para calcular probabilidades de distribuciones normales; y en lugar de usar las fórmulas para encontrar las probabilidades de distribuciones normales, se usará una tabla. A menudo estas fórmulas aparecen como identificación en la parte superior de las tablas de probabilidad normales. Así, es común que en los libros de estadística se incluya en un apéndice final de tablas estadísticas, la correspondiente a la curva normal estándar. Aunque hay variaciones en la forma de presentar esta tabla en los libros, en la misma nos será posible consultar para un rango de valores comprendido habitualmente entre -3 y 3, el cuál es el valor de probabilidad y de probabilidad acumulada correspondiente a esos valores.

USO DE LAS TABLAS DE LA DISTRIBUCIÓN NORMAL.

Se dispone de una tabla de la N(0,1), en la que aparecen las probabilidades de la forma p(Z<k), siendo k positivo. Haciendo uso de la simetría de la distribución normal, y de que el área total es 1, se pueden calcular todos los casos que se presenten. La normal N(0;1) se encuentra tabulada, para valores a partir de 0 y hasta 3.9.

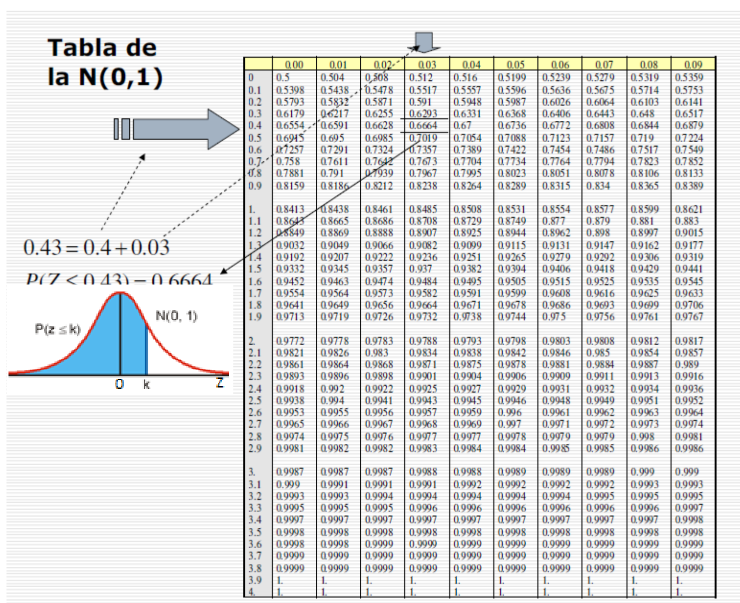
Es importante entender que la curva de Gauss es una curva de probabilidad acumulada. Esto quiere decir, que dado un valor de la variable (que se sitúa en el eje de las x), toda el área comprendida en la gráfica hasta ese punto es la probabilidad de que esa variable valga dicho valor o menos. Por ejemplo, si la variable aleatoria X tomase el valor 0.43, se ubica en la gráfica de la siguiente forma:



El área sombreada es la probabilidad que la variable X valga 0.43 o menos.

Si consultamos la tabla, nos indica que dicha probabilidad vale 0,6664. Expresado de manera matemática:

P (X< 0,43) = *0,6664.*



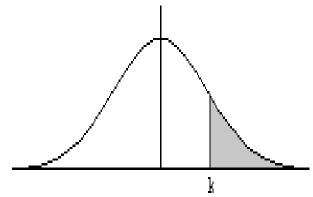
	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5	0.504	0.508	0.512	0.516	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.591	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.648	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.67	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.695	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.719	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.758	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.791	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.834	0.8365	0.8389
1.	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.877	0.879	0.881	0.883
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.898	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.937	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.975	0.9756	0.9761	0.9767
2.	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.983	0.9834	0.9838	0.9842	0.9846	0.985	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.989
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.992	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.994	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.996	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.997	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.998	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.999	0.999
3.1	0.999	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.
4.	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.

Tabla 3: Tabla de distribución normal estándar

La tabla indica probabilidades para distintos valores. Para leerla, se mira en la primera columna el número y su primer decimal, y en la primera fila, el segundo decimal. El punto donde se corten esos dos valores, nos da una probabilidad entre 0.5 y 1.

