Primera Edición

FUNDAMENTOS BÁSICOS DE ESTADÍSTICA

CECILIA SALAZAR P.
SANTIAGO DEL CASTILLO G.

2018

FUNDAMENTOS BÁSICOS DE ESTADÍSTICA

PRIMERA EDICIÓN
2018

ISBN: 978-9942-30-616-6

QUI-052578

DEDICATORIA

A los estudiantes, quienes necesitan comprender los fundamentos estadísticos para relacionarlos con otras ciencias.

Cuando escucho… entiendo,

Cuando veo… comprendo,

Cuando hago… aprendo.

María Montessori

PRÓLOGO

El mayor problema que enfrenta un ser humano, es el de entenderse a sí mismo. En todos los campos la comprensión de conceptos y la abstracción de propiedades de los entes y fenómenos sujetos de estudio, son la actividad intelectual primordial que con el desarrollo de la ciencia y a partir de ésta, permite el desarrollo de la tecnología.

Los procesos complejos de la mente, por siglos estudiados, están gobernados por mecanismos que de alguna manera se logran formalizar en las ciencias cuánticas, como la Lógica Matemática, Estadística e Investigación operativa. Las cuales nos permiten comprender muchas herramientas intelectuales, como los razonamientos deductivos e inductivos, e incluso pensar acerca del propio pensamiento.

Por todo ello, iniciar los cursos de Estadística, permite a las jóvenes generaciones adquirir poderosas herramientas de análisis, descripción, y presentación de datos, los cuales les permitirán comprender los conceptos de qué es, y por qué se estudia estadística; que son el espíritu mismo del análisis de la estadística descriptiva e inferencial para establecer el enfoque de probabilidades a utilizarse en situaciones reales, para progreso propio y de la sociedad en general.

"Educar…Es dar al intelecto del estudiante toda belleza y perfección de la que ellos son susceptibles"

Sócrates

ÍNDICE GENERAL

ÍNDICE DE FIGURAS

ÍNDICE DE TABLAS

PREFACIO

Nuestra experiencia como docentes universitarios a través de los años, nos han permitido cubrir diferentes ramas de la Matemática, tales como Cálculo, Estadística, Optimización e Investigación Operativa. Este antecedente nos ha motivado a preparar un texto que se ajuste a los programas de estudio basados en una secuencia lógica entre los temas tratados y que sea de fácil asimilación para los estudiantes. Nuestra preocupación y ocupación es cubrir con didáctica los contenidos, tratando de llegar cuidadosamente a una enseñanza personalizada.

Este libro es para las personas que inician sus estudios de Estadística, les servirá para dar sus primeros pasos y a poder desarrollar la capacidad de autoformarse tomando en cuenta la actualización de los cambios científicos y tecnológicos.

Los Autores

INTRODUCCIÓN

Historia de la Estadística

La palabra Estadística procede del vocablo "Estado", pues era función principal de los Gobiernos de los Estados establecer registros de población, nacimientos, defunciones, impuestos, cosechas... La necesidad de poseer datos cifrados sobre la población y sus condiciones materiales de existencia han debido hacerse sentir desde que se establecieron sociedades humanas organizadas.

Es difícil conocer los orígenes de la Estadística. Desde los comienzos de la civilización han existido formas sencillas de estadística, pues ya se utilizaban representaciones gráficas y otros símbolos en pieles, rocas, palos de madera y paredes de cuevas para contar el número de personas, animales o ciertas cosas.

Su origen empieza posiblemente en la isla de Cerdeña, donde existen monumentos prehistóricos pertenecientes a los Nuragas, las primeros habitantes de la isla; estos monumentos constan de bloques de basalto superpuestos sin mortero y en cuyas paredes de encontraban grabados toscos signos que han sido interpretados con mucha verosimilidad como muescas que servían para llevar la cuenta del ganado y la caza.

Hacia el año 3.000 a.C. los babilonios usaban ya pequeñas tablillas de arcilla para recopilar datos en tablas sobre la producción agrícola y los géneros vendidos o cambiados mediante trueque.

Los egipcios ya analizaban los datos de la población y la renta del país mucho antes de construir las pirámides. En los antiguos monumentos egipcios se encontraron interesantes documentos en que demuestran la sabia organización y administración de este pueblo; ellos llevaban cuenta de los movimientos poblacionales y continuamente hacían censos. Tal era su dedicación por llevar siempre una relación de todo que hasta tenían a la diosa Safnkit, diosa de los libros y las cuentas.

En la Biblia observamos en uno de los libros del Pentateuco, bajo el nombre de Números, el censo que realizó Moisés después de la salida de Egipto. Textualmente dice: "Censo de las tribus: El día primero del segundo año después de la salida de Egipto, habló Yahvé a Moisés en el desierto de Sinaí en el tabernáculo de la reunión, diciendo: "Haz un censo general de toda la asamblea de los hijos de Israel, por familias y por linajes, describiendo por cabezas los nombres de todos los varones aptos para el servicio de armas en Israel". En el libro bíblico Crónicas describe el bienestar material de las diversas tribus judías.

En China existían los censos chinos ordenados por el emperador Tao hacia el año 2.200 a.C.

Posteriormente, hacia el año 500 a.C., se realizaron censos en Roma para conocer la población existente en aquel momento. Se erigió la figura del censor, cuya misión consistía en controlar el número de habitantes y su distribución por los distintos territorios.

En la Edad Media, en el año 762, Carlomagno ordenó la creación de un registro de todas sus propiedades, así como de los bienes de la iglesia.

Después de la conquista normanda de Inglaterra en 1.066, el rey Guillermo I, el Conquistador, elaboró un catastro que puede considerarse el primero de Europa.

Los Reyes Católicos ordenaron a Alonso de Quintanilla en 1.482 el recuento de fuegos (hogares) de las provincias de Castilla.

En 1.662 un mercader de lencería londinense, John Graunt, publicó un tratado con las observaciones políticas y naturales, donde Graunt pone de manifiesto las cifras brutas de nacimientos y defunciones ocurridas en Londres durante el periodo 1.604-1.661, así como las influencias que ejercían las causas naturales, sociales y políticas de dichos acontecimientos. Puede considerarse el primer trabajo estadístico serio sobre la población.

Curiosamente, Graunt no conocía los trabajos de B. Pascal » (1.623-1.662) ni de C. Huygens (1.629- 1.695) sobre estos mismos temas. Un poco más tarde, el astrónomo Edmund Halley (1.656- 1.742) presenta la primera tabla de mortalidad que se puede considerar como base de los estudios contemporáneos. En dicho trabajo se intenta establecer el precio de las anualidades a satisfacer a las compañías de seguros. Es decir, en Londres y en París se estaban construyendo, casi de manera simultánea, las dos disciplinas que actualmente llamamos estadística y probabilidad.

En el siglo XIX, la estadística entra en una nueva fase de su desarrollo con la generalización del método para estudiar fenómenos de las ciencias naturales y sociales. Galton » (1.822-1.911) y Pearson (1.857-1936) se pueden considerar como los padres de la estadística moderna, pues a ellos se debe el paso de la estadística deductiva a la estadística inductiva.

Los fundamentos de la estadística actual y muchos de los métodos de inferencia son debidos a R. A. Fisher. Se interesó primeramente por la eugenesia, lo que le conduce, siguiendo los pasos de Galton a la investigación estadística, sus trabajos culminan con la publicación de la obra Métodos estadísticos para investigaciones. En él aparece la metodología estadística tal y como hoy la conocemos.

A partir de mediados del siglo XX comienza lo que podemos denominar la estadística moderna, uno de los factores determinantes es la aparición y popularización de los computadores. El centro de gravedad de la metodología estadística se empieza a desplazar técnicas de computación intensiva aplicadas a grandes masas de datos, y se empieza a considerar el método estadístico como un proceso iterativo de búsqueda del modelo ideal.

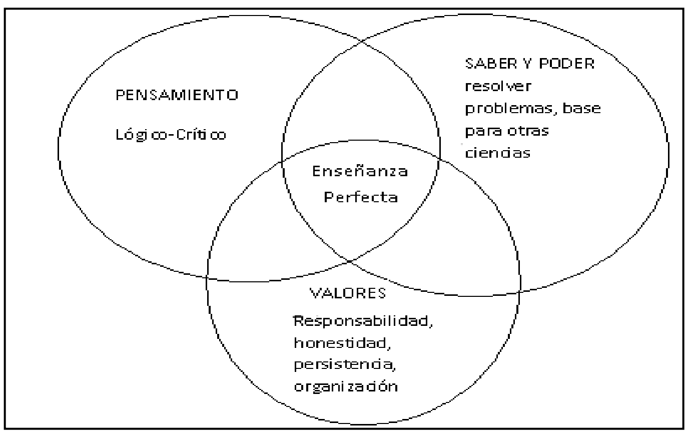
Las aplicaciones en este periodo de la Estadística a la Economía conducen a una disciplina con contenido propio: la Econometría. La investigación estadística en problemas militares durante la segunda guerra mundial y los nuevos métodos de programación matemática, dan lugar a la Investigación Operativa.

Presentación del texto

La presente obra, no pretende hacer un estudio exhaustivo de estadística, se presenta más bien como un resumen de lo estudiado en el nivel medio al que se le ha adicionado las aplicaciones más comunes dentro del ámbito económico y que privilegia la imprescindible e impostergable tarea de enseñar a recolectar, organizar y representar datos estadísticos, pues ya no es hora de la simple trasmisión de conocimientos al estudiante en este tiempo donde la creatividad es la rectora de todas las actividades; entonces una de las tareas fundamentales es tener la posibilidad de resolución de problemas probabilísticos y es precisamente esta función donde se fundamenta la enseñanza de la estadística.

Para la resolución de los problemas del mundo real no bastan únicamente conocimientos teóricos, más bien se debe recurrir a un pensamiento lógico-crítico que basado en un marco conceptual pueda deducir la solución. El proceso de enseñanza-aprendizaje en general, debe posibilitar a cada estudiante conseguir: persistencia, autoestima, honestidad, responsabilidad y pensamiento lógico- crítico. En especial para la Estadística, se puede señalar como objetivos los que se observan en la *Figura 1*.

Figura 1. Objetivos de la Estadística



Elaboración: Autores

En cada unidad se han resuelto y propuesto ejercicios a fin de que el estudiante pueda advertir y poner en práctica los conocimiento que vaya adquiriendo en la solución de problemas y casos que se presenta en la vida diaria. Al final de cada capítulo se presenta un resumen de los temas tratados que sirven de retroalimentación para una mayor fijación de los conocimientos y una autoevaluación, que si es realizada en forma responsable y adecuada, permite al estudiante observar los logros académicos.

Se ha considerado adicionalmente las facilidades que la tecnología actualmente pone a nuestra disposición por lo tanto en varias partes del texto se describe su utilización especialmente al uso adecuado de la hoja electrónica Excel en el uso principal de la funciones estadísticas y las herramientas de análisis de datos. También se proveerá la información sobre el uso de la calculadora científica en el e manejo de información en el modo estadístico y de regresión.

Importancia del estudio de la Estadística

Podríamos dedicar muchas páginas al desarrollo evolutivo de esta materia, pero ese no es el objetivo principal del texto. Para terminar con esta nota introductoria, basta con decir que a más del valor intrínseco que tiene el conocimiento de la materia, existen por lo menos cinco motivos importantes, debido a los cuales tanto estudiantes como profesionales necesitan del aporte de la estadística en su vida diaria.

1. Como herramienta de trabajo. En todas las ciencias, la estadística aporta sus métodos para sintetizar, representar y establecer conclusiones sobre el comportamiento de datos.

2. En la solución de problemas. En los procesos investigativos, el aporte que brinda la estadística es fundamental, para absolver las preguntas: ¿Cómo mejorar el ensayo? ¿Entre variables de estudio, existe alguna relación?

3. En la investigación teórica. Ayudan a la generación de teorías que permiten predecir el comportamiento bajo circunstancias determinadas, especialmente en circunstancias donde los eventos no están regidos por leyes físicas o determinísticas.

4. Utilización de la investigación. En todo ámbito ayudan a los profesionales a comprender la información que se genera en la investigación teórica o aplicada, toda vez que se genera cuantiosa información cuantitativa, la misma que es analizada a través de la teoría estadística.

5. Satisfacción personal. Al inicio, los estudiantes tienden a pensar que el proceso de la recolección de datos y su análisis no es muy ameno, posiblemente debido a que por desconocimiento se crea que los procesos son muy complejos; felizmente con la ayuda de la tecnología, todos los procesos sistemáticos y repetitivos son realizados casi instantáneamente. Al final con la obtención de los resultados y conclusiones que se pueden dar, se genera un ambiente de satisfacción.

El texto consta de las siguientes unidades:

Unidad I Análisis Descripción y Representación de datos: Incluyen el lenguaje y fórmulas de la estadística, así como operaciones estadísticas.

Unidad II Probabilidades y distribución normal: Permite a los estudiantes adquirir poderosas herramientas de cálculo de probabilidades, las mismas que les permitirán comprender con mayor claridad los conceptos, criterios y análisis.

UNIDAD 1

Análisis y descripción de representación de datos

Objetivos:

- Definir estadística y diferenciar entre estadística descriptiva e inferencial.
- Conocer los tipos de variables con los cuales se trabaja en estadística, los niveles de medición y los datos que se generan.
- Resumir la información obtenida atreves de la generación de tablas y gráficos estadísticos.
- Calcular e interpretar las principales medidas de centralización, como la media aritmética, mediana, moda y media geométrica.
- Calcular e interpretar las principales medidas de variabilidad y forma de un conjunto de datos.
- Escoger la medida más representativa de un estudio estadístico de acuerdo al tipo de variable utilizado.
- Comprender el teorema de Chebyshev y la regla normal o empírica y saber aplicarlos de acuerdo al conjunto de datos.

1.1. Estadística básica

Estadística: Es la ciencia que se encarga de la recolección, ordenamiento, representación, análisis e interpretación de datos generados en una investigación sobre hechos, individuos o grupos de los mismos, para deducir de ello conclusiones precisas o estimaciones futuras.

Población

Es el colectivo que abarca a todos los elementos cuya característica o características queremos estudiar; dicho de otra manera, es el conjunto entero al que se desea describir o del que se necesita establecer conclusiones. Como ejemplos de poblaciones, podemos citar: todos los estudiantes de la Universidad Central del Ecuador, o los artículos producidos en una semana en una determinada fábrica.

Por su tamaño, las poblaciones pueden ser finitas o infinitas.

Muestra

Es un conjunto de elementos seleccionados de una población de acuerdo a un plan de acción previamente establecido (muestreo), para obtener conclusiones que pueden ser extensivas hacia toda la población. Ejemplos constituyen las muestras que escogen las empresas encuestadoras en estudios de sondeos de opinión, o la selección de un grupo de artículos recibidos en una bodega para estimar las condiciones de todo un embarque.

Censo

Es el estudio de todos y cada uno de los elementos de una población. Esta condición hace que este tipo de estudios no sean muy frecuentes, por cuanto la recolección de toda esa información, sobre todo cuando el tamaño de la población es muy grande o sus elementos se encuentran muy dispersos, sea muy costosa. Ejemplo: último censo de población y vivienda que se realizó en Ecuador en noviembre de 2010.

Muestreo

Es la técnica que nos permite seleccionar muestras adecuadas de una población de estudio. El muestreo debe conducir a la obtención de una muestra representativa de la población de donde proviene, esta condición establece que cada elemento de la población tiene la misma probabilidad de ser incluida en la muestra. El estudio de selección de muestras, en sí constituye todo un estudio pormenorizado, que no atañe al estudio en este texto.

Parámetro

Es cualquier medida descriptiva o representativa de una población. Generalmente se utilizan las letras griegas como símbolo. Ejemplos: media aritmética poblacional (μ) *(mu)*, desviación estándar poblacional (σ) *(sigma)*.

Estadístico

Constituyen cualquiera de las medidas descriptivas de una muestra. Se las simboliza con letras minúsculas de nuestro alfabeto. Ejemplos: media aritmética (x), desviación estándar (s).

1.1.1. División de la Estadística

Básicamente la estadística se divide en dos grandes ramas: estadística descriptiva o matemática y estadística inferencial, estas dos divisiones se articulan adecuadamente mediante las probabilidades.

Estadística descriptiva

Es la parte de la estadística que permite analizar todo un conjunto de datos, de los cuales se extraen conclusiones valederas, únicamente para ese conjunto. Para realizar este análisis se procede a la recolección y representación de la información obtenida. Como ejemplo de estas estadísticas podemos citar a aquellas que se obtienen generalmente en los deportes, en los rendimientos académicos de los estudiantes de una determinada materia, en los negocios al determinar las ventas obtenidas mensualmente en un determinado año por una empresa en particular.

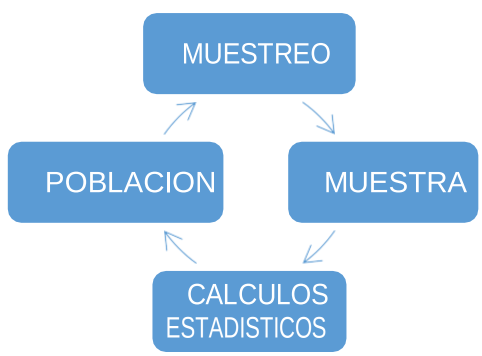
Estadística inferencial

En esta rama de la estadística, lo que se pretende es obtener conclusiones generales de una determinada población, mediante el estudio de una muestra representativa sacada de ella, dicho de otra manera, lo que se trata es que, con el valor de los estadísticos obtenidos, podamos establecer los valores de los parámetros. Entonces podemos concluir que la estadística inferencial analiza o investiga a una población, valiéndose de los datos y resultados que se obtienen de una muestra. Ejemplos muy claros de este tipo de estadística constituyen la aplicación de nuevos tratamientos con nuevos fármacos, o las proyecciones que pueden hacer los investigadores de mercado sobre cómo influye la publicidad en ciertos segmentos de mercado.

Esta condición vista anteriormente, ha permitido que la estadística inferencial, tenga un crecimiento cada vez mayor, por cuanto sus aplicaciones son cada vez más eficientes en el manejo de poblaciones; por tal motivo es que existen métodos muy variados para poder realizar la generalización de los resultados obtenidos en el muestreo. (Pruebas de hipótesis, predicciones futuras y más).

En la *Figura 2*, se puede visualizar el procedimiento que se sigue para realizar inferencia estadística.

Figura 2. Inferencia estadística

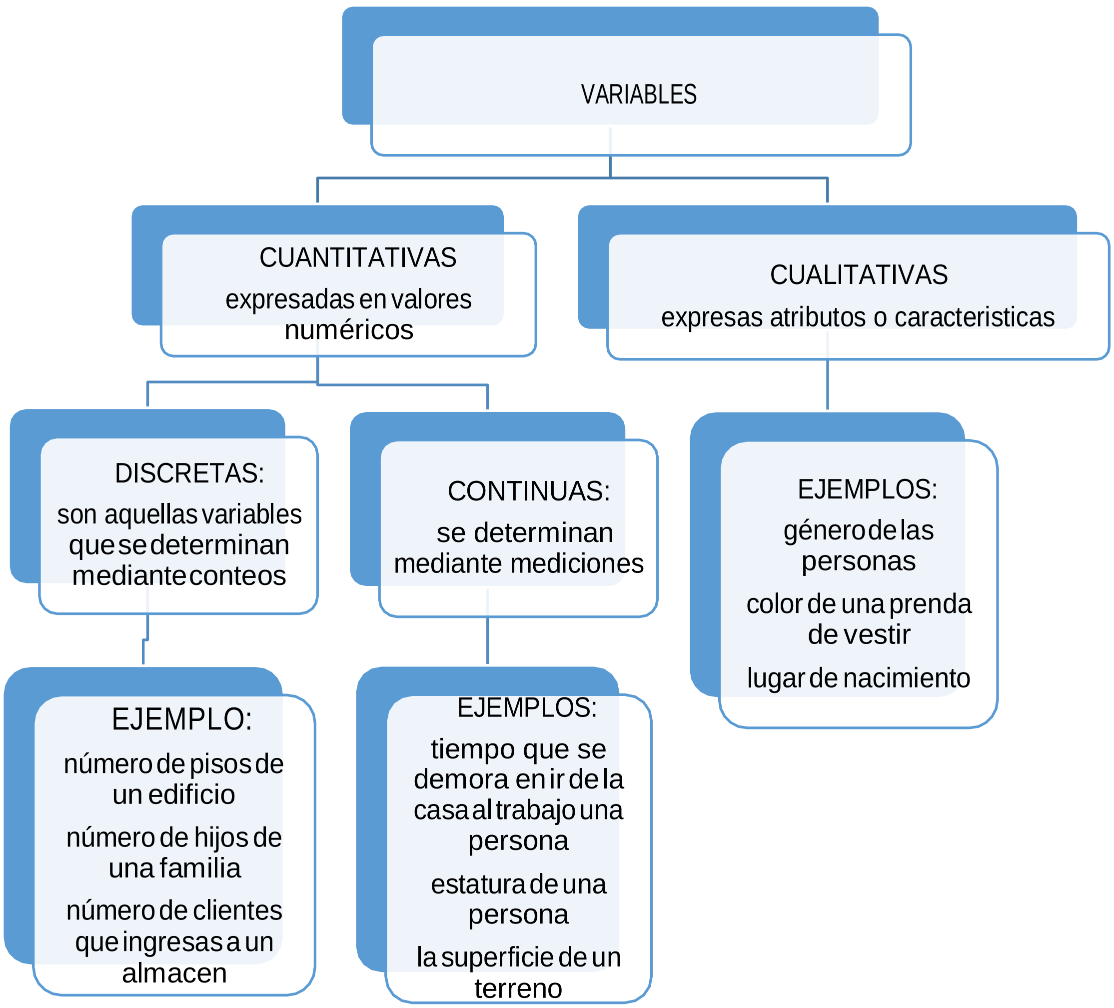


Elaboración: Autores

1.2. Variables de datos

La información que se obtiene de un estudio estadístico, proviene de variables, las mismas que están determinadas con el interés que se tenga sobre los elementos de Observación. Estas variables están categorizadas en dos grandes grupos, tal como se aprecia en la *Figura 3*.

Figura 3. Variables



Elaboración: Autores

1.2.1. Tipo de Variable

De tal manera los datos que se generan en un estudio estadístico, serán de la misma categorización de la variable que se está estudiando, por lo tanto, éstos pueden ser cualitativos o cuantitativos y a su vez discretos o continuos.

En la *Tabla 1*, se puede observar la categorización de los tipos de variables que se pueden determinar en un elemento de observación específico.

Ejemplo: Se desea observar a las casas unifamiliares del sector sur de la ciudad de Quito

Tabla 1. Ejemplo de variables

VARIABLES		
CUALITATIVAS	**CUANTITATIVAS DISCRETAS**	**CUANTITATIVAS CONTINUAS**
Barrio donde se halla ubicada Tipo de estructura disponibilidad de garaje Tipo de cubierta Material utilizado en las paredes Disponibilidad de los servicios básicos	Número de dormitorios Número de baños Número de pisos Antigüedad de la construcción	Área de construcción Área del terreno sobre el que está ubicada Área de patios y jardines exterior Longitud del terreno de la casa Altura de la casa

Elaboración: Autores

En la *Tabla 1*, constan algunas de las variables que pueden ser observadas en el elemento de observación propuesto, pueden determinarse muchas más dependiendo del interés de la investigación.

Variable cualitativa

Las variables cualitativas se refieren a características o cualidades que no pueden ser medidas con números. Podemos distinguir dos tipos:

- **Variable cualitativa nominal**

Una variable cualitativa nominal presenta modalidades no numéricas que no admiten un criterio de orden. Por ejemplo:

El estado civil, con las siguientes modalidades: soltero, casado, separado, divorciado y viudo.

- **Variable cualitativa ordinal o variable cuasi cuantitativa**

Una variable cualitativa ordinal presenta modalidades no numéricas, en las que existe un orden. Por ejemplo:

- La nota en un examen: suspenso, aprobado, notable, sobresaliente.
- Puesto conseguido en una prueba deportiva: 1º, 2º, 3º,...

- Medallas de una prueba deportiva: oro, plata, bronce.

Variable cuantitativa

Una variable cuantitativa es la que se expresa mediante un número, por tanto se pueden realizar operaciones aritméticas con ella. Podemos distinguir dos tipos:

Variable discreta

Una variable discreta es aquella que toma valores aislados, es decir no admite valores intermedios entre dos valores específicos. Por ejemplo:

El número de hermanos de 5 amigos: 2, 1, 0, 1, 3.

Variable continúa

Una variable continua es aquella que puede tomar valores comprendidos entre dos números. Por ejemplo: La altura de los 5 amigos: 1.73, 1.82, 1.77, 1.69, 1.75.

En la práctica medimos la altura con dos decimales, pero también se podría dar con tres.

Aplicación

1. Proponga dos elementos de estudio que sean de su interés y anote algunas variables de cada tipo que desearía conocer.

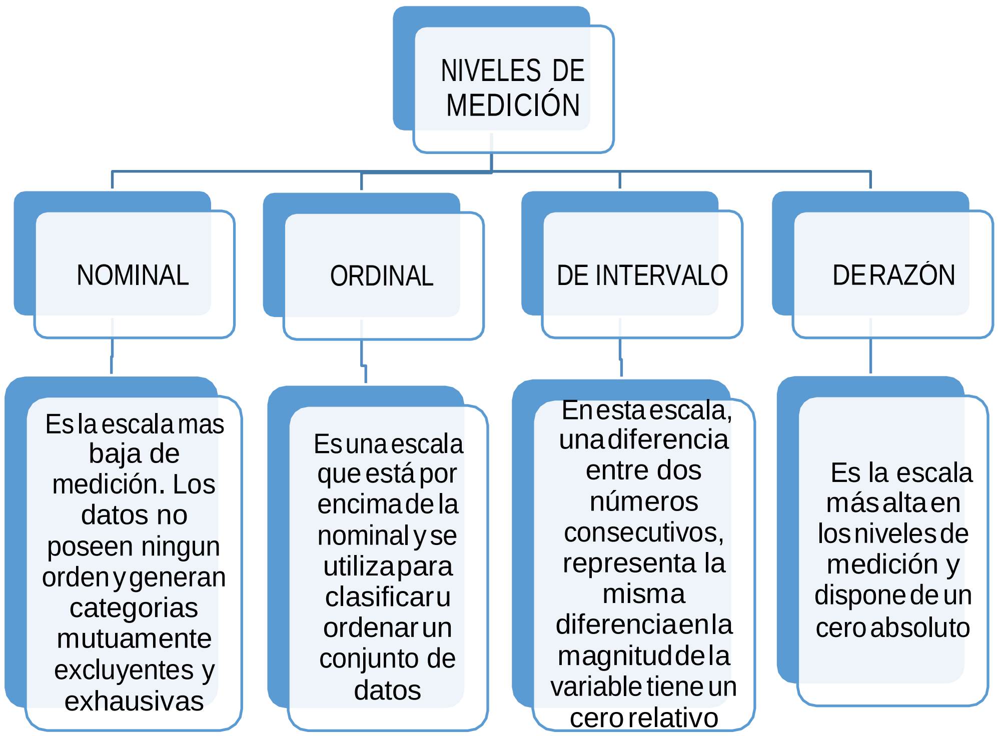
2. En un estudio estadístico se observó el tiempo que toma en llegar a su lugar de trabajo desde su lugar de residencia. Determine cuál es el elemento de observación, cuál es la variable de estudio y de qué tipo es.

1.2.2. Escalas De Medición

Es un instrumento de medida, de acuerdo al cual se asignan valores a los datos estadísticos.

Se reconocen cuatro niveles de medición: nominal, ordinal, de intervalo y de razón. En la *Figura 4*, se resumen las principales características de cada nivel.

Figura 3. Escala de medición



Elaboración: Autores

En la escala nominal, los valores obtenidos en las observaciones, pueden ser codificados mediante números, donde estos números no tienen valor intrínseco, sino solo la de etiquetar; por lo tanto, no pueden ser ordenados o medidos. Generalmente los datos que provienen de variables cualitativas son representados en esta escala de medición. Ejemplo: si se está determinando el estado civil de las personas, este estado puede ser etiquetado como: casado (1), soltero (2), divorciado (3), unión libre (4).

Los datos provenientes de un nivel ordinal, también generan categorías mutuamente excluyentes (significa que una observación en particular no pude pertenecer a dos o más categorías al mismo tiempo). Las encuestas que utilizan una escala tipo Likert (ver *Tabla 2*), utilizan datos medidos en escala ordinal. Ejemplo: se puede pedir que el encuestado diga cómo percibe el presente libro de estudio, para lo cual puede contestar con cualquiera de las siguientes opciones.

Tabla 2 Ejemplo de escala tipo Likert

Moderadamente			Extremadamente	
No útil	**Poco útil**	**Útil**	**Muy útil**	**útil**
1	2	3	4	5

Elaboración: Autores

En este caso, no se pueden establecer diferencias entre los valores numéricos, toda vez que estos representan una etiqueta o código, pero si responden a un orden preestablecido.

Ejemplos de datos medidos en escala de intervalo constituyen las escalas de temperatura, excepto la medida en grados Kelvin. Con los datos que están medidos en esta escala se puede establecer las operaciones de suma y substracción, pero no de multiplicación y división. Ejemplo: Al decir que en una ciudad la temperatura ambiente es de 20º C y en otra es de 40º C, se puede establecer que la segunda ciudad tiene una temperatura más alta de 20º C, pero no se puede concluir que la temperatura de ésta es el doble de la otra.

Finalmente los datos que se obtuvieron mediante un nivel de razón, a más de poseer todas las características anteriores, también puede trabajar con los datos en operaciones de multiplicación y división.

Ejemplo: Dos Personas perciben como salario, la A: $300 y la B: $600. Con esta información se puede concluir que B gana $300 más que A y que además gana el doble que A. Esto es posible por cuanto esta variable dispone de un cero absoluto.

Clasificación de las escalas

Las Escalas al igual que las variables se clasifican en cualitativas y cuantitativas, al mismo tiempo se subdividen en:

- **Cualitativas:**
  - Escala nominal o clasificatoria
  - Escala ordinal
- **Cuantitativas:**
  - Escala discreta o discontinua
  - Escala concretao continua
- **Otras escalas:**
  - Escala cronológica
  - Escala intervalar
  - Escala de razón

Escala nominal o clasificatoria

Este tipo de variables, no presentan un ordenamiento previo, más al contrario es arbitraria, de ahí que se haya ideado tres parámetros para entender mejor este tipo de escala; variable, escala y diferencia, por Ejemplo:

- Variable: Profesión
  Escala: Ingeniero, Médico, Abogado, Enfermero, Odontólogo
  Diferencia: No existe diferencia entre los profesionales en cuanto a nivel de estudio (tercer nivel)
- Variable: Provincias de la Costa del Ecuador
  Escala: Esmeraldas, Manabí, Santo Domingo de los Tsáchilas, Guayas, Santa Elena, El Oro
  Diferencia: Ninguna
- Variable: Sexo
  Escala: Masculino, Femenino
  Diferencia: Ninguna
- Variable: Estado civil
  Escala: Soltero, Casado, Divorciado, Viudo, Unión libre
  Diferencia: Ninguna

Escala ordinal

Son variables susceptibles de ser medidas siguiendo un ordenamiento (orden), formada por una clase mutuamente excluyente, que se agrupan de acuerdo a un orden preasignado. Por ejemplo:

- Variable: Grado de instrucción
  Escala: Primaria, Secundaria, Superior, Posgrado
  Diferencia: Existe diferencia entre diferentes niveles de la escala entre los estudiantes no solo años de experiencia, sino de conocimiento
- Variable: Grado de militar y/o policial
  Escala: Soldado, Sargento, Suboficial, Oficial, General
  Diferencia: Existe diferencia entre los grados jerárquicos no solo en Años de experiencia, sino en tiempo de estudio
- Variable: Jerarquía familiar
  Escala: Hijo menor, Hijo intermedio, Hijo mayor, Madre o Padre, Padre o Madre
  Diferencia: Existe diferencia entre los grados jerárquicos de la familia no solo en años de experiencia, sino en edad

Escala discreta o discontinua

Se dice que si la variable medida es susceptible a ser contada, se puede construir una escala discreta, formada por números ENTEROS con incrementos fijos, donde las fracciones no son consideradas; para esto, se debe considerar la magnitud de los números expuestos. Por ejemplo:

- Variable: Número de hijos
  Escala: 1 hijo, 2 hijos, 3 hijos, 4 hijos

Amplitud: Entre 4 y 1 hijos, existe una amplitud de 3 hijos

- Variable: Número de visitas
  Escala: De 1 a 3 visitas, De 4 a 6 visitas, De 7 a 9 visitas, De 10 a 12 visitas
  Amplitud: Entre 1 y 3, existe una amplitud de 2

- Variable: Número de caries dental
  Escala: De 1 a 3 caries, De 4 a 6 caries, De 7 a 9 caries
  Amplitud: Entre 1 y 3 caries, existe una amplitud de 2 caries

Escala concreta o continua

Cuando se cuenta con variables de tipo cuantitativo continuo o concreto, se pude utilizar este tipo de escala, cuyo requisito es el de poder presentar números relativos o racionales (fraccionados, porcentuales y/o decimales) siendo esta medición aproximada. Por ejemplo:

- Variable: Estatura
  Escala: 1,65 m
  1,66 m
  1,67 m
  1,68 m
  1,69 m
  Amplitud: Entre 1,65 y 1,69 m, existe una amplitud de 0,5 m

- Variable: Peso
  Escala: 6,5 Kg
  7,5 Kg
  8,5 Kg
  9,5 Kg
  Amplitud: Entre 6,5 y 9,5 kg, existe una amplitud de 4,0 kg

- Variable: Tiempo
  Escala: 1,10 h
  2,10 h
  3,10 h
  4,10 h
  Amplitud: Entre 1,10 y 4,10 h, existe una amplitud de 4 h.

Escala dicotómica

Es aquella escala que presenta tan solo dos opciones para medir la variable, siendo esta variable de tipo cualitativo o cuantitativo, dependiendo de la información o resultado que se busque. Por ejemplo:

- Variable Preferencia por un equipo de fútbol
  Escala de medición: Liga y Barcelona

Respuesta: Liga o Barcelona

- Variable. Sexo de un estudiante de la Universidad Central
  Escala de medición: Masculino y femenino

  Respuesta: Masculino o femenino

Escala cronológica

Es un tipo de escala cuantitativa continua, se la utiliza para estudiar algunos fenómenos en función al tiempo, algunos autores la tratan como si fuera una escala de variable independiente, permite conocer un determinado fenómeno a través del tiempo, es decir permite un seguimiento temporalizado (en el pasado, en el presente o en el futuro). Por ejemplo:

- Variable. Cambios físicos de una persona
  Escala de medición: al 1 año, 5 años, 25 años, 50 años, 75 años, 100 años
  Respuesta: descripción de las alteraciones físicas durante su vida

Escala intervalar

Las categorías se ordenan en unidades igualmente espaciadas, siendo posible medir las diferencias relativas en cada punto de la escala, no existe el cero absoluto. Por ejemplo:

- Variable: Medición de la temperatura corporal
  Escala: Grados centígrados o Celsius (37º)

  Diferencia: Números mayores o menores de 37º en la escala de temperatura

Escala de razón

En esta escala, si existe el cero absoluto y la magnitud de diferencia entre los valores numéricos entre sí. Por ejemplo:

- Variable: Relación entre edades
  Escala: Juan: 0 años (recién nacido)
  José: 9 años
  Joaquín: 18 años
  Diferencia: Joaquín 18 años (9 años más que José y 18 años más que Juan)

Aplicación

1. ¿Cuál es el nivel de medición en cada una de las siguientes variables?
   a. Calificaciones de estudiantes en una evaluación de Estadística
   b. Distancia recorrida por los estudiantes desde sus domicilios hasta la universidad.
   c. Determinación del nivel o semestre en el que se hallan los estudiantes.

d. Cantidad de horas semanales que dedican al autoestudio.
e. Grupo musical o cantante de su preferencia.

2. Determine el nivel de medición utilizado en los siguientes conceptos relacionados con el negocio de distribución de revistas quincenales.

a. La cantidad de periódicos vendidos cada mes, durante el año 2017 en la ciudad de Quito.
b. La cantidad de puestos de venta de periódicos por sector (norte, centro, sur).
c. Determinación de los nombres de los periódicos más vendidos.
d. Ubicación de los puestos de distribución.

3. Indique cuatro ejemplos en los cuales aparezca un nivel de medición diferente.

1.3. Distribución de frecuencias

Distribución de frecuencias: Es una tabla estadística donde se presentan los datos resumidos, de tal manera que se puede en una visión panorámica establecer un criterio sobre su comportamiento, entendiéndose por comportamiento, la determinación aproximada de los valores centrales, la variabilidad que presentan y si son o no relativamente simétricos con relación a un valor central.

En una tabla de frecuencias se pueden resumir cualquier tipo de datos, categóricos (nominales), ordinales, discretos y continuos, para este último tipo de datos, más adelante se verá un procedimiento para crear una distribución de frecuencias. Para los tres primeros tipos de datos, es decir, nominales, ordinales y discretos, la distribución de frecuencias constará básicamente de dos columnas, la izquierda reservada para las categorías (valores) que aparecen en el estudio respectivo y la de la derecha donde se ubica la frecuencia de clase respectiva. Antes de dar Ejemplos de estas distribuciones veamos dos criterios que son muy utilizados.