Al momento de las aplicación de las distribución normal se pueden presentar otros casos que requieren ciertos "retoques", ya que la tabla de probabilidades solo sirve para los casos en que Z valga igual o menor que un valor. Se pueden dar los siguientes casos:

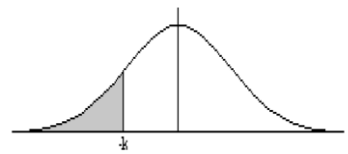
Caso 1: Valores de Z mayores o iguales a un valor P(Z≥k)



Si k es positivo y se quiere calcular $p(Z \geq k)$, es decir el área sombreada; basta pasar al complementario, es decir:

$P(Z \geq k) = 1 - p(Z \leq k)$ y esta última probabilidad ya se encuentra tabulada.

Caso 2: Valores de Z negativos menores o iguales a un valor P(Z≤-k)



Las probabilidades de valores negativos no están tabuladas.

Si k es positivo y se quiere calcular $p(Z \leq -k)$, es decir el área: por simetría, $P(Z \leq -k) = P(Z \geq k)$ y esta se calcula como en el caso anterior. Se puede observar la igualdad de áreas en la figura:

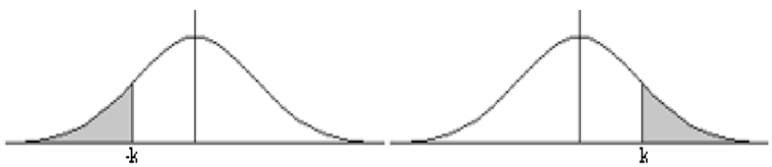
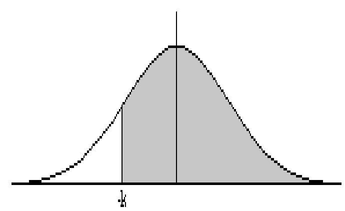


Figura 8: $P(Z \leq -k) = P(Z \geq k)$. La simetría permite reducir este caso al anterior

Caso 3: Valores de Z negativos mayores o iguales a un valor $P(Z \geq -k)$



Si k es positivo y queremos calcular $p(Z \geq -k)$, es decir el área sombreada; entonces, por simetría $p(Z \geq -k) = p(Z \leq k)$.

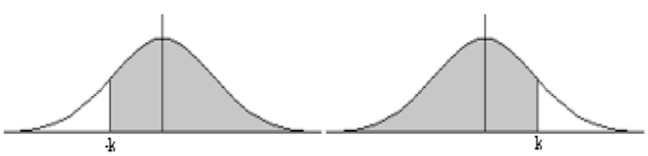
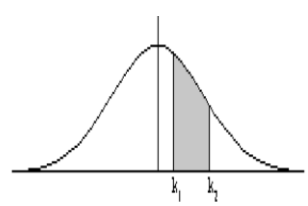


Figura 9: $P(Z \geq -k) = P(Z \leq k)$.La simetría permite reducir este caso al que ya esta tabulado

Que es el caso más simple que se describió al principio en el ejemplo, se obtiene directamente de la tabla.

Caso 4: Valores de Z comprendidos entre dos valores $P(k_1 \leq Z \leq k_2)$



Probabilidades comprendidas entre dos valores, $p(k_1 \leq Z \leq k_2)$,es decir el área sombreada; en este casos se tiene que calcular dos áreas de probabilidad, y restar a la mayor la menor. Es decir:

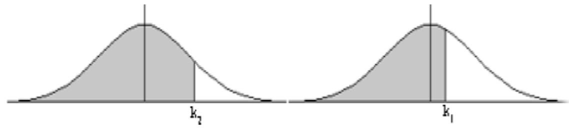


Figura 10: $p(Z \leq k_2)$ en la primera imagen. $p(Z \leq k_1)$ en la segunda. Al restar obtenemos el área pedida.

$$P(k_1 \leq Z \leq k_2) = P(Z \leq k_2) - P(Z \leq k_1)$$

Caso 5: Valores de Z concretos $P(Z=k)$

En este tipo de distribuciones no tiene sentido plantearse probabilidades del tipo p(Z=k), ya que siempre valen 0, al no encerrar ningún aérea. Este tipo de distribuciones en las cuales la probabilidad de tomar un valor concreto es 0 se denominan distribuciones continuas, para diferenciarlas de otras en las que esto no ocurre, como por ejemplo la binomial, que es una distribución discreta.

Caso 6: Valores de Z muy grandes

¿Qué pasa si el valor que nos piden de Z es tan alto que no aparece en la tabla? . En estos casos, la probabilidad vale 1. Observar que los últimos valores de probabilidad de la tabla valen prácticamente uno. Se supone que valores de Z más altos valen directamente uno.