Observaciones:

- Frecuencia es el número de veces que un elemento se repite en un estudio estadístico
- Frecuencia de clase, es el número de elementos que tiene cada una de las categorías mutuamente excluyente de una distribución

En algunas ocasiones estos dos criterios pueden ser sinónimos, especialmente cuando se trata de distribuciones de frecuencias de datos nominales, ordinales y discretos, para datos continuos los términos antes referidos no tienen el mismo significado. De la misma manera si la información de un estudio estadístico no está categorizada, entonces no aparecerán frecuencias de clase y solo habrá frecuencia en caso de elementos que se repiten.

La distribución de frecuencias deberá contar con un título que informe sobre qué versa el estudio y referencias (fuente) en caso de ser tomado de estudios anteriores. Ejemplo:

Estudio sobre tipo de automotores que circularon en quince minutos en la intersección de las calles “10 de Agosto” y “Colón”

Tabla 3 Tipo de automotores que circularon en quince minutos en la intersección de las calles “10 de Agosto” y “Colón”

TIPO	Nº DE AUTOMOTORES
Automóvil	25
Bus	12
Camioneta	10
Camión	5
Furgoneta	8
TOTAL	**60**

Elaboración: Autores

La *Tabla 3*, representa un estudio realizado con una variable cualitativa (tipo de automotor), por lo tanto, los datos son de carácter nominal, se puede observar que la categorización representa a los datos que fueron observados y no existe un orden predeterminado para realizar la categorización; además cabe resaltar que no aparece otro tipo de automotor ya que en los quince minutos en los cuales se hizo el estudio, no pasaron por la intersección ningún otro tipo de vehículo.

Observación:

Las distribuciones de frecuencia están conformadas por categorías mutuamente excluyentes y exhaustivas, esto significa que no existe la posibilidad de que un elemento pertenezca a varias categorías a la vez y que si un elemento fue observado, este debe constar en alguna de las categorías

Ejemplo:

Percepción del servicio recibido en la atención de la ventanilla de reclamos de una empresa.

Tabla 4 Percepción del servicio recibido en la atención de la ventanilla de reclamos de una empresa.

SERVICIO	Nº PERSONAS
Excelente	10
Muy bueno	15
Bueno	20
Regular	18
Malo	17
TOTAL	**80**

Elaboración: Autores

En la *Tabla 4*, los datos de la variable de estudio (**calidad de servicio**) están medidos en escala ordinal, por lo tanto las categorías constan en la tabla con un orden establecido. De la misma manera que en la *Tabla 3*, las categorías son mutuamente excluyentes y exhaustivas.

Ejemplo:

Estudio sobre la cantidad de hijos que tiene una muestra de familias de la ciudad de Quito

Tabla 5 Estudio sobre la cantidad de hijos que tiene una muestra de familias de la ciudad de Quito

NÚMERO DE HIJOS	Nº DE FAMILIAS
0	3
1	8
2	12
3	13
4	6
5	3
TOTAL	**45**

Elaboración: Autores

La *Tabla 5*, representa el estudio de una variable cuantitativa discreta, por lo tanto los datos que genera el estudio tienen la misma característica, es decir ser discretos. Las categorías tienen un orden establecido y siguen siendo mutuamente excluyentes yexhaustivas.

1.3.1. Distribuciones de frecuencia para variables continuas

Para este tipo de variables es necesario entrar en un proceso de categorización, en la cual las categorías están representadas mediante intervalos, que agrupan a unconjunto de valores que están incluidos en cada uno de ellos. Esto es necesario ya que la característica de estas variables, es que pueden en teoría tomar cualquier valor entre dos valores enteros consecutivos, esto hace que sea muy improbable que existan observaciones que tengan exactamente los mismos valores. Por ejemplo, si estamos determinando la estatura de una muestra de cincuenta personas es muy probable que tengamos cincuenta estaturas diferentes y si categorizamos con estas diferentes mediciones, no estaríamos cumpliendo el objetivo de las distribuciones de frecuencia, el cual es, resumir la información obtenida en el estudio.

A continuación se indica un procedimiento para la construcción de una distribución de frecuencias.

1. Determinar el rango de valores entre los cuales se sitúan todas las observaciones, es decir se observa el valor máximo y el mínimo y se obtiene su diferencia.

RANGO = MÁXIMO - MÍNIMO

2. Se decide el número de categorías que va a tener la distribución, este número se sitúa para la mayoría de los investigadores entre 5 y 15. La cantidad de categorías se la establece de acuerdo al número (***n***) de observaciones del estudio. De una manera tentativa se determina esta cantidad al aplicar este criterio:

Nº DE CATEGORÍAS = 1 + 3.3*LOG(n)

El resultado se aproxima al entero inmediato menor o mayor.

3. Se calcula la amplitud (***i***) de cada intervalo de clase, dividiendo el rango entre el número de categorías. Diferentes mediciones, no estaríamos cumpliendo el objetivo de las distribuciones de frecuencia, el cual es, resumir la información obtenida en el estudio.

A continuación se indica un procedimiento para la construcción de una distribución de frecuencias.

1. Determinar el rango de valores entre los cuales se sitúan todas las observaciones, es decir se observa el valor máximo y el mínimo y se obtiene su diferencia.

RANGO = MÁXIMO - MÍNIMO

2. Se decide el número de categorías que va a tener la distribución, este número se sitúa para la mayoría de los investigadores entre 5 y 15. La cantidad de categorías se la establece de acuerdo al número (n) de observaciones del estudio. De una manera tentativa se determina esta cantidad al aplicar este criterio:

Nº DE CATEGORÍAS = 1 + 3.3*LOG(n)

3. El resultado se aproxima al entero inmediato menor o mayor. Se calcula la amplitud (i) de cada intervalo de clase, dividiendo el rango entre el número de categorías.

i =RANGO/(N° CATEGORÍAS)

El resultado se aproxima al valor inmediato superior

4. Se establecen los límites de cada categoría, para esto se determina el límite inferior de la primera categoría, que puede ser igual o menor al valor mínimo de todo el estudio, a este valor se suma la amplitud del intervalo y se obtiene el límite inferior de la siguiente categoría y así sucesivamente sumamos la amplitud del intervalo a cada nuevo límite obtenido, hasta completar el número de categorías establecido. Los límites superiores de las categorías quedan determinados al establecer que serán los valores menores que el límite inferior de la categoría inmediata siguiente, esto permite que las categorías sean mutuamente excluyentes.

5. Finalmente se completa la categorización con la determinación de las frecuencias de clase de cada categoría, esto es contar cuántos elementos están incluidos en cada una de ellas. Recordar que la sumatoria de estas frecuencias debe ser igual que el total de elementos observados en el estudio. (Condición de exhaustivo: ningún elemento puede quedar sin ser considerado en alguna de las categorías).

Para este paso, vamos a utilizar una herramienta de Excel, que permite contar automáticamente los elementos que tiene cada categoría, sin necesidad de que los datos se encuentren ordenados. Si no utilizamos esta herramienta debemos ordenar previamente los datos y realizar un conteo manual.

A continuación vamos a realizar la categorización del siguiente conjunto de datos obtenidos en la determinación del tiempo (minutos) que un cliente se mantuvo en la fila antes de ser atendido en una sucursal bancaria de la localidad.

12	14	9	10	8	26	27	14	13	14
3	5	10	8	7	7	6	13	12	21
25	27	22	7	12	12	13	19	18	17
28	30	25	21	15	15	16	21	20	14
14	16	11	18	21	8	9	10	9	9
7	9	4	32	20	4	5	8	7	18
6	8	12	11	16	31	24	26	25	26
4	6	18	12	14	22	23	31	30	12

Como los datos no se encuentran ordenados, para la determinación del total de observaciones, así como de los valores máximo y mínimo, vamos a recurrir a la hoja electrónica de Excel y utilizar las funciones: **CONTAR, MÁXIMO, MÍNIMO.** Entonces es necesario que estos datos sean ingresados en un grupo de celdas seleccionado.

12	14	9	10	8	26	27	14	13	14
3	5	10	8	7	7	6	13	12	21
25	27	22	7	12	12	13	19	18	17
28	30	25	21	15	15	16	21	20	14
14	16	11	18	21	8	9	10	9	9

7	9	4	32	20	4	5	8	7	18
6	8	12	11	16	31	24	26	25	26
4	6	18	12	14	22	23	31	30	12

En la barra de fórmulas activar insertar función ***(fx)*** a continuación seleccionar en la opción de usadas recientemente **Estadística** y en la lista que se genera escoger cada una de las funciones indicadas: CONTAR, MÁXIMO, MÍNIMO. Activando para cada una de ellas celdas diferentes.

NÚMERO DE ELEMENTOS	80
MÁXIMO	32
MÍNIMO	3

Determinación del rango: Rango = máximo − mínimo, entonces R = 32

− 3 = 29

2. Cálculo del número ce categorías = 1 + 3.3*log (80) → 1 + 3.3 (1.903) → 7.28 por lo tanto podemos escoger 7 u ocho categorías.

3. Amplitud de ancho del intervalo: si escogemos siete:

Si escogemos ocho: entonces lo más conveniente es trabajar con ocho categorías, toda vez que en la aproximación al entero más cercano, cometemos el menos error.

4. Pasamos a conformar los límites de las categorías: escogemos como límite inferior de la primera categoría (2) es un valor menor que el mínimo observado, a este valor sumamos el ancho del intervalo (4) y nos da (6), que será el límite inferior de la siguiente categoría y así seguimos sumando
(4) hasta llegar al límite inferior de la octava categoría.

Límite inferior	Ancho Del intervalo	Límite siguiente
2	+4	6
6	+4	10
10	+4	14
14	+4	18
18	+4	22
22	+4	26
26	+4	30
30	+4	34

Entonces los límites de las categorías quedan estructurados de la siguiente manera a través de intervalos semiderruidos (a, b] o [a, b):

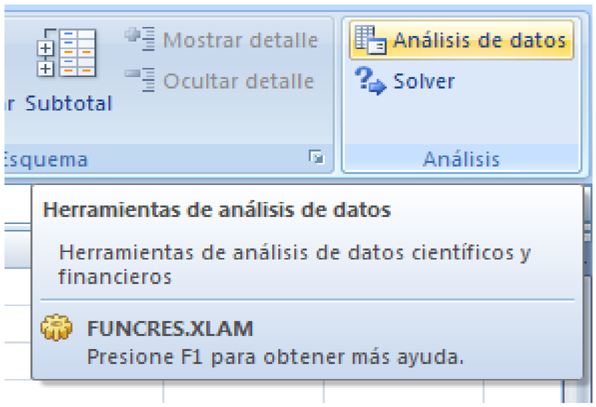
Límite inferior	Límite Superior	Intervalo		Límite Inferior	Límite Superior	Intervalo
2	<6	[2, 6)		>2	6	(2, 6]
6	<10	[6, 10)		>6	10	(6, 10]
10	<14	[10, 14)		>10	14	(10, 14]
14	<18	[14, 18)		>14	18	(14, 18]
18	<22	[18, 22)		>18	22	(18, 22]
22	<26	[22, 26)		>22	26	(22, 26]
26	<30	[26, 30)		>26	30	(26, 30]
30	<34	[30, 34)		>30	34	(30, 34]

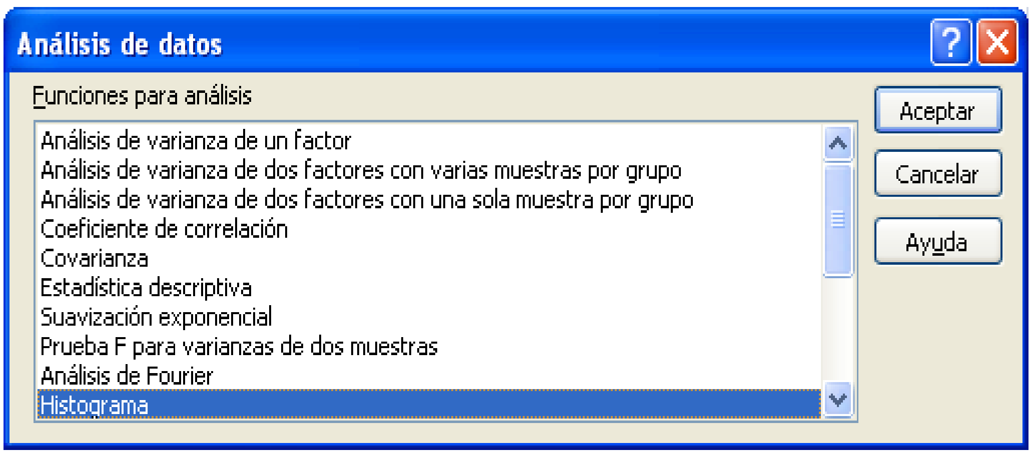
En la práctica se prescinde de los signos "menor que" de los límites superiores, ya que hay una aceptación general que la lectura de estos límites será siempre menor que el valor anotado, o de la misma manera se prescinde de los símbolos "mayor que" de los límites inferiores. Con las citadas conformaciones, se cumple la condición de que las categorías sean mutuamente excluyentes.

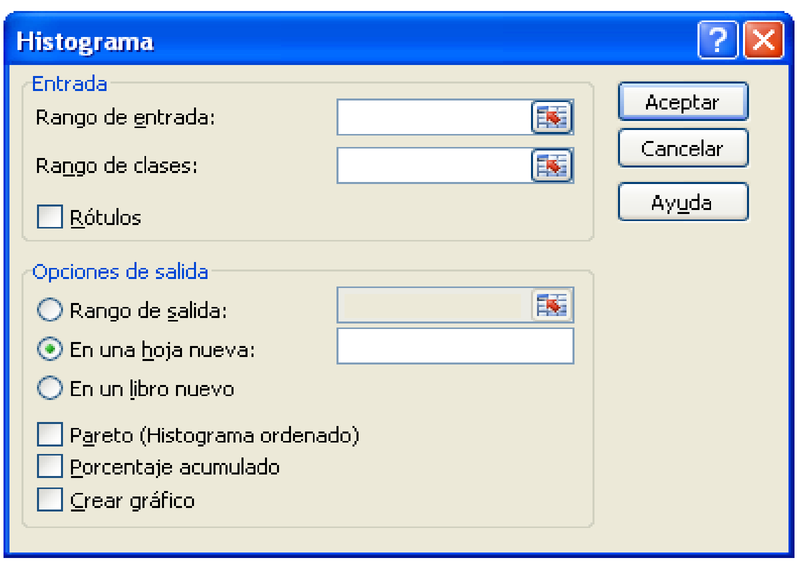
5. Para determinar el número de elementos que tiene cada categoría, se debe Realizar el conteo respectivo en los datos originales, esto es, si se sigue un proceso manual, pero podemos utilizar la computadora para efectuar el proceso más rápido al recurrir a una herramienta estadística que está dentro de la hoja electrónica de Excel. Para acceder a la misma, debemos cerciorarnos que al activar la pestaña ***Datos***, en la esquina superior derecha de la hoja (barra de íconos) debe aparecer la opción de análisis de datos.

Si esta opción no está activada será necesario que se active en ***complementos*** esta alternativa (Microsoft Excel 2007); para versiones anteriores activar en ***Herramientas*** este completo (***Análisis de datos***).

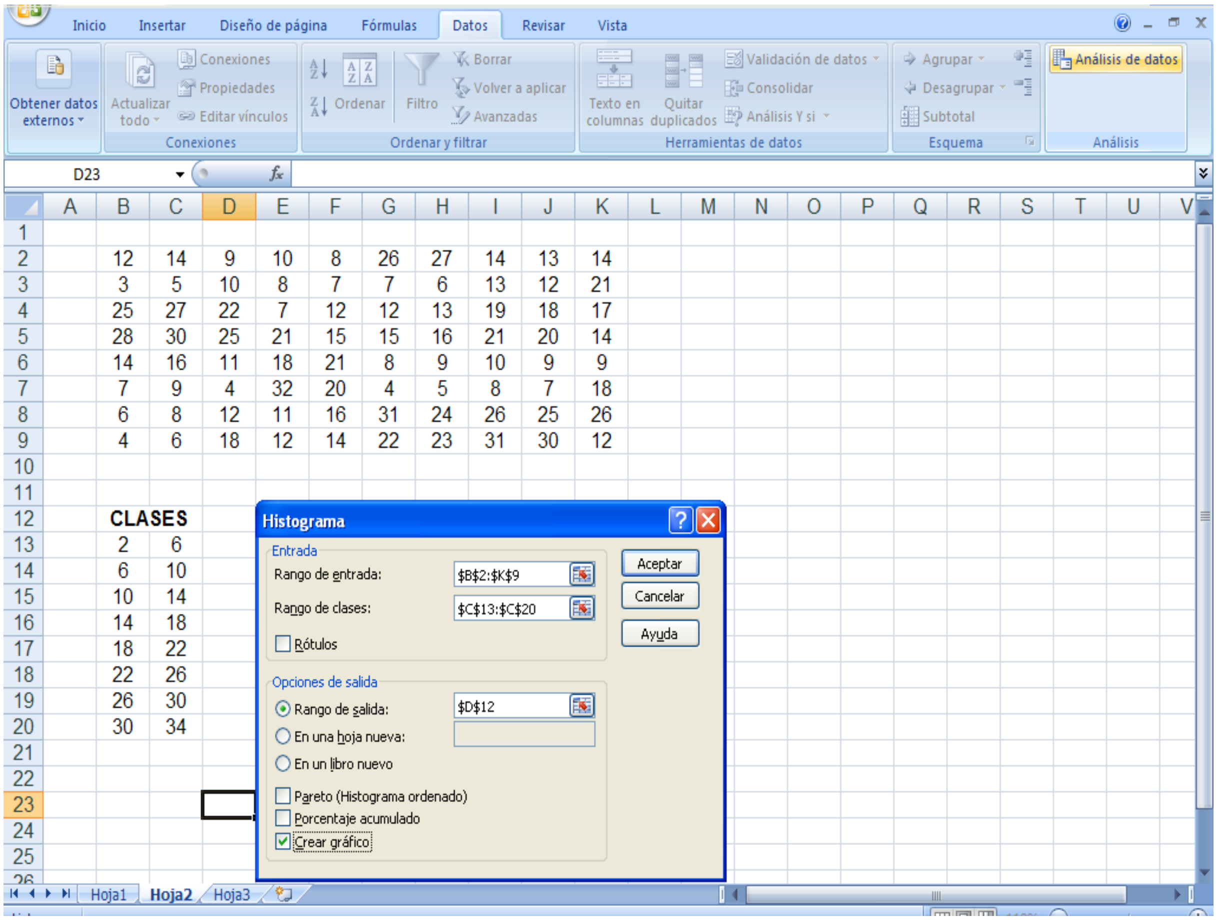
Previamente en la hoja de Excel, donde están los datos originales, ubicamos los límites de las categorías en celdas separadas, con la creación de esta matriz estamos listos para utilizar la herramienta que nos permitirá realizar el conteo automático de datos. Al escoger ***Análisis de datos*** se abre una lista con las funciones estadísticas incluidas, seleccionamos "***Histograma***" y se da un clic en aceptar, se abre un cuadro de diálogo en el que nos pide, primero el rango de entrada de datos, esto es, debemos señalar las celdas donde se hallan los datos originales; segundo nos pide el rango de clases, para lo cual marcamos la columna correspondiente a los límites superiores de las categorías, marcamos a continuación el rango de salida, es decir seleccionamos una celda diferente de donde se encuentran nuestros datos y finalmente activamos la opción de creación de gráfico, a fin de que nos reproduzca el gráfico respectivo (***histograma***).





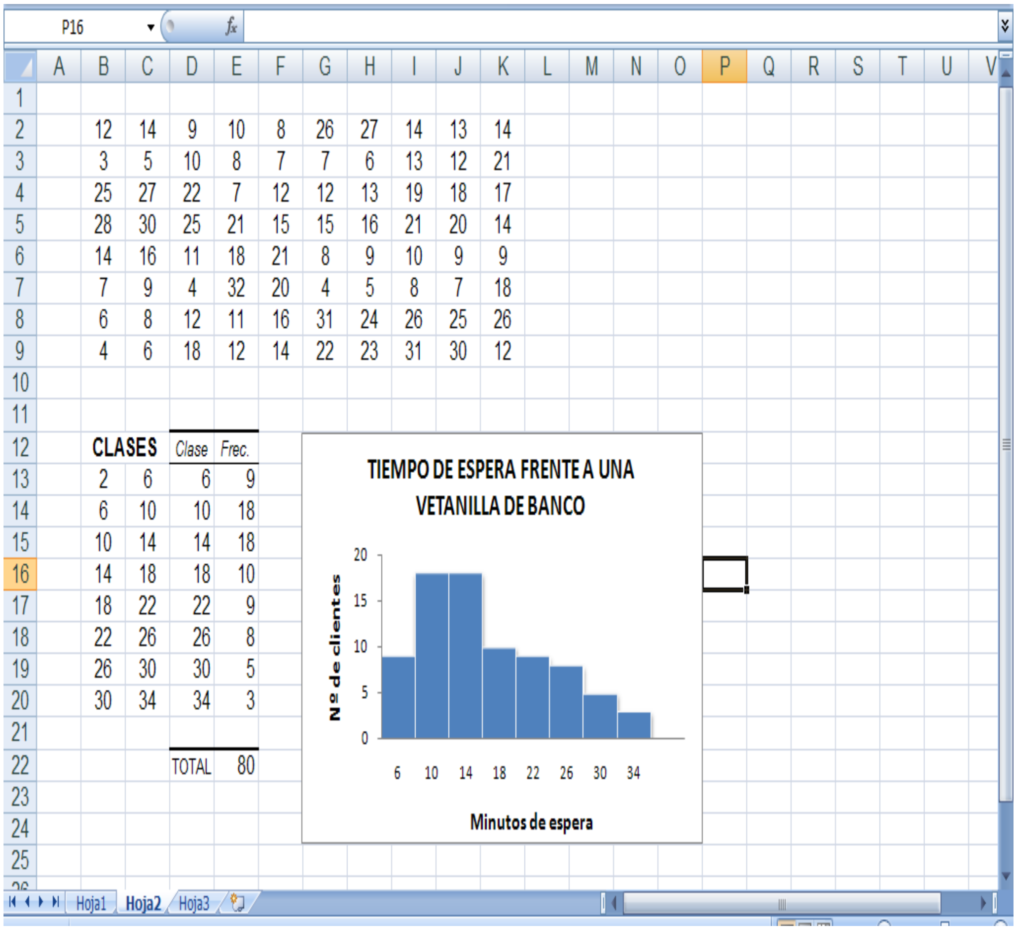


El proceso descrito anteriormente en el numeral (5), se lo puede observar con claridad en los siguientes reportes de pantalla de Excel.



Podemos ver que en el rango de entrada están marcadas las celdas desde la B2, hasta la K9, el rango de clases las celdas desde la C13 hasta la C20; el rango de salida se ha marcado la celda D12 y se encuentra activada la opción “Crear gráfico”. En esta instancia lo que hacemos es dar un clic en **Aceptar**, ante lo cual el programa nos devolverá los resultados que requerimos.

En la siguiente pantalla aparecen los resultados obtenidos. En el gráfico se ha trabajado previamente un poco a fin de darle el tamaño adecuado. De la misma manera en las clases se ha eliminado la última categoría que aparece como mayor y con frecuencia cero y se ha determinado el total de elementos, que en este caso es igual a ochenta.



	B	C	D	E	F	G	H	I	J	K
2	12	14	9	10	8	26	27	14	13	14
3	3	5	10	8	7	7	6	13	12	21
4	25	27	22	7	12	12	13	19	18	17
5	28	30	25	21	15	15	16	21	20	14
6	14	16	11	18	21	8	9	10	9	9
7	7	9	4	32	20	4	5	8	7	18
8	6	8	12	11	16	31	24	26	25	26
9	4	6	18	12	14	22	23	31	30	12

CLASES		Clase	Frec.
2	6	6	9
6	10	10	18
10	14	14	18
14	18	18	10
18	22	22	9
22	26	26	8
26	30	30	5
30	34	34	3
		TOTAL	80

De tal manera que podemos resumir nuestro estudio, generando una distribución de frecuencias bien estructurada y un gráfico representativo de ella

Luego de este trabajo inicial, donde hemos obtenido la frecuencia absoluta de clases (*f*), se puede determinar otro tipo de frecuencias que son muy utilizadas en estadística. Estas frecuencias son**: frecuencia relativa (fr), frecuencia porcentual (f%), frecuencia acumulada menor que (fa<) o mayor que (fa>) y frecuencias acumuladas porcentuales (fa%). Estas frecuencias se las calcula de la siguiente manera:**

Para las frecuencias acumuladas se debe responder al criterio que se quiere establecer “menor que” o “mayor que” para proceder a la acumulación de elementos, en el caso de requerir la frecuencia acumulada menor que, se procede de la manera siguiente, se comienza con la primera categoría y se observa cuántos elementos están por debajo o son menores que el límite superior de esta categoría, luego se contabiliza la cantidad de elementos que son menores que el segundo límite superior y así sucesivamente, hasta llegar a la última categoría que tendrá una frecuencia acumulada igual al total de observaciones (***n***).

Para el caso de la frecuencia acumulada “mayor que” se procede de la misma manera pero respondiendo a la pregunta de cuántos elementos son mayores o iguales al límite inferior de cada categoría. Si es necesario que las frecuencias acumuladas se presenten en forma relativa o porcentual, tendrán que dividirse para el total de elementos del estudio y en el caso porcentual, multiplicar por 100.

También se puede representar a los intervalos de las distribuciones de frecuencias, mediante las ***marcas de clase*** (se les suele simbolizar con la variable ***x***) o ***puntos medios de las clases***, la obtención de estos valores se consigue al sumar los límites de cada categoría y dividirlos para dos.

Observación:
Marca de clase (□) = $\frac{(\text{limite inferior}+\text{limite superior})\ 2}{}$

En la *Tabla 6*, se han calculado las diferentes frecuencias descritas anteriormente, para el ejemplo que se viene desarrollando.

Tabla 6 . Frecuencias

Límites		**Marca de clase**	**f**	**fr**	**f%**	**fa"<"**	**fa">"**	**fa%"<"**	**fa%">"**
2	6	4	9	0,1125	11,3	9	80	11,3	100
6	10	8	18	0,225	22,5	27	71	33,8	88,8
10	14	12	18	0,225	22,5	45	53	56,3	66,3
14	18	16	10	0,125	12,5	55	35	68,8	43,8
18	22	20	9	0,1125	11,3	64	25	80	31,3
22	26	24	8	0,1	10	72	16	90	20
26	30	28	5	0,0625	6,3	77	8	96,3	10
30	34	32	3	0,0375	3,8	80	3	100	3,8
TOTAL			80	1	100				

Elaboración: Autores

Es necesario indicar que estas frecuencias no siempre se calculan todas al mismo tiempo, pues habrá ocasiones en las que no se necesitan conocer todos los valores al mismo tiempo, sino tal vez una o dos de las frecuencias adicionales calculadas.

Aplicación

Las calificaciones de 50 alumnos en Matemáticas han sido las siguientes:

5, 2, 4, 9, 7, 4, 5, 6, 5, 7, 7, 5, 5, 2, 10, 5, 6, 5, 4, 5, 8, 8, 4, 0, 8, 4, 8, 6, 6, 3, 6, 7, 6, 6, 7, 6, 7, 3, 5, 6, 9, 6, 1, 4, 6, 3, 5, 5, 6, 7.

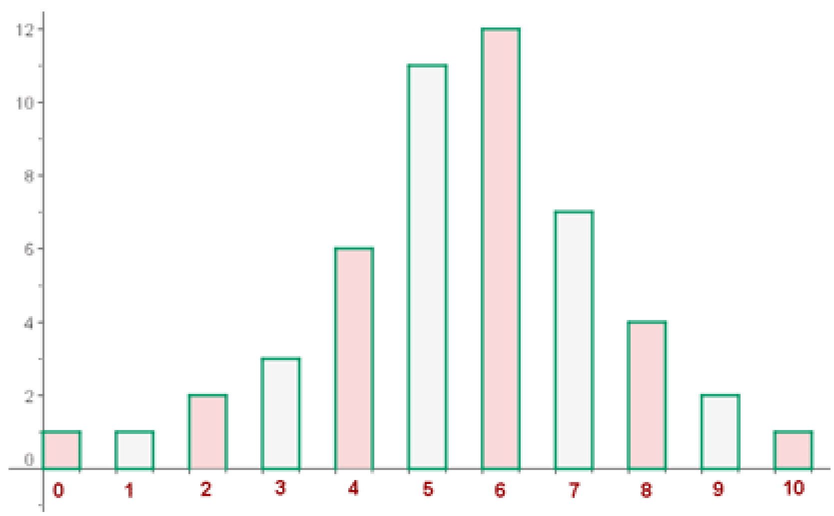
Construir la tabla de distribución de frecuencias y dibuja el diagrama de barras.

Tabla 7 Tabla de frecuencias

x_i	f_i	F_i	n_i	N_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
	50		1.00	

Elaboración: Autores

D i a g r a m a d e b a r r a s



a. Datos Estadísticos

Existen muchos gráficos estadísticos, nosotros centraremos nuestra atención en los más utilizados de acuerdo a la variable que se esté estudiando, para todos los demás el lector puede profundizar su estudio, bien utilizando la función de ayuda de Excel para cada gráfico que puede crear esta hoja de cálculo.

Los principales gráficos estadísticos los podemos resumir en:

- Diagramas de barras o columnas
- Diagrama de sectores
- Diagramas lineales
- Diagramas de dispersión
- Histogramas
- Polígonos de frecuencia
- Ojivas

Los dos primeros diagramas son utilizados para variables cualitativas y también para cuantitativas, pero cuando los datos son discretos, el diagrama lineal es muy utilizado.

Sobre todo para estudiar el comportamiento de variables en el tiempo; el diagrama de dispersión es usado para analizar la relación de dos variables (correlación, regresión) y finalmente los tres últimos diagramas se usan para representar a datos categorizados mediante intervalos. A continuación se presentarán Ejemplos de cada uno de ellos. Los gráficos presentados han sido creados en Excel.

Diagrama de barras o columnas

Es un tipo de gráfico que constan de dos ejes, de los cuales se escoge a uno de ellos para representar a la variable de estudio de acuerdo a la distribución de frecuencias generada y el otro para representar la frecuencia de cada categoría, si es el eje vertical en el que se marcan las frecuencias el diagrama será de **columnas**, por el contrario si es en el eje horizontal donde se representan las frecuencia, el diagrama será de **barras**.

En este diagrama, bien, la altura de las columnas o la longitud de las barras, vienen establecidas por la frecuencia de cada categoría.

Existen otras posibilidades para estos diagramas, como columnas apiladas y columnas apiladas al 100%, dejamos al lector la investigación de cuándo usar estas alternativas.

- Ejemplo

Se ha realizado un estudio sobre la provincia de nacimiento de una muestra de 60 personas, obteniéndose la siguiente distribución:

Tabla 8 estudio sobre la provincia de nacimiento de una muestra de 60 personas

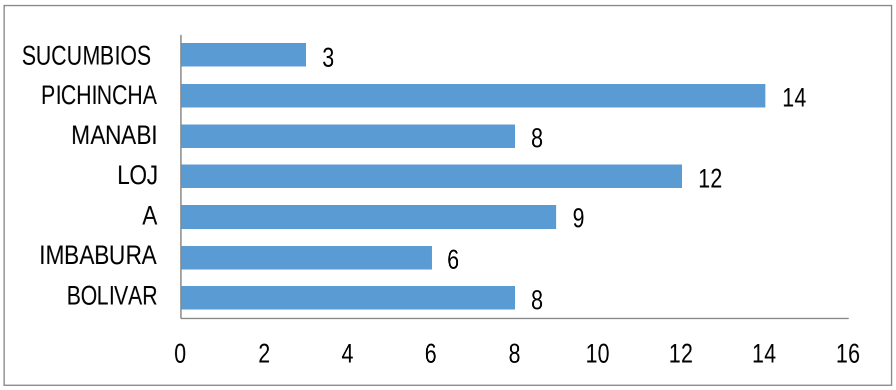
PROVINCIA	Nº DE PERSONAS
Azuay	8
Bolívar	6
Imbabura	9
Loja	12
Manabí	8
Pichincha	14
Sucumbíos	3
TOTAL	**60**

Elaboración: Autores

Como la variable de estudio (provincia de nacimiento), es de tipo cualitativo, se escogerá el diagrama de columnas para representar a este estudio y su gráfico se ve en la **figura.**

Si los mismos datos son representados mediante barras, se obtendrá el gráfico

Figura 4 estudio sobre la provincia de nacimiento de una muestra de 60 personas



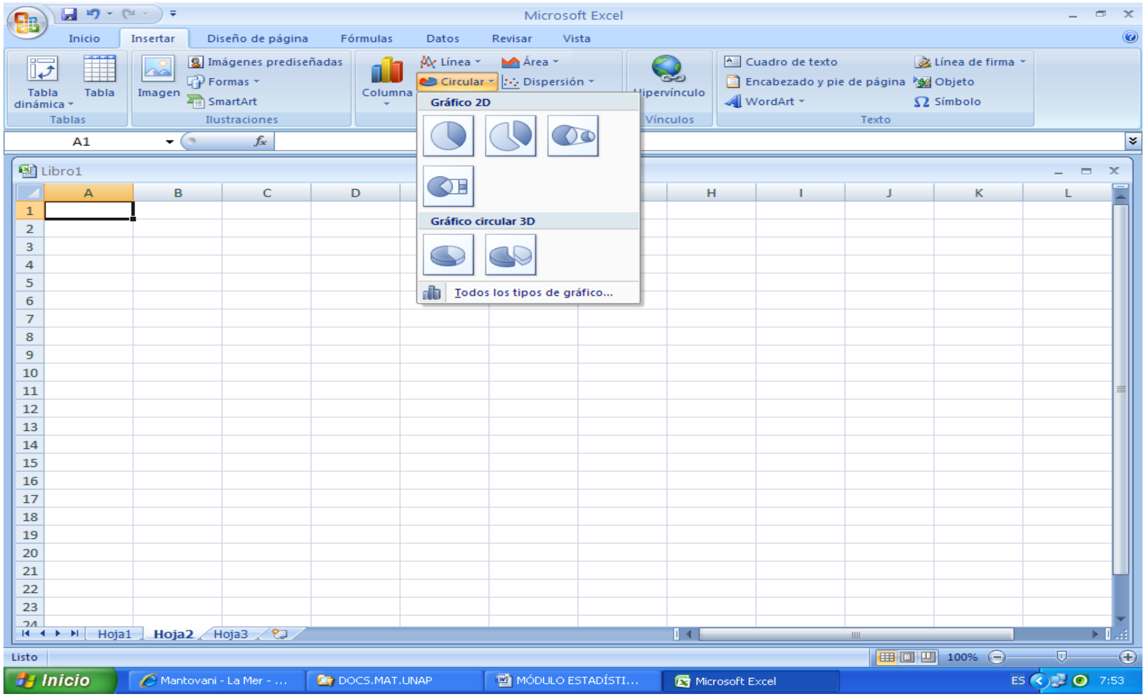
Elaboración: Autores

Diagrama de sectores

Es un gráfico que se presenta como un círculo en el cual constan divisiones o sectores que representan a las diferentes categorías que tiene la distribución, es un gráfico que es apropiado para variables cualitativas o cuantitativas discretas. En Excel, se tiene

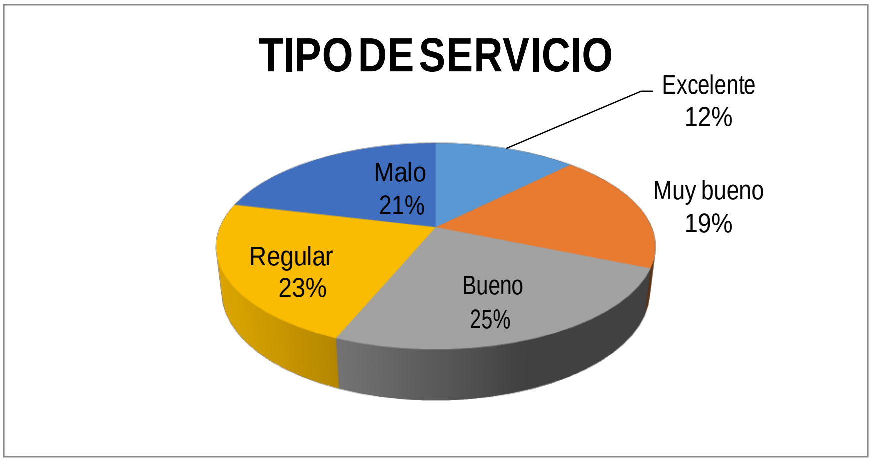
Varias opciones para la generación del gráfico, puede ser presentado como una figura

Bidimensional en el plano o en perspectiva, entonces se presenta como un disco, adicionalmente tiene otro tipo de representación como pueden ser anillos o coronas circulares, pero en esencia son equivalentes. Si el número de categorías es muy alto, es preferible seleccionar la alternativa de gráfico circular con subgráfico circular de barras.



TIPO SERVICIO	Nº PERSONAS
Excelente	10
Muy bueno	15
Bueno	20
Regular	18
Malo	17
TOTAL	**80**

Figura 5 Diagrama circular o de sectores



Elaboración: Autores

En el presente Ejemplo, como existió pocas categorías no fue necesario el uso de gráficos con subgráfico desagregado. La variedad de presentaciones que tiene esta opción de gráfico es muy amplia y cada una presenta alternativas que satisfacen las necesidades del investigador.