Caso 7: Probabilidad y encontrar Z

También puede ser que el dato que nos den sea la probabilidad, y nos pregunten para que valor de Z se corresponde esa probabilidad en la curva de Gauss. Para resolver estos casos, se busca en la tabla el valor de probabilidad más cercano al que nos dan; se observa la fila y columna correspondiente para formar el valor de Z que le corresponden. Es importante saber si la probabilidad que nos dan como dato esta a la izquierda o la derecha de la línea del cero. Si es mayor que 0.5, se resuelve como se acaba de decir. Si es menor de 0.5, estaría a la izquierda:

APLICACIONES DE LA DISTRIBUCIÓN NORMAL

La tabla de la normal, como hemos dicho, corresponde a una variable cuya media es igual a cero y su desviación típica igual a 1. Sin embargo, la gran mayoría de las variables del mundo real no son de este tipo. Como no es posible crear una tabla para cada distribución posible, lo que se hace es modificar, “tipificar” nuestra variable real para adaptarla y poder usar con ella la tabla de N(0,1). Es decir, convertir la distribución real en una distribución normal estándar utilizando un valor llamado Z, o estadístico Z que será la distancia entre un valor seleccionado, designado X, y la media μ, dividida por la desviación estándar σ.

Formalmente, si $X \sim N(\mu,\sigma)$, entonces la variable aleatoria $Z=\frac{X-\mu}{\sigma}$ se distribuya según una normal de media 0 y desviación estándar 1. Es decir, Z ~ N(0,1), que es la distribución llamada normal estándar o tipificada. Y por lo tanto:

$$P(X \leq x) = \left[\frac{X-\mu}{\sigma} < \frac{x-\mu}{\sigma}\right] = P\left[Z < \frac{x-\mu}{\sigma}\right]$$

De esta manera, un valor Z mide la distancia entre un valor especificado de X y la media aritmética, en las unidades de la desviación estándar. Al determinar el valor Z utilizando la expresión anterior, es posible encontrar el área de probabilidad bajo cualquier curva normal haciendo referencia a la distribución normal estándar en las tablas correspondientes.

Ejemplo: El llenado de las cajas de talco en la fábrica de una empresa de perfumería se hace automatizadamente, de forma que el peso neto de las cajas se distribuye normalmente, siendo el peso promedio de 15 onzas con una desviación típica de 0,8 onzas.

a) ¿Qué probabilidad hay que una caja tenga un peso neto inferior a 13 onzas?

b) ¿Qué proporción de las cajas tendrá pesos netos superiores a 16 onzas?

c) ¿Qué proporción de las cajas tendrá pesos netos entre 15 y 16 onzas?

d) ¿Cuál es el peso máximo del 20% de las cajas menos pesadas?

e) ¿Cuál es el peso mínimo del 10% de las cajas más pesadas?

Solución.

Datos

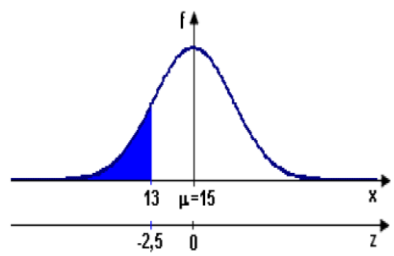
μ=15 onzas y σ= 0.8 onzas **Variable** X: Peso neto de las cajas de talco.

$X \sim N(15,0.8)$ $Z=\frac{X-\mu}{\sigma}$ $Z \sim N(0,1)$

Solución a) ¿Qué probabilidad hay de que una caja tenga un peso neto inferior a 13 onzas?

$$P(X<13)=P\left(\frac{X-\mu}{\sigma}<\frac{13-15}{0.8}\right)=P\left(Z<\frac{-2}{0.8}\right)=P(Z<-2.50)$$

$$=P(Z>2.50)=1-P(Z<2.50)=1-0.9938=0.0062$$

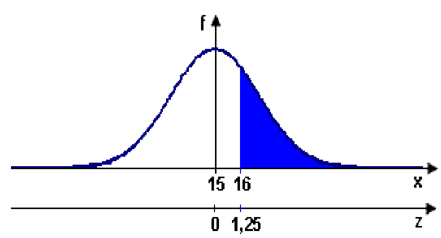


La probabilidad de que una caja tenga un peso neto inferior a 13 onzas es: 0.0062

Solución b) ¿Qué proporción de las cajas tendrá pesos netos superiores a 16 onzas?

$$P(X>16)=P\left(\frac{X-\mu}{\sigma}>\frac{16-15}{0.8}\right)=P\left(Z>\frac{1}{0.8}\right)=P(Z>1.25)$$

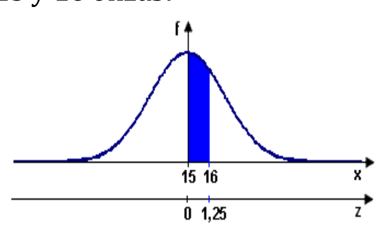
$$=1-P(Z<1.25)=1-0.8944=0.1056$$



El 10,6% de las cajas tendrá pesos netos mayores de 16 onzas.

Solución c) ¿Qué proporción de las cajas tendrá pesos netos entre 15 y 16 onzas?

$$P(15<X<16)=P\left(\frac{15-15}{0.8}<\frac{X-\mu}{\sigma}<\frac{16-15}{0.8}\right)=P\left(0<Z<\frac{1}{0.8}\right)$$

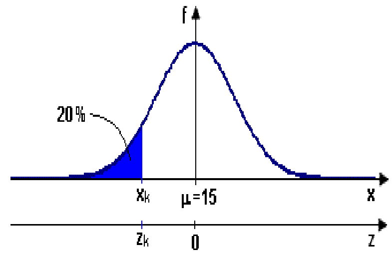


$= P(0 < Z < 1.25)$=P(Z<1.25)-P(Z<0)=0.8944-0.500 =0.3944

El 39,4% de las cajas tendrán pesos netos entre 15 y 16 onzas.

Solución d) ¿Cuál es el peso máximo del 20% de las cajas menos pesadas?

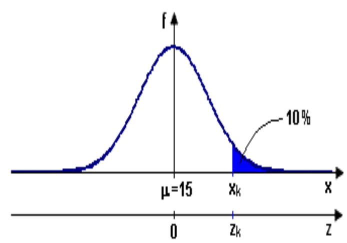
Para resolver esto lo primero es ubicar las cajas menos pesadas, que son aquellas ubicadas en la cola o extremo izquierdo de la curva. De ellas interesan las que representan el 20% del total, y se quiere determinar el peso (X_k) que acota superiormente a ese 20% de cajas; por tanto, puede plantearse que:$P(X < X_k) = 0,20$. Entonces, de la misma manera se tiene que: $P(Z < Zk) = 0,20$ Y una forma de representar ese valor Z_k es: $Z_k = z_{0,20}$. Con esto se quiere decir que es el valor de de una variable Z que ha acumulado un 20% de probabilidad.



Encontrar mediante la tabla el valor de Z que acumula un 20% de probabilidad implica buscar en el interior de la misma el número más cercano a 0,20 (que es 0,2005), y de su encabezado de fila y columna se llega a que: $Z_k = Z_{0,20} = -0,84$. Conocido el valor Z_k se puede hallar X_k, despejando de: $Z=\frac{X-\mu}{\sigma}$, $Xk = Zk\ \sigma + \mu = -0,84*0,8 + 15 = 15 - 0,672 = 14.328$. Se concluye, pues, que el peso máximo para el 20% de las cajas menos pesadas es de 14.328 onzas.

Solución e) ¿Cuál es el peso mínimo del 10% de las cajas más pesadas?

Ahora interesan las cajas más pesadas, que son las ubicadas en la cola o extremo derecho de la curva, y de ellas importa las que representan el 10% del total. O sea, se quiere determinar el peso (X_k) que acota inferiormente a ese 10% de cajas; y puede plantearse que: $P(X > Xk) = 0.10$.



Así, se tiene también que: $P(Z > Z_k) = 0,10$. Pero esto no constituye un valor de probabilidad acumulada, pues la probabilidad acumulada es la que está por debajo del punto, y para Z_k sería, haciendo uso de la regla del complemento: $P(Z < Z_k) = 1 - 0.10 = 0.90$ ó $Z_k = Z_{1-0,10} = Z_{0,90}$.

Buscando en la tabla el valor de Z que acumula un 90% de probabilidad se encuentra que el valor más cercano a 0.90 en el interior de la misma es 0.8997, y de su encabezado de fila y columna se llega a que: $Z_k = Z_{0.90} = 1.28$.

Y despejando X_k: $X_k = Z_k *\sigma + \mu = 1.28 * 0.8 + 15 = 15 + 1,024 = 1,.024$

Se concluye, pues, que el peso mínimo para el 10% de las cajas más pesadas es de 16,024 onzas.