Diagramas lineales

Este tipo de gráfico es utilizado de manera usual en la representación de una variable en el tiempo, estableciéndose siempre que la variable tiempo estará ubicada en el eje horizontal y la otra variable en el eje vertical. El diagrama permite ver ciertas tendencias que pueden tener los datos en el transcurso del tiempo, en relación específicamente a si existe crecimiento o decrecimiento de la variable de estudio, durante el período de estudio.

El análisis de estos datos permitirá el cálculo de alguna función matemática que mejor represente el comportamiento de los datos, de tal manera de poder realizar pronósticos a un futuro inmediato.

- Ejemplo

Una empresa exportadora de flores ha determinado durante 12 años, el valor anual de sus exportaciones, expresado en millones de dólares, los datos se muestran en la siguiente tabla:

Tabla 9 valor anual de sus exportaciones

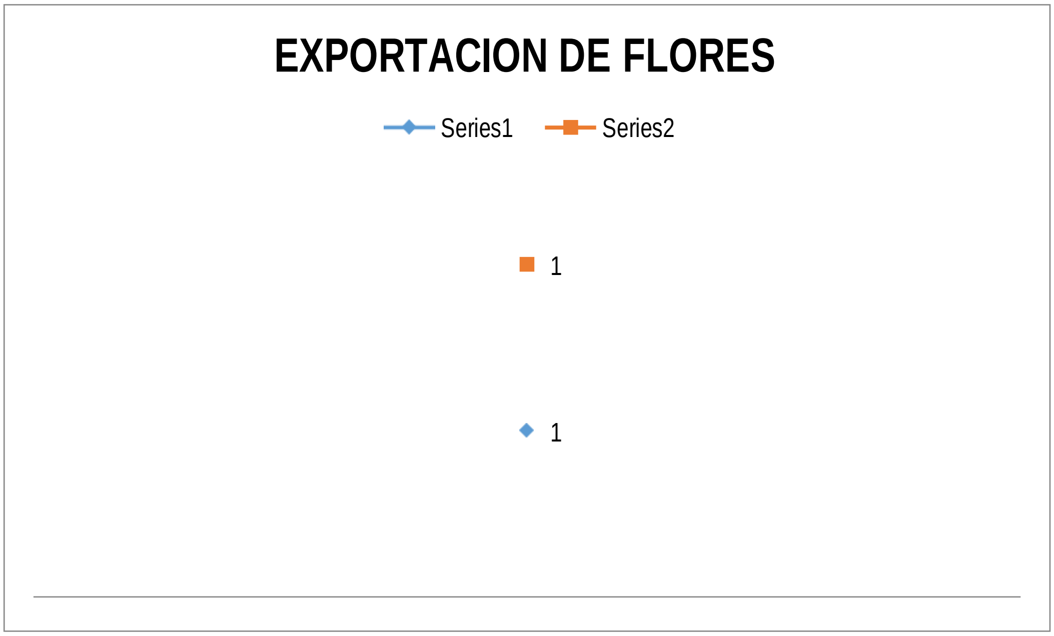
EXPORTACION DE FLORES(MILLONES$)			
AÑO	VALOR	AÑO	VALOR
1998	0.85	2004	1.35
1999	1.02	2005	1.28

2000	1.15	2006	1.16
2001	1.08	2007	1.38
2002	1.16	2008	1.45
2003	1.20	2009	1.55

Elaboración: Autores

El diagrama que se genera al insertar gráfico "líneas", se aprecia en la figura 13.6

Figura 6 valor anual de sus exportaciones



Elaboración: Autores

Del análisis de la figura 7 podemos concluir que a pesar de que hubo en los años 2001, 2005 y 2006 una disminución en la exportación de flores, con relación a los años inmediatos anteriores, la tendencia general de esta serie de datos de crecimiento, pudiendo concluir que las exportaciones de flores desde el período inicial hasta el final, casi se han duplicado.

Diagrama de dispersión

Es un gráfico que nos permite visualizar la relación entre dos variables de estudio, una con carácter de independiente y la otra dependiente de la primera, la representación sigue la convención establecida que, la variable independiente se ubicará en el eje horizontal y la dependiente en el eje vertical. Normalmente se presentan los valores como puntos ubicados en el plano, se recomienda no unir los puntos mediante segmentos

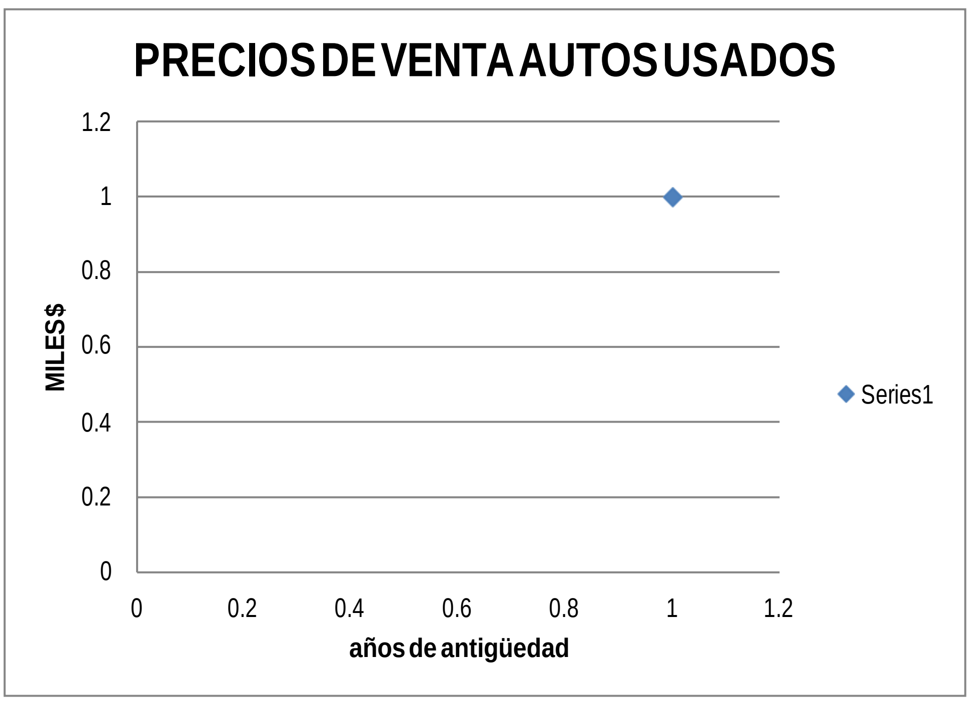
- Ejemplo

Tabla 10 precios de venta autos usados

ANTIGÜEDAD	PRECIO DE VENTA
9	7.8
7	8.1
11	4.8
10	5.9
8	8.0
7	9.5
8	7.9
11	5.2
10	6.3
12	4.2
5	10.5
6	9.2

Elaboración: Autores

Figura 7 precios de venta autos usados



Elaboración: Autores

Histograma

En esencia es un gráfico compuesto por una sucesión de rectángulos adyacentes, cada uno de los cuales representa a una categoría, con la condición de que el área de cada uno de ellos es igual o proporcional a la frecuencia de la categoría que representa. La variable de estudio se ubica en el eje horizontal y la frecuencia de clase (absoluta, relativa o porcentual) se ubica en el eje vertical. Cuando los intervalos de clase son iguales para todas las categorías, la altura de cada rectángulo es igual a la frecuencia de clase.

Tipo de gráfico es representativo de las distribuciones de frecuencia cuya variable de estudio es de tipo cuantitativa continua; como ya sabemos las clases o categorías de estas distribuciones están formadas mediante intervalos.

Al utilizar el asistente de gráficos de Excel, se escoge la opción de “diagramas de columnas” y seguidamente se opta por un subtipo de diagrama donde aparezcan los rectángulos juntos y no separados como en las columnas. Como valores representativos de la variable de estudio se ubican en el eje horizontal las marcas de clase.

Si el estudio comprende todo el proceso desde la creación de la distribución de frecuencias, entonces podemos utilizar la herramienta “Histograma” de la opción de análisis de datos, como ya vimos en páginas anteriores para la creación del gráfico respectivo.

- Ejemplo:

Se realizó un estudio sobre una muestra de 100 familias de cuatro integrantes, para determinar cuál es el gasto semanal en alimentación, obteniéndose la siguiente distribución:

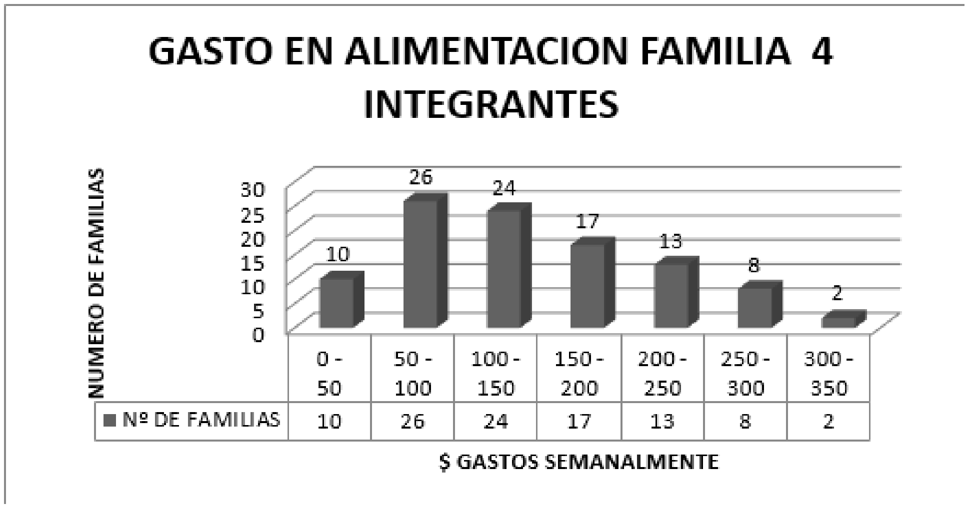
Tabla 11 Nº de familias

GASTO EN $	Nº DE FAMILIAS
0 - 50	10
50 - 100	26
100 - 150	24
150 - 200	17
200 - 250	13
250 - 300	8
300 – 350	2
TOTAL	**100**

Elaboración: Autores

Para generar el gráfico, debemos trabajar con la tabla anterior en Excel, en la cual se añadirá una columna donde se determinen las marcas de clase, para que sean estos valores los que representen a las categorías en el eje horizontal. A continuación se marcan las celdas correspondientes a las frecuencias de clase y pedimos Insertar: Diagrama de columnas, A continuación se escoge la opción de rectángulos adyacentes y el gráfico quedará generado. Así se lo observa.

Figura 8 Nº de familias



Elaboración: Autores

Polígono de frecuencias

Es otro gráfico representativo de las variables cuantitativas continuas, que consiste en un diagrama de segmentos rectos articulados en los puntos que se generan al ubicar en el plano a cada marca de clase con su respectiva frecuencia (**x**, **f**). La variable de estudio se ubica en el eje horizontal (representada por las marcas de clase) y la frecuencia de clase; absoluta relativa o porcentual en el eje vertical. El gráfico se concluye cuando se cierra el polígono con el eje horizontal, para el efecto se crean marcas de clase inmediatas anterior a la primera y posterior a la última, con frecuencias cero y con ancho de intervalo igual a las categorías que anteceden o suceden.

La principal característica de los polígonos de frecuencia es que el área que se encuentra bajo ellos y el eje horizontal es igual o proporcional al total de observaciones del estudio.

- Ejemplo

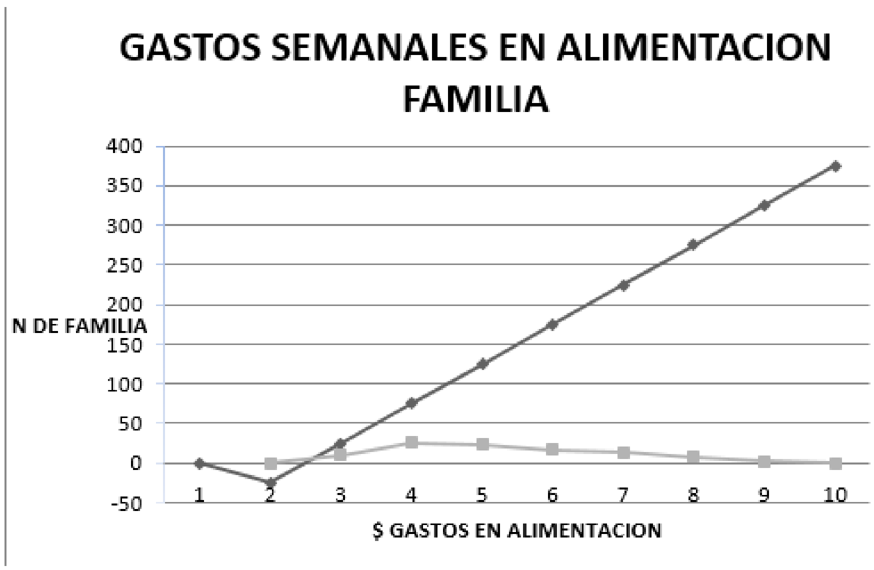
Vamos a trazar el polígono de frecuencias del estudio realizado sobre los gastos semanales familiares en alimentación, en la muestra de 100 familias compuestas por cuatro integrantes. Creamos la tabla respectiva en la cual se han adicionado las categorías anterior a la primera y posterior a la última con frecuencia cero.

Tabla 12 gastos semanales familiares en alimentación

MARCA DE CLASE	FRECUENCIA
-25	0
25	10
75	26
125	24
175	17
225	13
275	8
325	2
375	0

Elaboración: Autores

Figura 9 gastos semanales familiares en alimentación



Elaboración: Autores

Ojivas o polígono de frecuencias acumulada

También es un diagrama representativo de estudios estadísticos con variable cuantitativa continua; constituye un gráfico de tipo lineal que forma una línea poligonal abierta. Está representado por los segmentos rectos que se unen secuencialmente en los puntos que se generan al relacionar los límites inferior o superior de cada categoría con las respectivas frecuencias acumuladas “menor que” o “mayor que”.

- Ejemplo

Vamos a continuar trabajando con los datos obtenidos en el estudio anterior, a la distribución de frecuencias original, se adicionarán dos columnas, una para la frecuencia acumulada absoluta “menor que” y otra para la frecuencia acumulada porcentual “mayor que”, con estos datos adicionales se trazarán las ojivas respectivas.

La primera será una ojiva absoluta, ya que estamos trabajando con la frecuencia acumulada absoluta y la segunda una ojiva porcentual, por cuanto las frecuencias están expresadas en forma porcentual.

Tabla 13 gastos semanales familiares en alimentación

LÍMITES				
INFERIOR	**SUPERIOR**	**f**	**fa”<”**	**fa%”>”**
0	50	10	10	100
50	100	26	36	90
150	150	24	60	64
200	200	17	77	40
250	250	13	90	23
300	300	8	98	10
350	350	2	100	2

Elaboración: Autores

Para trabajar en Excel los datos de esta tabla deben estar en una hoja de cálculo, a fin de escoger las columnas respectivas y el tipo de gráfico que se quiere generar, para el presente estudio, se trabajará con las columnas de las frecuencias acumuladas una por una para conseguir la graficación de cada ojiva. Otra opción es utilizar un solo gráfico, donde consten las dos ojivas, pero para ello las dos frecuencias acumuladas deben estar expresadas en forma absoluta o porcentual. (Solo en este Ejemplo coinciden las frecuencias, toda vez que el total de elementos estudiados es 100). Los gráficos generados se los puede apreciar en las figuras.

Figura 10 Ojiva porcentual “ mayor que”

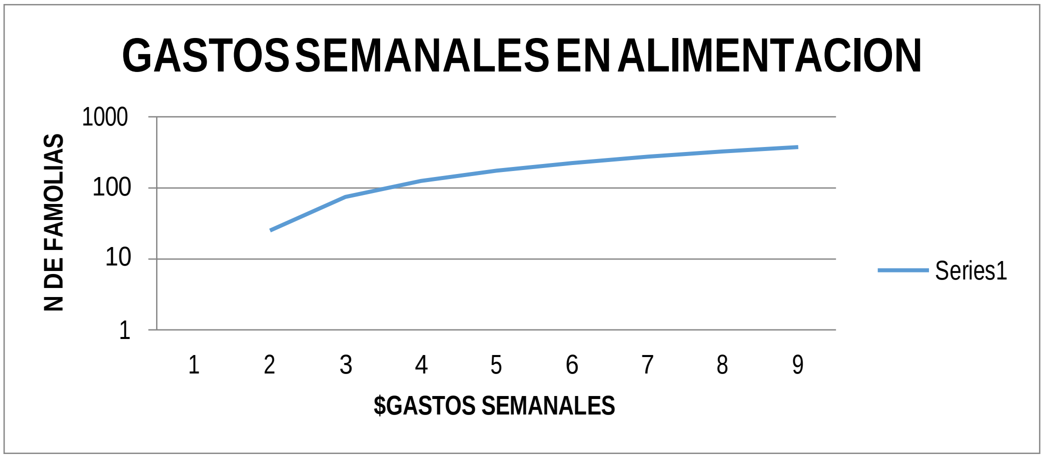
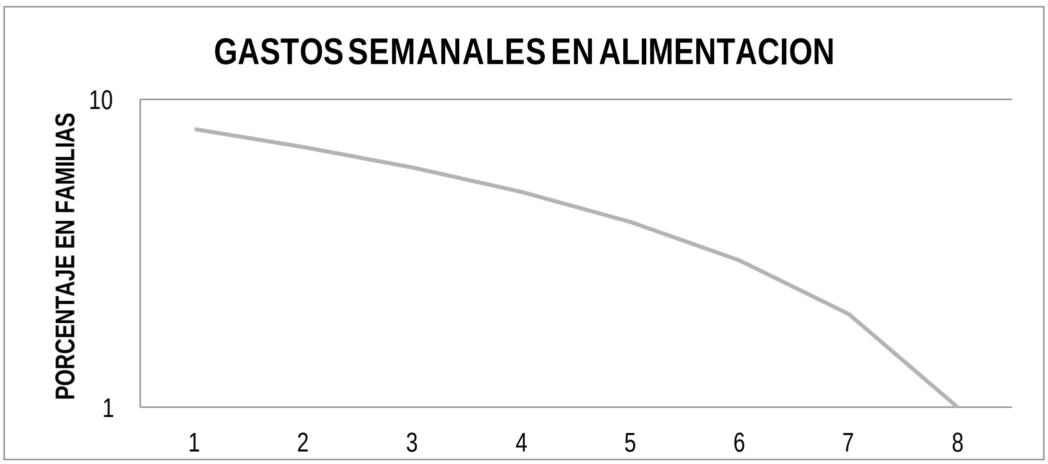


Figura 11 Ojiva "menor que"



Elaboración: Autores

Aplicaciones

1. Trace un diagrama de columnas para la siguiente distribución de frecuencias:

NÚMERO DE AUTOMÓVILES VENDIDOS POR MARCA EN EL MES DE MAYO

Tabla 14 número de automóviles vendidos en mayo

MARCA	CANTIDAD
CHEVROLET	115
KIA	72
HYUNDAI	63
TOYOTA	48

NISSAN	45
MAZDA	50
PEUGEOT	31
RENAULT	28
VOLSWAGEN	25
OTRAS	30

Elaboración: Autores

2. Para el siguiente conjunto de datos trace un diagrama de barras

RENDIMIENTO (%) DE LAS ACCIONES DE EMPRESAS EN LOS AÑOS 2003-2004-2005

Tabla 15 rendimiento (%) de las acciones de empresas en los años 2003-2004-2005

Empresa	**AÑOS**		
	2003	**2004**	**2005**
COMPAÑÍA1	7,5	8,8	12,3
COMPAÑÍA2	10,8	9,5	14,5
COMPAÑÍA3	8,9	8,3	10,5
COMPAÑÍA4	12,3	13,4	14,1
COMPAÑÍA5	15,6	15,1	15,4
COMPAÑÍA6	16,8	17	16,8
COMPAÑÍA7	9,3	10,8	13,4

Elaboración: Autores

Con los datos del ejercicio trace además un diagrama de barras apiladas y comente los resultados obtenidos.

3. Para los siguientes datos elabore un diagrama de sectores.

Estudiantes de último año de cinco facultades

Tabla 16 Estudiantes de último año de cinco facultades

Facultad	N° estudiantes
Derecho	180

Medicina	157
Ciencias	148
Administración	110
Filosofía y Letras	95

Elaboración: Autores

4. En la Carrera de Ciencias Económico Administrativas de la Unap, se tomó una muestra de quince estudiantes, a los cuales se les registro las notas obtenidas en las materias de Matemáticas 1 y Estadística Matemática, a fin de realizar un estudio que determine si existe alguna relación entre estas dos asignaturas. Con los datos que constan a continuación, trace un diagrama de dispersión, analícelo y exponga un criterio sobre si existe alguna relación.

ESTUDIO DE UNA MUESTRA DE LAS NOTAS OBTENIDAS EN MATEMÁTICAS 1 Y ESTADÍSTICA DE 15 ESTUDIANTES

Tabla 17 estudio de una muestra de las notas obtenidas en matemáticas 1 y estadística de 15 estudiantes

Est	matemáticas 1	ESTADÍSTICA. MATEMÁTICA
1	75	72
2	80	80
3	78	75
4	72	78
5	85	82
6	90	92
7	91	89
8	70	78
9	81	82
10	82	80
11	86	90
12	78	80
13	79	81
14	82	90
15	95	92

Elaboración: Autores

5. Con los datos que en la siguiente tabla constan, realizar distribuciones de frecuencia para cada fabricante, posteriormente elaborar los siguientes gráficos:

a. Histogramas respectivos
b. Polígono de frecuencias relativas (en un mismo gráfico deben constar los dos estudios, afín de poder comparar el comportamiento de cada fabricante).
c. Ojivas porcentuales "menores que" de ambos fabricantes.
d. Realizar un pequeño informe sobre los gráficos obtenidos, señalando lo más notorio de cada estudio y estableciendo las debidas comparaciones que puedan darse.

MUESTRA DE LA DURACIÓN (HORAS) DE FOCOS DE 100 W. DE DOS FABRICANTES.

Tabla 18 muestra de la duración (horas) de focos de 100 w. De dos fabricantes.

FABRICANTE "A "				FABRICANTE "B"			
684	860	926	1005	819	959	1020	1113
831	899	946	1080	907	1004	1077	1188
859	924	984	821	952	1018	1113	903
893	943	1052	852	994	1072	1174	943
922	977	773	876	1016	1100	897	992
939	1041	852	911	1038	1154	942	1015
972	720	870	938	1096	888	986	1034
1016	848	909	971	1153	918	1007	1082
697	868	926	1014	836	962	1022	1116
835	905	954	1093	912	1005	1077	1230
1025	698	852	745	715	854	698	738
1158	785	870	735	700	855	710	1050
980	1350			874	705		

Elaboración: Autores

6. TIPOS DE VARIABLES

Clasificar las siguientes variables en cualitativas y cuantitativas discretas o continuas.

1. La nacionalidad de una persona.

    Cualitativa

2. Número de litros de agua contenidos en un depósito.

    Cuantitativa continúa.

3. Número de libro en un estante de librería.

    Cuantitativa discreta.

4. Suma de puntos tenidos en el lanzamiento de un par de dados.

    Cuantitativa discreta.

5. La profesión de una persona.

    Cualitativa.

6. El área de las distintas baldosas de un edificio.

    Cuantitativa continúa.

b. Análisis y descripción de los datos

En la unidad anterior nos preocupamos de aprender a resumir la información proporcionada por un estudio estadístico a través de tablas estadísticas o distribuciones de frecuencias y gráficos estadísticos acordes a la naturaleza del estudio. En esta unidad entraremos a un proceso que nos permita describir a un conjunto de datos categorizados o no, mediante el cálculo de medidas, tales como promedios o medidas de centralización que nos brindan información sobre el valor o valores que se ubican generalmente en el centro de los datos ordenados; medidas de variabilidad o dispersión, que proporcionan información referente a cuán disperso se hallan los datos frente a una medida de tendencia central, esto nos faculta a determinar si las medidas anteriores son o no representativas del estudio y medidas de forma que contribuirán a informarnos sobre la cómo los datos se encuentran frente a estudios referentes tales como la distribución normal.

Medidas de tendencia central

Al describir grupos de observaciones, con frecuencia es conveniente resumir la información con un solo número. Este número que, para tal fin, suele situarse hacia el centro de la distribución de datos se denomina medida o parámetro de tendencia central o de centralización. Cuando se hace referencia únicamente a la posición de estos parámetros dentro de la distribución, independientemente de que ésta esté más o menos centrada, se habla de estas medidas como medidas de posición. En este caso se incluyen también los cuantiles entre estas medidas. Se debe tener en cuenta que existen variables cualitativas y variables cuantitativas, por lo que las medidas de posición o medidas de tendencia se usan de acuerdo al tipo de variable que se está observando, en este caso se observan *variables cuantitativas*.

Entre las medidas de tendencia central tenemos:

- Media .
- Media ponderada.
- Media geométrica.
- Media armónica.
- Mediana.
- Moda.

1. Medidas de tendencia central de datos no agrupados

Las principales medidas de centralización que vamos a estudiar en este acápite son: media aritmética, mediana, moda y media geométrica. Existen otras medidas que no se aplican dentro del ámbito administrativo, por lo tanto no serán estudiadas. De las tres primeras medidas, posiblemente la media aritmética sea la más utilizada y como se verá más adelante, es fundamental en otros muchos estudios y cálculos estadísticos. Por otra parte la media geométrica tiene importancia específica sobretodoenestudioseconómicosyfinancieros.

Media Aritmética

Paraestamedidavamosaconsiderardosalternativas,noporqueelconceptoyelcriteriode cálculosean diferentes,másbienrespondeacondicionessobre,silamediaes obtenida de unapoblación o de una muestra, lo único que cambiará es la simbología utilizada para la representación (μ: media aritméticadepoblación)mediaaritméticademuestra.

En general podemos indicar que, la media aritmética es el valor que resulta de dividir la suma de todos los valores observados entre el número de datos considerados. Utilizando un lenguaje simbólico, se tiene lo siguiente:

$$\mu = \frac{\sum x}{N} = \frac{x_1 + x_2 + x_3 + \cdots .. + x_N}{N}$$

$$\bar{x} = \frac{\sum x}{n} = \frac{x_1 + x_2 + x_3 + \cdots .. + x_n}{n}$$

Dónde:

Σ: sumatoria (letra mayúscula sigma)

N: Tamañodepoblación(númerodeelementosdelapoblación) n:

Tamañodelamuestra(númerodeelementosdelamuestra) μ: media

aritméticadepoblación(letragriegamu)

$\bar{x}$: media aritmética de muestra (equis barra)

Esta medida de tendencia central es la más utilizada e inclusive el lector ya debe estar familiarizado con ella, toda vez que el lector en su avance estudiantil desde la educación básica obtuvo “promedios” de sus calificaciones, que no fueron sino la mediaaritméticadesusaportesy evaluaciones.

Ejemplo 1:

En una muestra de diez envases de refrescos se obtuvieron los siguientes valores (cm^3): 251, 248.5, 250.8, 249.7, 249, 251.2, 248.8, 249.2, 250.5, 249.3, determinar el contenido medio de esta muestra:

$$\bar{x} = \frac{\Sigma x}{n} = \frac{251,248.5,250.8,249.7,249,251.2,248.8,249.2,250.5,249.3}{10} = \frac{2498}{10} = 249.8\ cm^3$$

Propiedades de la media aritmética:

1. Los datos medidos en escala de intervalo o de razón, tienen una media aritmética.

2. El valor de la media aritmética es único, es decir, un conjunto de datos tiene un solo valor de media aritmética.

3. Para el cálculo de la media aritmética se consideran todos los datos observados. Esta propiedad determina que la media aritmética sea sensible a la presencia de valores extremos.

4. Es una medida muy útil cuando se necesita comparar estudios estadísticos de la misma naturaleza.

5. La media aritmética es la única medida de tendencia central, donde la suma de las desviaciones de los elementos con respecto a ella, siempre es cero.

Expresado simbólicamente:

$$\Sigma(x - \bar{x}) = 0$$

Ejemplo 1:

La media de 2, 10 y 3, es 5.

Entonces:

$$\Sigma(x - \bar{x}) = (2 - 5) + (10 - 5) + (3 - 5) = -3 + 5 - 2 = 0$$

Es necesario señalar que la media aritmética tiene algunas desventajas, entre ellas la principal es que al ser afectada por la presencia de valores extremos (altos o bajos), pierde representatividad del conjunto de datos. Como Ejemplo de esta desventaja podemos citar al siguiente: suponga que un estudiante obtuvo las siguientes 4 notas que fueron evaluadas sobre 100. 86, 90, 94 y 6, por lo tanto la media aritmética o promedio de estas notas es: 69, resulta obvio que el valor medio no es representativo de la notas obtenidas, ya que tres de las cuatro notas están dentro de un intervalo comprendido entre 86 y 94. En este caso el extremo inferior afecta completamente al valor de la media. Si el conjunto de datos ha sido categorizado mediante intervalos y uno de ellos es abierto, no se podrá determinar la media aritmética.

Media Aritmética Ponderada: Constituye un caso especial de la media aritmética y ocurre cuando los datos individuales, están categorizados de acuerdo a la frecuencia o factores de ponderación.

En estos casos la variable está representada por cada valor observado y los pesos constituyen las frecuencias o los factores de ponderación de cada uno de ellos. Para este caso, el cálculo de la media aritmética se reduce a encontrar la suma de los productos de cada valor observado con su respectiva frecuencia y dividirla entre la suma de las frecuencias.

La siguiente expresión simboliza el cálculo descrito.

$$\bar{x} = \frac{\Sigma(x * f)}{\Sigma f} = \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 + \cdots .. + x_n f_n}{f_1 + f_2 + f_3 + \ldots \ldots . + f_n}$$

Donde "**x**" representa la variable y "**f**" frecuencia.

Ejemplo 1:

Un negocio de refrescos vende tres tipos de contenidos, pequeños, medianos y grandes a 40, 60 y 80 centavos de dólar cada uno, en un día en particular vendió 50 pequeños, 45 medianos y 60 grandes, se quiere determinar cuál es el precio promedio de venta de cada refresco.

TIPO	P.VENTA Centavos x	Nº Refrescos f	X*f
Pequeño	40	50	2000
Mediano	60	45	2700
Grande	80	60	4800
TOTAL		**155**	**9500**

$$\bar{x} = \frac{\Sigma(x * f)}{\Sigma f} = \frac{9500}{155} = 61,29\ \text{ctvs}.$$

Mediana

Es el punto medio del total de observaciones, luego de que han sido ordenados y que deja al mismo número de observaciones por debajo de su valor, así como por arriba de él.

La mediana es una importante medida de ubicación, en casos en que la media aritmética no es representativa de un conjunto de datos, esta situación se da cuando existe la presencia de valores extremos altos o bajos, en cuyo caso la **mediana** proporciona un valor más representativo de la tendencia central. Para la determinación de la mediana es necesario que los datos se encuentren previamente ordenados y su valor coincide con aquel que deja el mismo número de observaciones por debajo y por Encima de él.

Para la determinación de la mediana, únicamente se recurre a la determinación del valor medio, existen dos posibilidades, cuando el número de observaciones es impar y cuando este número es par. En el primer caso la ubicación del elemento central es directa escogiendo el elemento que ocupa la posición ***(n + 1) / 2***. Es decir al total de observaciones se le suma una unidad y a este resultado se lo divide entre 2, dando como resultado la ubicación del elemento central. En el segundo caso es necesaria la determinación de dos valores centrales, (**n / 2**) y (**n / 2 + 1**), una vez determinados se encuentra la media aritmética de estos valores, que a su vez constituye la mediana del conjunto de datos.

Ejemplo 1:

Determinar la mediana del siguiente conjunto de datos: 8, 10,

18, 14, 15, 13, 11, 16, 17

Ejemplo 2:

b. Determinar la mediana del siguiente conjunto de datos: 21, 15,

18, 20, 16, 19.

Ordenamos los datos: 8, 10, 11, 13, **14**, 15, 16, 17, 18. Como el número de elementos es impar (***n = 9***), ubicamos al elemento central

(**9 +1**)/**2=5**. De tal manera que debemos escoger el quinto elemento del ordenamiento de los datos, para el caso presente este quinto elemento es el ***14 (mediana).*** Se verifica que este valor de acuerdo al concepto de la mediana, deja por igual el mismo número de elementos por debajo y por encima de él, en el presente Ejemplo, cuatro.

Ordenamos los datos: 15, 16, **18, 19**, 20, 21. Al ser el número de elementos par (**n = 6**), ubicamos los dos valores centrales, que, en este caso son: 18 (**n /2**) y 19 (**n / 2 + 1**). Luego encontramos la media aritmética de estos dos valores, este resultado será la mediana del conjunto de datos. (18 + 19) / 2 = **18.5.**

Se comprueba que existe la misma cantidad de elemento que son menores a la mediana y mayores a ella; en este caso tres.

Propiedades de la mediana:

Al igual que la media aritmética, su valor es único, entonces, un conjunto de datos posee una sola mediana.

No se ve afectada por la presencia de valores extremos bajos o altos, en el caso del Ejemplo anterior en el literal (a.) puede ser el último dato un valor tan alto como se quisiese, que la mediana seguirá siendo la misma.

Puede ser determinada para distribuciones de frecuencia que tengan intervalos abiertos, siempre y cuando la mediana no se encuentre en esa categoría.

Puede determinarse para datos que han sido medidos en escala de intervalo, de razón u ordinal.

Moda

Es el valor de la observación o elemento que tiene la mayor frecuencia.

La moda es otra medida de tendencia central, que es muy útil para describir conjuntos de datos nominales y ordinales y su determinación es sencilla, toda vez que queda fijada por la ubicación del elemento que mayor frecuencia tiene, es decir, el que más veces aparece en el estudio.

En definitiva la moda puede determinarse para cualquier conjunto de datos y al igual que la mediana no se ve afectada por la presencia de valores extremos y puede ser determinada para categorías con intervalos abiertos. Sin embargo la moda tiene una desventaja, la cual hace que no sea muy utilizada, principalmente para datos numéricos y es que muchos estudios no poseen moda no hay elementos con mayor frecuencia o puedan tener varias modas (cuando dos o más elementos tienen la misma mayor frecuencia), dando lugar en este último caso a que los estudios sean bimodales o plurimodales.

Ejemplo 1:

Para los siguientes datos, determinar la moda:

12, 10, 13, 9, 12, 11, 14, 13, 12, 15, 8, 12, 14. Al ordenar los datos obtenemos: 8, 9, 10, 11, **12, 12, 12, 12**, 13, 13, 14, 14, 15, podemos observar que el elemento que mayor frecuencia tiene es el valor 12 el cual se repite 4 veces, por lo tanto es la moda de este conjunto de datos.

32, 30, 28, 29, 31, 33, 35, 36. Igualmente ordenado los datos se obtiene: 28, 29, 30, 31, 32, 33, 35, 36 y podemos observar que no existe ningún elemento que tenga mayor frecuencia, todos tienen frecuencia 1, por lo tanto este conjunto de datos no posee moda.

Media Geométrica

Es de gran utilidad cuando se quiere establecer el promedio de porcentajes, razones, índices o tasa de crecimiento. Su uso es ampliamente demandado en economía y en demografía.

Pues proporciona el cambio porcentual de ventas, sueldos, o cifras como tasa de inflación, crecimiento del Producto Nacional Bruto u otras. La expresión matemática que define a esta medida es:

$$MG = \sqrt[n]{x_1 * x_2 * x_3 * \ldots\ldots * x_n}$$

Entonces podemos indicar que la media geométrica es la raíz enésima del producto de todos los elementos de un estudio, con la condición de que estos elementos deben ser siempre positivos. Se verifica adicionalmente que la media geométrica siempre será menor, o máximo igual a la media aritmética (**MG ≤ µ**).

Ejemplo1:

Las ganancias obtenidas por una empresa en los cuatro últimos años fueron de 8%, 6%, 7.5% y 9%. ¿Cuál es la media geométrica de las ganancias?

$$MG = \sqrt[n]{x_1 * x_2 * x_3 * \ldots\ldots * x_n} = \sqrt[4]{8 * 6 * 7{,}5 * 9} = 7{,}545\%$$

Como una segunda aplicación de la media geométrica podemos citar a la determinación de una tasa promedio de crecimiento en un intervalo de tiempo, cuando se conoce el valor inicial y final del período. Este valor queda determinado a partir del uso de la siguiente expresión:

AUMENTO PORCENTUAL PROMEDIO EN UN PERIODO DADO

$$MG = \left(\sqrt[n]{\frac{Valor\ final}{Valor\ inicial}} - 1\right) * 100$$

Ejemplo 2:

Suponga que una ciudad en el año 1985, tuvo 250000 habitantes y en el año 2010 400000 pobladores, determinar cuál es la tasa promedio de crecimiento promedio anual de la población de esta ciudad.

Valor final: 400000

Valor inicial: 250000

Duración del período (n): 2010 – 1985 = 25

$$MG = \left(\sqrt[25]{\frac{400000}{250000}} - 1\right) * 100$$

Esto significa que la población de esta ciudad ha crecido a una tasa promedio anual de aproximadamente 1.9%.