APLICANDO LO APRENDIDO

1) Se lanza un dado equilibrado produciendo el espacio equiprobable. E= {1,2,3,4,5,6}. Sea X el doble del número que aparece. Encuentre la distribución ƒ, la media, la varianza y la desviación estándar de X.
2) La distribución de probabilidad de una variable aleatoria discreta **x** viene dada por:

Xi	3	4	5	6	7
Pi	0.1	0.2	k	0.25	0.3

Calcular: a) Calcula el valor de **k** b) $P(x > 5)$ c) $P(x < 3)$ d) la media. e) La desviación típica.

3) Explica para cada una de estas situaciones si se trata de una distribución binomial. En caso afirmativo, identifica los valores de n y p.
a) El 2% de las naranjas que se empaquetan en un cierto lugar están estropeadas. Se empaquetan en bolsas de 10 naranjas cada una. Nos preguntamos por el número de naranjas estropeadas de una bolsa elegida al azar.
b) En una urna hay 2 bolas rojas, 3 blancas y 2 verdes. Sacamos una bola, anotamos su color y la devolvemos a la urna. Repetimos la experiencia 10 veces y estamos interesados en saber el número de bolas blancas que hemos extraído.
4) La probabilidad de que un alumno de 1º de Bachillerato repita curso es de 0,3. Elegimos 20 alumnos al azar. ¿Cuál es la probabilidad de que haya exactamente 4 alumnos repetidores? R/ 0.13
5) El 20 % de los tornillos de un gran lote son defectuosos. Se cogen tres tornillos al azar y se pide calcular razonadamente:
a) Defina claramente la variable aleatoria en estudio.
b) La probabilidad de que los tres sean defectuosos. R/0.008
c) La probabilidad de que ninguno sea defectuoso. R/0.512
d) La probabilidad de que solamente uno sea defectuoso. R/0.384
6) Se ha estudiado que 1/3 de los alumnos de Bachillerato no leen nunca la prensa diaria. Tomando una muestra al azar de 10 alumnos estudiar las probabilidades siguientes:
a) Defina claramente la variable aleatoria en estudio.
b) Encontrar dos alumnos que no leen la prensa. R/0.1951
c) Más de tres alumnos que no leen la prensa. R/0.4408
d) Por lo menos cinco alumnos que no leen la prensa. R/0.9235
7) Si un estudiante responde al azar a un examen de 8 preguntas de verdadero o falso:
a) Defina claramente la variable aleatoria en estudio.
b) ¿Cuál es la probabilidad de que acierte 4?
c) ¿Cuál es la probabilidad de que acierte dos o menos?
d) ¿Cuál es la probabilidad de que acierte cinco o más?
e) ¿Cuánto valen la media y la varianza del número de preguntas acertadas?
8) Un avión de alto rendimiento contienen tres computadoras idénticas. Se utiliza únicamente una para operar el avión; las dos restantes son repuestos que pueden activarse en caso de que el sistema primario falle. Durante una hora de operación la probabilidad de que una falle en la computadora primaria (o de cualquiera de los sistemas de repuesto activados) es 0,0005.

Suponiendo que cada hora representa un ensayo independiente:

a) ¿Cuál es el tiempo promedio para que fallen las tres computadoras?
b) ¿Cuál es la probabilidad de que las tres computadoras fallen en un vuelo de 5 horas?

9) Sabiendo que la variable Z sigue una distribución Normal cero, uno, calcule las siguientes Probabilidades:

$P(Z < 93)$, $P(Z > 1{,}68)$, $P(Z < -2{,}27)(Z > -0{,}27)$, $P(Z > 0{,}62)$, $P(Z > 2{,}05)$ $P(Z > -1{,}07)$, $P(Z > -3{,}39)$, $P(0{,}56 < Z < 2{,}80)$, $P(-2{,}81 < Z < -0{,}33)$ $P(-0{,}85 < Z < 0{,}72)$

10) En la distribución normal N(0,1) calcular el valor de k en los siguientes casos:

a) $P(Z \leq k) = 0.7673$ b) $P(Z \leq k) = 0.9761$

c) $P(Z \geq k) = 0.1075$

11) La media de los pesos de 500 estudiantes de un colegio es 70 kg y la desviación típica 3 kg. Suponiendo que los pesos se distribuyen normalmente, hallar cuántos estudiantes pesan:

a) Entre 60 kg y 75 kg. b) Más de 90 kg.
c) Menos de 64 kg. d) 64 kg. e) 64 kg ó menos.