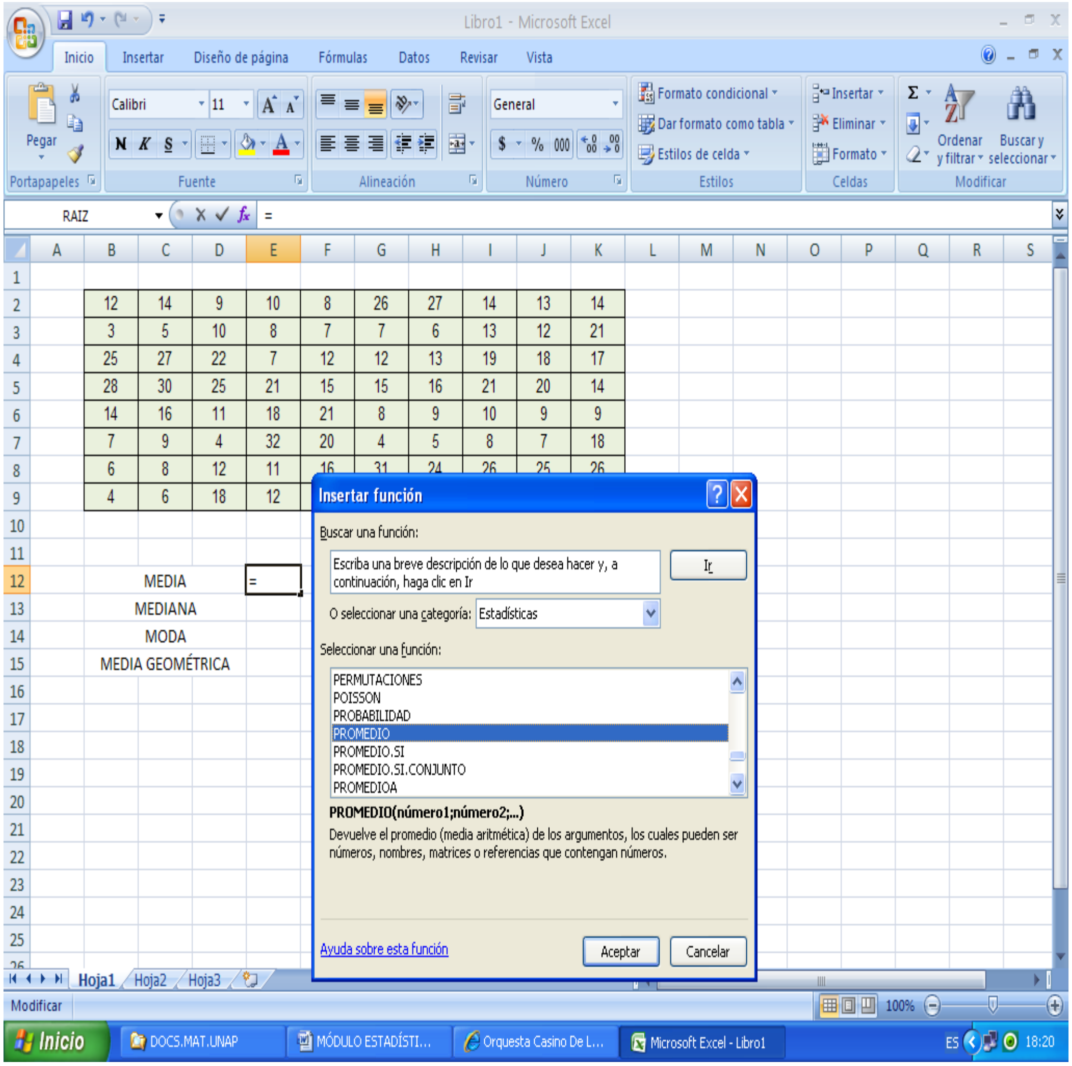
Uso de Excel para determinar las medidas de tendencia central

Si vamos a ocupar la hoja electrónica, o más bien, los datos originales se encuentran consignados en ella, entonces directamente podemos encontrar los valores correspondientes a estas medidas que hemos visto anteriormente. Simplemente se marca una celda donde queremos que aparezca el resultado y escogemos la opción de insertar función, luego se señala el tipo de función, en este caso será “Estadísticas” y aceptamos esta opción, ante lo cual se despliega toda la lista de funciones estadísticas que tiene el programa; en esta instancia con el cursor nos movilizamos hasta encontrar la función requerida. Una vez que se acepta esta función el programa pide que se seleccione el rango de datos, es decir, marcaremos las celdas donde se hallan ubicados, (no es necesario que estén anotados en una sola fila o columna, más bien es preferible que se hallen ubicados en una matriz de varias filas y columnas, especialmente si el número de datos es numeroso) y procedemos a aceptar, inmediatamente el programa nos devolverá el resultado solicitado. Cabe señalar que, para calcular la mediana y la moda no es necesario que los datos se hallen ordenados. La computadora internamente realiza el proceso y nos entrega el resultado.

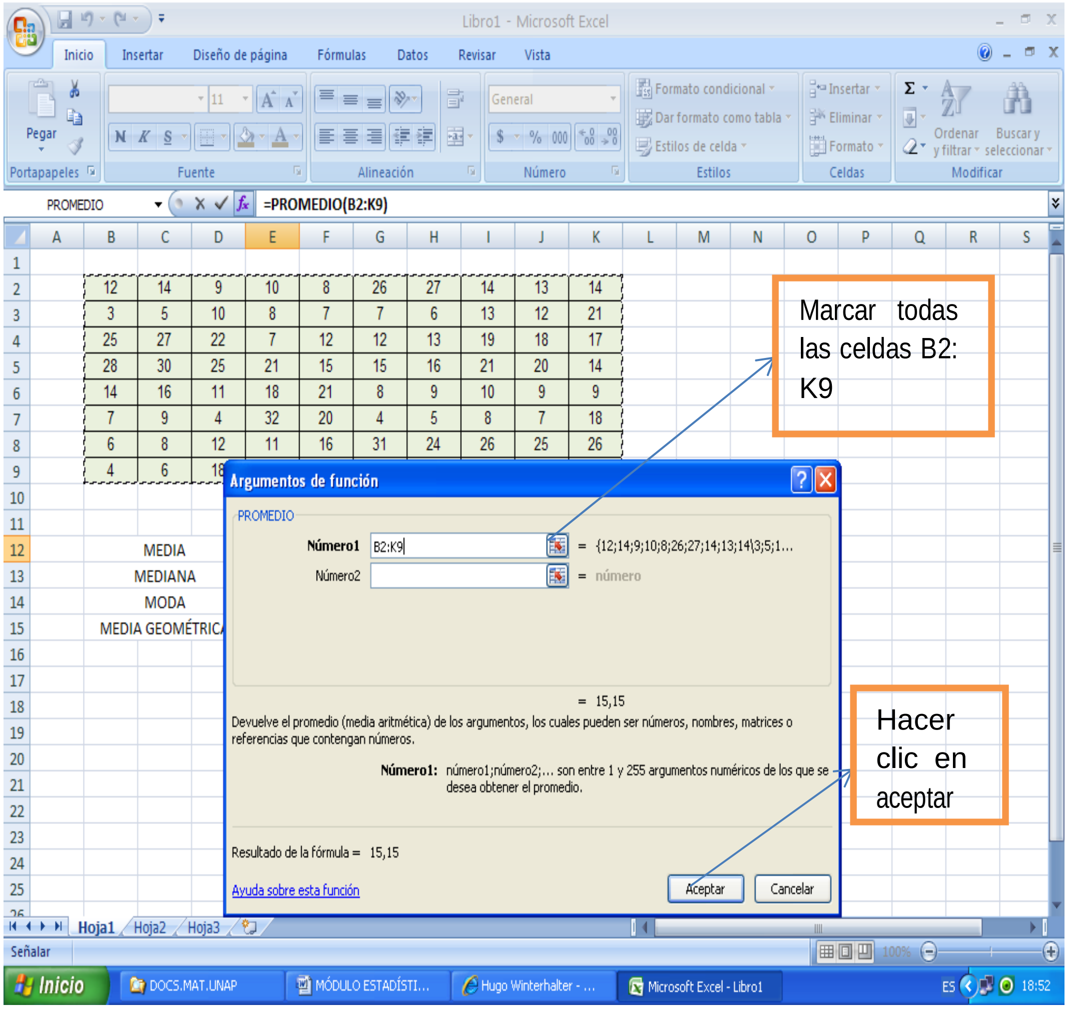
El nombre de las funciones estadísticas estudiadas, constan en el listado de las funciones estadísticas con los siguientes nombres:

MEDIDA	FUNCIÓN
Media Aritmética	Promedio
Mediana	Mediana
Moda	Moda
Media Geométrica	Media Geométrica

A continuación vamos a desarrollar un Ejemplo, en el que se observe lo dicho anteriormente. Se usarán los datos que se hallan en la página 14 y que corresponden al tiempo (minutos) en que un cliente es atendido en una ventanilla bancaria.



Al dar clic en aceptar se presenta el siguiente cuadro de diálogo, pide en la primera ventana que se marque las celdas donde se hallan los datos, a continuación damos un nuevo clic en **Aceptar** y automáticamente se desplegará el resultado en la celda donde se activó la función promedio, esto se ve con claridad en la pantalla que consta a continuación:



Al dar clic en aceptar, en la celda activa aparece el resultado, que para el presente caso es 15.15 minutos. De la misma forma se trabaja para las otras medidas, en cada caso el programa irá determinando el valor de la función seleccionada.

2. Medidas de tendencia central para datos agrupados

Cuando la información ha sido resumida en una tabla de distribución de frecuencias, normalmente se pierde la información de cuáles fueron los datos originales, en tales circunstancias es necesario conocer técnicas que permitan determinar las medidas de tendencia central como la media aritmética, la mediana y la moda, que representen al conjunto de datos que fue categorizado.

Media Aritmética

Para calcular la media aritmética de una distribución de frecuencias cuyas categorías están representadas por intervalos, es necesario primero determinar las marcas de clase o los puntos medios de cada intervalo y luego se aplica la fórmula de la media aritmética ponderada, en este caso los factores de ponderación constituyen las frecuencias de clase y la variable está representada por la marca de clase de cada categoría, entonces la media aritmética se calcula con la siguiente expresión:

$$\bar{x} = \frac{\Sigma(x * f)}{\Sigma f}$$

Dónde:

$\bar{x}$: representa la media aritmética

X: es la marca de clase de cada categoría f: es

la frecuencia de cada categoría

∑(x.f): sumatoria de los productos de las marcas de clase con su frecuencia(n)

∑f: Sumatoria de las frecuencias de clase, que, es igual al total de elementos

Ejemplo 1:

Calcular la media aritmética de la siguiente distribución de frecuencias obtenida de un estudio sobre el precio de ventas de acciones que se negociaron en una semana en la bolsa de valores.

En la siguiente tabla constan los cálculos realizados tales como la obtención de la marca de clase y los productos de la marca de clase de cada categoría con su respectiva frecuencia.

Precio de venta acciones($)		**Nº de acciones(f)**	**Marca de clase(x)**	**x.f**
0	5	310	2,5	775
5	10	430	7,5	3225
10	15	480	12,5	6000
15	20	520	17,5	9100
20	25	500	22,5	11250
25	30	490	27,5	13475
30	35	420	32,5	13650
35	40	370	37,5	13875
40	45	260	42,5	11050
45	50	110	47,5	5225
Total:(Σf)		3890	**Total:(Σx.f)**	87625

$$\bar{x} = \frac{\Sigma(x * f)}{\Sigma f} = \frac{87625}{3890} = \$22.52$$

Este resultado nos proporcionara como valor representativo del estudio$ 22.525, el cual nos indica que en general todas las acciones se vendieron en ese precio promedio.

Mediana

Debemos recordar que esta medida de centralización es una medida posicional que se ubica en el centro de un conjunto de datos ordenados. Si los datos se hallan agrupados, muchas de las informaciones originales no está disponible, en consecuencia no es posible calcular el valor

Para esta estimación es preciso ubicar la categoría donde se encuentra ubicada el elemento mediano (n /2), para proceder a continuación con una interpolación dentro de esta categoría; la razón de este procedimiento radica en la suposición de que todos los elementos que se encuentran en esa categoría, están espaciados de manera uniforme.

La expresión matemática que nos permite calcular este valor estimado es:

$$Mediana = L + \frac{\frac{n}{2} - Fa}{f} * (c)$$

Dónde:

L: Límite inferior de la clase mediana

n: Posición de la mediana (permite ubicar a la categoría donde se halla 2 ubicada utilizando la frecuencia acumulada”<”

FA: Frecuencia acumulada de todas las categorías anteriores a la categoría f:

Frecuencia de la clase media

i: Ancho del intervalo de la categoría mediana.

Ejemplo 1:

Determinar la mediana de los datos categorizados del Ejemplo anterior, referentes a los precios de venta de las acciones en una semana en la bolsa de valores.

Solución: Se procede a calcular la frecuencia acumulada "<", para determinar la categoría mediana, con la ubicación del elemento (n/2); (n/2 = 3890/2 = 1945).

Este elemento se halla ubicado en la quinta categoría, ya que en ella están desde el elemento 1741 hasta el 2240. Esto se visualiza claramente en la siguiente tabla:

Tabla 19 mediana de los datos categorizados

Precio de venta de acciones ($)		Nº de acciones (f)	Frecuencia acumulada
0	5	310	310
5	10	430	740
10	15	480	1220
15	20	520	1740
20	25	500	2240 cat. Mediana
25	30	490	2730
30	35	420	3150
35	40	370	3520
40	45	260	3780
45	50	110	3890

Elaboración: Autores

$$Mediana = L + \frac{\frac{n}{2} - FA}{f} * (i) = 20 + \frac{\frac{3890}{2} - 1740}{500} * (5) = \$22.05$$

La interpretación de este valor nos indica que 1945 precios son menores que $22.05 y 1945 precios son mayores que el valor calculado.

> Observación:
>
> La mediana puede ser calculada para distribuciones de Frecuencias que tienen extremos no determinados, los mismos que generalmente se encuentran en la primera y / o última categoría, y el cálculo de la mediana ocupa solo los datos alrededor de la clase mediana, que normalmente se encuentra aproximadamente en el centro de la distribución.

Moda:

> Generalmente el valor de la moda para datos cuantitativos agrupados tiene un sentido más informativo que representativo, toda vez que bien la media aritmética o la mediana, son valores más representativos de este tipo de estudios categorizados mediante intervalos.

Sin embargo se es necesaria su determinación, ésta se la hace mediante la selección de la marca de clase de la categoría con la mayor frecuencia observada. En caso de existir dos categorías con la misma mayor frecuencia, se escogerán ambas marcas de clase y el estudio será bimodal; si la distribución tiene tres o más categorías con la misma mayor frecuencia, el estudio es primordial, en tal caso estos valores carecen de representantes del estudio. Algunos autores prefieren calcular la moda mediante una fórmula que permite interpolar de una manera más fina. Sin embargo se es necesaria su determinación, ésta se la hace mediante la selección de la marca de clase de la categoría con la mayor frecuencia observada. En caso de existir dos categorías con la misma mayor frecuencia, se escogerán ambas marcas de clase y el estudio será bimodal; si la distribución tiene tres o más categorías con la misma mayor frecuencia, el estudio es plurimodal, en tal caso estos valores carecen de representantes del estudio. Algunos autores prefieren calcular la moda mediante una fórmula que permite interpolar de una manera más fina. La fórmula que se aplica es la siguiente:

$$Moda = L + \frac{d_1}{d_1 + d_2} * (i)$$

Dónde:

L: Límite inferior de la clase modal (categoría con la mayor frecuencia)

d_1: Diferencia entre la frecuencia de la clase modal y la frecuencia de la categoría anterior a la clase moral.

d_2: Diferencia entre la frecuencia de la clase modal y la frecuencia de la categoría anterior a la clase modal.

i : Ancho del intervalo de la clase modal.

Ejemplo 1:

Vamos a determinar la moda para el conjunto de datos trabajados en los Ejemplos anteriores. Se observa en la tabla siguiente que la cuarta categoría es la clase que tiene la mayor frecuencia, por lo tanto esta es la **categoría modal.**

Tabla 20 moda

Precio de venta de acciones ($)		Nº de acciones (f)
0	5	310
5	10	430
10	15	480
15	20	**520 C. modal**
20	25	500
25	30	490
30	35	420
35	40	370
40	45	260
45	50	110

Elaboración: Autores

Primer Procedimiento:

Moda=Marca de clase de la categoría modal

$$Moda = \frac{(15+20)}{2}$$

$$Moda = \$\ 17.50$$

Segundo procedimiento:

$$\square\square\square\square = \square + \frac{\square_1}{\square_1 + \square_2} * (\square) = 15 + \frac{(520-480)}{(520-480)+(520-500)} * (5) = \$18.33$$

Aplicación

1. La Empresa Eléctrica Quito, seleccionó 20 clientes residenciales al azar. Los valores obtenidos del último consumo mensual, expresados en dólares constan a continuación. Determine la media

aritmética, la mediana y la moda de estos datos e indique si las medidas obtenidas son estadísticos o parámetros. ¿Por qué?

54	47	55	50	65	43	47	68	50	70
65	68	38	43	50	74	39	65	48	41

Centiles (porcentiles)

Los percentiles son, tal vez, las medidas más utilizadas para propósitos de ubicación o clasificación de las personas cuando atienden características tales como peso, estatura, etc.

Los percentiles son ciertos números que dividen la sucesión de datos ordenados en cien partes porcentualmente iguales. Estos son los 99 valores que dividen en cien partes iguales el conjunto de datos ordenados. Los percentiles (P1, P2,... P99), leídos primer percentil,..., percentil 99.

Datos Agrupados

Cuando los datos están agrupados en una tabla de frecuencias, se calculan mediante la fórmula:

$$P_k = L_k + \frac{k\left(\frac{n}{100}\right) - F_k}{f_k} * c$$

k= 1,2,3,... 99

Donde:

Lk = Límite real inferior de la clase del decil *k* *n* =

Número de datos

Fk = Frecuencia acumulada de la clase que antecede a la clase del decil *k*. *fk* =

Frecuencia de la clase del decil *k*

c = Longitud del intervalo de la clase del decil *k*

Otra forma para calcular los percentiles es:

Primer percentil, que supera al uno por ciento de los valores y es superado por el noventa y nueve por ciento restante.

$$P = \frac{1n}{100}$$

$$P_1 = l_i + \frac{P - f_{a-1}}{f_1} * I_c$$

El 60 percentil, es aquel valor de la variable que supera al 60% de las observaciones y es superado por el 40% de las observaciones.

$$P_{60} = l_i + \frac{P - f_{a-1}}{f_1} * I_c \quad P = \frac{60n}{100}$$

$$P_{99} = l_i + \frac{P - f_{a-1}}{f_1} * I_c \quad P = \frac{99n}{100}$$

El percentil 99 supera 99% de los datos y es superado a su vez por el 1% restante.