12) El tiempo que una persona sana invierte en recorrer 10 km está normalmente distribuido con una media de 60 minutos y una desviación típica de 9 minutos.

a) Calcula la probabilidad de que una persona sana invierta menos de 50 minutos.
b) Calcula la probabilidad de que una persona sana invierta menos de 55 minutos o más de 65 minutos.
c) En una fiesta de animación al deporte participan 500 personas sanas. Calcula cuantas de ellas invertirán en hacer el recorrido entre 50 y 60 minutos.

13) En un estudio realizado por una empresa hotelera, la distribución del tiempo de estancia del viajero en el hotel fue normal, con una media de 3,7 días y una desviación típica de 1,1 días.

a) ¿Qué probabilidad habrá de que un viajero permanezca en el hotel entre 2 y 5 días?
b) De 500 viajeros, ¿Cuántos habrán permanecido entre 4 y 7 días?

BIBLIOGRAFÍA

Johnson R., Cuby P. (1999), *Estadística Elemental*. México: Internacional Thomson Editores, S.A de S.V.

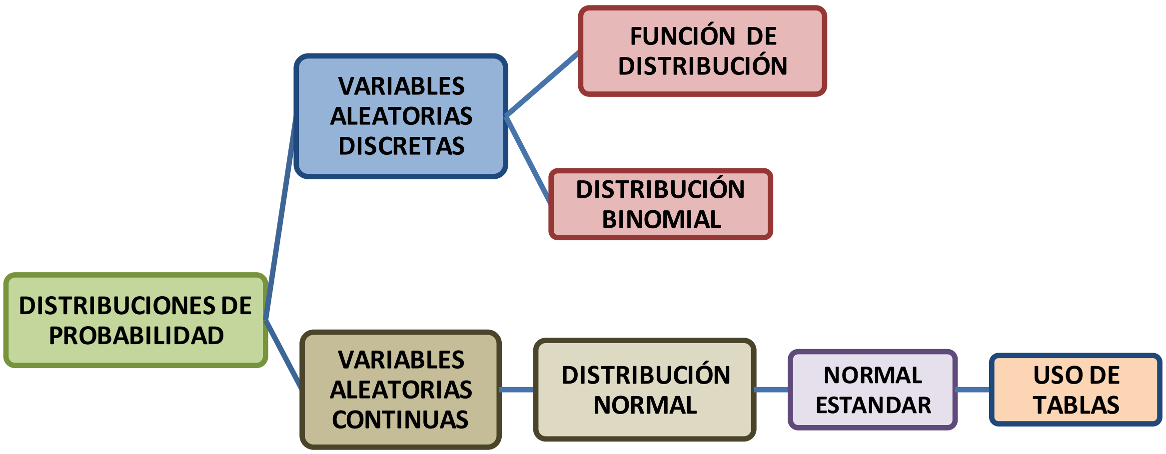
Pérez C., (2003).*Estadística. Problemas Resueltos y Aplicaciones*. Madrid: Pearson Educación, S.A.

Pérez-T. H.E ((2007), Estadística para las Ciencias Sociales, del Comportamiento y de la Salud. (3ª. Ed).México: Impreso Edamsa Impresiones, S.A. de C.V.

Sarabia J.M. (2000), *Curso Práctico de Estadística*. (2da ed.) . España: Impreso por Gráficas Rogar, S.A Navalcarnero (Madrid).

Triola, M., (2009). *Estadística.* (10a ed.). México: Pearson Educación.

DIAGRAMA DE CONTENIDOS



Viceministerio de Ciencia y Tecnología

Gerencia de Educación en Ciencia Tecnología e Innovación

Este material de Autoformación e Innovación Docente es un esfuerzo del Gobierno de El Salvador (Gestión 2009-2014) para desarrollar y potenciar la creatividad de todos los salvadoreños y salvadoreñas, desde una visión que contempla la Ciencia y la Tecnología de una manera "viva" en el currículo nacional, la visión CTI (Ciencia, Tecnología e Innovación)

Ministerio de Educación

Gobierno de El Salvador
Unir, crecer, incluir

Viceministerio Ciencia y Tecnología

GECTI
Gerencia de Educación en Ciencia, Tecnología e Innovación

ProEduca
Programa de Fortalecimiento de la Cohesión Social en el Sector Educativo de El Salvador