Fórmulas Datos No Agrupados

Si se tienen una serie de valores X1, X2, X3 ... Xn, se localiza mediante las siguientes fórmulas: Para los percentiles, cuando n es par:

$$\frac{A*n}{10}$$

$$\frac{A(n+1)}{100}$$ Cuando n es impar:

Siendo A, el número del percentil.

Es fácil ver que el primer cuartil coincide con el percentil 25; el segundo cuartil con el percentil 50 y el tercer cuartil con el percentil 75.

3. EJEMPLO

Tabla 21 Determinación del primer cuartil, el séptimo decil y el 30 percentil, de la siguiente tabla:

Salarios	No. De	fa
(I. De Clases)	Empleados (f1)	
200-299	85	85
300-299	90	175
400-499	120	295
500-599	70	365
600-699	62	427
700-800	36	463

Elaboración: Autores

Como son datos agrupados, se utiliza la fórmula

$$P = l_i + \frac{P - f_{a-1}}{f_1} * I_c$$

Siendo,

$$P = \frac{n}{4}$$ La posición del primer cuartil.

$$P = \frac{7n}{10}$$

La posición del 7 decil.

$$P = \frac{30n}{100}$$

La posición del percentil 30.

Entonces,

$$\frac{463}{4} = 115.5$$ El primer cuartil:

115.5 – 85 = 30.75

Li = 300, *Ic* = 100 , *fi* = 90

$$Q_1 = 300 + \frac{30.75}{90} * 100 = 334$$

El 7 decil:

$$\frac{7(463)}{10} = \frac{3241}{10} = 324.1$$

Posición:

324.1 – 295 = 29.1

Li = 500, *fi* = 70

$$D_7 = 500 + \frac{29.1}{70} * 100 = 541.57$$

El percentil 30

Posición:

$$\frac{30(463)}{100} = \frac{13890}{100} = 138.9$$

138.9 – 85 = 53.9

fi = 90

$$P_{30} = 300 + \frac{53.9}{90} * 100 = 359.88$$

Estos resultados nos indican que el 25% de los empleados ganan salarios por debajo de $334; que bajo 541.57 gana el 57% de los empleados y sobre $359.88, gana el 70% de los empleados.

3. Medidas de dispersión de datos no agrupados

Estas medidas son necesarias para la mejor comprensión de la distribución de un conjunto de observaciones realizadas en un estudio estadístico y se complementan con las medidas de centralización vistas anteriormente, toda vez que proporcionan conjuntamente una descripción numérica más completa de los datos.

Recordemos que las medidas de tendencia central localizan generalmente a un valor que se halla ubicado en el centro de la distribución, pero no informa sobre el grado de dispersión o variabilidad del conjunto de datos. El análisis de los resultados de estas medidas también permite comparar los grados de dispersión entre dos o más distribuciones.

Las principales medidas de variabilidad que estudiaremos son:

- Rango o amplitud de variación
- Desviación media
- Desviación estándar
- Varianza
- Coeficiente de variación

Amplitud de variación (Rango)

Es la medida más simple de dispersión y se obtiene al establecer la diferencia entre el máximo y el mínimo de los datos cuantitativos.

Amplitud de variación = Máximo – Mínimo

El valor obtenido nos brinda la información en relación al intervalo entre los valores límites en los que se observaron los datos; su utilización está más ligada al control estadísticos de procesos y no es muy utilizada como medida de dispersión, ya que se ve muy influenciada por la presencia de los valores extremos tanto inicial como final.

Ejemplo 1:

1.-Determinar la amplitud de variación de las siguientes edades:

25 43 28 32 27 39 40 29 28 33 36 30

Máximo = 43

Mínimo = 25

Amplitud de variación = 43 – 25 = 18

Este valor nos permite también comparar con estudios de la misma naturaleza y establecer con buen criterio que el conjunto de datos que tenga la menor amplitud de variación, será el que tenga menos variabilidad o menor dispersión.

Desviación media

Es la medida de dispersión que mide más exactamente el grado de dispersión de un conjunto de datos con relación a la media aritmética. En otras palabras es la medida que nos determina en cuantas unidades en promedio los datos se hallan desviados o alejados de la media aritmética.

El uso de esta medida no es muy generalizado por cuanto para su determinación se utiliza el valor absoluto de las desviaciones, esta situación no permite un trabajo algebraico mayor, por lo tanto se utiliza con mayor frecuencia la desviación estándar para representar a la dispersión de los datos frente a la media. Por medio de la siguiente expresión se calcula la desviación media (DM).

$$DM = \frac{\sum |x - \bar{x}|}{n}$$

Dónde:

x: Valor de cada observación

$\bar{x}$: Media aritmética de las

observaciones n: observaciones del

estudio

$| \ |$: Valor absoluto

Ejemplo 1:

En un almacén se determinó en una semana el ingreso de clientes por día, obteniéndose los siguientes resultados, lunes: 250, martes: 265, miércoles: 243, jueves: 225, viernes: 274 y sábado 294. Calcular la desviación media de estos datos.

Solución: Primero se calcula la media aritmética y luego se determina los valores de absolutos de las desviaciones de las observaciones frente a la media aritmética.

$$\bar{x} = \frac{250 + 265 + 243 + 225 + 274 + 294}{6} = 258.5 \text{ clientes}$$

Número de clientes	(X - X̄)	Desviación absoluta
250	250 – 258.5 = - 8.5	8.5
265	265 – 258.5 = 6.5	6.5
243	243 – 258.5 = -15.5	15.5
225	225 – 258.5 = - 33.5	33.5
274	274 – 258.5 = 15.5	15.5
294	294 – 258.5 = 35.5	35.5
Total		**115**

$$DM = \frac{\sum|x - \bar{x}|}{n} = \frac{115}{6} = 19,2 \text{ clientes}$$

La interpretación de este resultado, nos indica que en promedio 19.2 clientes por día están alejados o dispersos los datos obtenidos en este estudio, en relación con la media diaria de visitas que es de 258.5 clientes por día.

Desviación estándar (varianza):

Estas dos medidas de dispersión se basan en los cuadrados de las desviaciones de los elementos con relación a la media aritmética y podemos indicar que la varianza es la media aritmética de las desviaciones cuadráticas con relación a la media aritmética general, mientras que la desviación estándar constituye la raíz cuadrada positiva de la varianza.

Medida	Población	Muestra
Varianza	σ^2	S^2
Desviación estándar	σ	S

σ: letra griega sigma minúscula

La interpretación de la desviación estándar es la misma que se le dio a la desviación media, es decir, proporciona el valor promedio de las desviaciones de los elementos, con relación a la media aritmética; a pesar de que los valores no son iguales. La interpretación para la varianza es más compleja toda vez que las unidades están al cuadrado, sin embargo es una mediad muy útil cuando se comparan estudios estadísticos de la misma naturaleza. Las expresiones de cálculo que nos permitirán determinar sus valores se expresan en la siguiente tabla:

Tabla 22 expresiones de cálculo

MEDIDA	POBLACIÓN	MUESTRA
Varianza	$\sigma^2 = \frac{\sum(x-\mu)^2}{N}$	$s^2 = \frac{\sum(x-\bar{x})^2}{n-1}$
Desviación estándar	$\sigma = \sqrt{\frac{\sum(x-\mu)^2}{N}}$	$s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$

Elaboración: Autores

Como se puede observar en las fórmulas de cálculo de la varianza y desviación estándar de muestra el denominador es algo diferente a las expresiones de la población, ya que al total de elementos de la muestra, se descuenta una unidad, esto se realiza con el objeto de que el estadístico sea un mejor estimador del parámetro.

Las expresiones anotadas anteriormente se basan en el concepto de estas medidas, sin embargo se utilizan operativamente otras expresiones equivalentes que resultan de un manejo algebraico de las anteriores ya que facilitan sustantivamente los cálculos.

Las fórmulas alternativas son:

Tabla 23 expresiones de cálculo

Varianza muestral:

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

Desviación estándar muestral:

$$s = \sqrt{\frac{n\sum x^2 - (\sum x)^2}{n(n-1)}}$$

Elaboración: Autores

Ejemplo 1:

La producción diaria de una fábrica de mesas fue de: lunes 15, martes 18, miércoles 19, jueves 21 y viernes 16, si se considera a estas observaciones como una unidad poblacional, calcular la varianza y la desviación estándar poblacional

Solución:

Calculamos primeramente la media aritmética, para luego proceder a calcular las deviaciones cuadráticas y aplicar la fórmula de cálculo.

$$\mu = \frac{15 + 18 + 19 + 21 + 16}{5} = 17.8 \text{ mesas diarias.}$$

DÍA	Producción (x)	X – μ	$(X – \mu)^2$
Lunes	15	- 2.8	7.84
Martes	18	0.2	0.04
Miércoles	19	1.2	1.44
Jueves	21	3.2	10.24
Viernes	16	-1.8	3.24
TOTAL	79	0	22.80

$$\square^2 = \frac{\square(\square - \square)^2}{\square} = \frac{22.80}{5} = 4.56 \;\square\square\;\square\square\square\square\square\;\square\square\;\square\square\square\square\square\square\square\square$$

$$\square = \sqrt{\frac{\square(\square - \square)^2}{\square}} = \sqrt{\frac{22.8}{5}} = \sqrt{4.56} = 2.14 \;\square\square\square\square\square$$

La interpretación de la desviación estándar, nos indica que en promedio la producción diaria de esta unidad poblacional varía en 2.14 mesas por día. Nótese que la varianza es más difícil de ser interpretada.

4. Medidas de dispersión para datos agrupados

Cuando los datos se encuentran categorizados mediante distribuciones de frecuencia, será necesario que se trabaje con la marca de clase de cada categoría y la frecuencia de clase respectiva para poder calcular las medidas de dispersión, en especial, se pondrá énfasis en la desviación estándar.

Desviación Estándar

Igual que en el cálculo de la desviación estándar de datos no agrupados, podemos utilizar dos procedimientos que responden al criterio conceptual el uno y a la forma simplificada el otro. Se debe en primera instancia determinar la marca de clase de cada categoría con el objeto de calcular las desviaciones de la marca de clase con relación a la media aritmética, o los cuadrados de la marca de clase.

Las siguientes expresiones nos permitirán realizar el respectivo cálculo:

$$\square = \sqrt{\frac{\square(\square - \overline{\square})^2 * \square}{\square - 1}}$$

$$\square = \sqrt{\frac{\square\square(\square - \overline{\square})^2 * \square - (\square\square\square)^2}{\square(\square - 1)}}$$

En el siguiente Ejemplo calcularemos la desviación estándar, siguiendo los dos procedimientos anotados.

Ejemplo:

1.-El reporte de una muestra de 100 envases de refresco sabor a limón

determinó la siguiente distribución de frecuencias, con esta información calcular la desviación estándar, e interpretar los resultados:

Tabla 24 reporte de muestra de 100 envases

CONTENIDO (cm3)	Nº DE ENVASES
497 –498	5
498 –499	23
499 –500	40
500 –501	22
501 –502	10
TOTAL	**100**

Elaboración: Autores

Primer procedimiento: Calculamos las respectivas marcas de clase, para luego calcular la media aritmética de la distribución. Posteriormente se determinan las desviaciones cuadráticas de las marcas de clase, que se las multiplica por la frecuencia de clase.

CONTENIDO (cm^3)	Nº DE ENVASES	MARCA CLASE	*x*f*	*(x-x)²*	*(x-x)²f*
497 – 498	5	497,5	2487,5	4,3681	21,84
498 – 499	23323	498,5	11465,5	1,1881	27,326
499 – 500	40	499,5	19980	0,0081	0,324
500 – 501	22	500,5	11011	0,8281	18,218
501 – 502	10	501,5	5015	3,6481	36,481
TOTAL	**1000**		**49959**		**104,109**

$$\bar{x} = \frac{497.5 * 498.5 * 23 + 499.5 * 40 + 500.5 * 22 + 501.5 * 10}{\square\square\square} = \square\square\square.\square\square\ cm^3$$

$$s = \sqrt{\frac{\sum(x-\bar{x})^2 * f}{n-1}} = \sqrt{\frac{104.19}{100-1}} = 1.026\ cm^3$$

Segundo Procedimiento: Se calculan las marcas de clase, a continuación se determinan los cuadrados de estas marcas de clase. Luego se establecen los productos entre las marcas de clase y las frecuencias y los cuadrados de las marcas de clase por las frecuencias.

CONTENIDO (cm^3)	**Nº DE ENVASES**	**MARCA CLASE**	***x^2***	***x*f***	***x^2*f***
497 – 498	5	497,5	247506	2487,5	1237531,3
498 – 499	23	498,5	248502	11465,5	5715551,8
499 – 500	40	499,5	249500	19980	9980010
500 – 501	22	500,5	250500	11011	5511005,5
501 – 502	10	501,5	251502	5015	2515022,5
TOTAL	**100**			**49959**	**24959121**

$$s = \sqrt{\frac{n\sum(x-\bar{x})^2 * f - (\sum xf)^2}{n(n-1)}} = \sqrt{\frac{100(24959121) - (49959)^2}{100(100-1)}} = 1.026\ cm^3$$

Como se puede advertir, los resultados obtenidos en ambos procedimientos son iguales. La interpretación que se da es que los envases tienen en promedio una desviación de más o menos 1,026 cm3 con relación a la media aritmética.

Si necesita determinar la varianza de este conjunto de datos, simplemente obtenemos el cuadrado de la desviación estándar. S2 = (1.026)2 = 1.0524 cm6.

La desviación estándar además de permitirnos conocer el grado de dispersión de un conjunto de datos, tiene otras aplicaciones muy importantes que a continuación se establecen.

Teorema de Chebyshev

Este teorema establece que para cualquier conjunto de datos estadísticos, la proporción (p) mínima de elementos que se hallan ubicados entre la media aritmética y más o menos “k” desviaciones estándares (μ ± k σ) viene dado por la siguiente expresión:

$$P \geq 1 - \frac{1}{K^2} \quad \text{donde } k > 1$$

Específicamente el teorema determina que el al menos el 75% de los datos encuentran en el intervalo[μ-2σ;μ+2σ], porque si k=2, entonces 1–1/k² = 1–¼= ¾ = 0.75, que expresado porcentualmente es 75%. Pero hay que tomar en consideración que este valor es un límite inferior; por lo tanto puede ser que un mayor porcentaje de elementos se hallen dentro de este intervalo.

Si aplicamos el teorema en mención (cuando k = 2) al Ejemplo anterior, podemos establecer los límites del intervalo de la siguiente forma:

$\mu = 499.59 cm^3, \sigma = 1.026 cm^3$ μ-

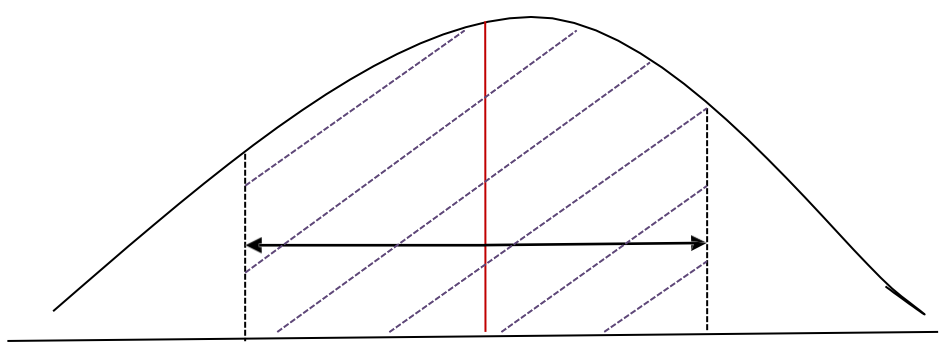
$2\sigma = 497.54\ cm^3$

$\mu + 2\sigma = 501.64\ cm^3$

Podemos observar que estos dos valores son aproximadamente igual a las marcas de clase de la primera categoría (497.5) y la última (501.5), por lo tanto podemos observar que casi el 100% de los envases se encuentran dentro de estos límites. Esto hace que el teorema sea verdadero.

Cabe resaltar que el teorema es válido si la constante (***k***) es mayor a uno, caso contrario el teorema no tiene validez. En la siguiente figura se puede observar el contenido del teorema que fue creado por el matemático ruso Pafnuty L. Chebyshev (1821–1894).

Figura 12 Teorema de Chebyshev



Al menos el 75%

Aplicación del Teorema de Chebyshev para 2σ

Regla empírica

Si los datos se encuentran distribuidos normalmente, la regla empírica provee una mejor aproximación de la concentración de datos entre la media aritmética y un valor de más o menos "***k***" desviaciones estándares. Esta regla se aplica cuando ***k* = 1, 2 o 3.** En una unidad posterior se estudiará con más detalle la distribución normal y sus aplicaciones, donde el valor de ***k*** puede tomar otros valores y no solo los enteros 1, 2, y 3.

Tabla 25 Teorema de Chebyshev

REGLA EMPIRICA	INTERVALO	PORCENTAJE DE ELEMENTOS
	μ±1σ	68%
	μ±2σ	95%
	μ±3σ	99.7%

Elaboración: Autores

Entonces la lectura de estos intervalos es, si los datos se encuentran distribuidos normalmente o se aproximan a esta distribución, el 68% de los datos se hallan entre la media aritmética y más o menos una desviación estándar; el 95% de los elementos observados se encuentran entre la media y más o menos dos desviaciones estándares y el 99.7% de las observaciones entre la media y tres desviaciones estándares.

En la figura que consta a continuación se puede observar con claridad lo mencionado en el párrafo anterior. Se determinó que en cinco semanas los precios de una acción A fueron: 55, 70, 63, 69, 72. Los precios de una acción B en esas mismas cinco semanas fueron: 15, 18, 14,
10, 8.

¿Cómo asesor bursátil cuál acción recomendaría comprar? Calculamos la media aritmética y la desviación estándar de cada acción:

MEDIDAS	ACCIONES	
	A	B
Media aritmética	$ 65,80	$ 13,00
Desviación estándar	$ 6,91	$ 4,00
Coeficiente de variación	10,50%	30,77%

Podemos advertir que si nos centramos en el análisis de la desviación estándar, tomaríamos la decisión de comprar las acciones B, pero al analizar los coeficientes de variación de estos precios, observamos que la acción A tiene menor variabilidad, esto significa que los precios de la acción son más estables y existe menos probabilidad de sufrir pérdidas por descensos bruscos en los precios. Entonces sin lugar a dudas nuestra recomendación será adquirir acciones tipo A.

5. Medidas de forma

Son valores que representan o dan a conocer la manera en que los datos se hallan distribuidos en relación a la media aritmética y toman como referente de comparación a la distribución normal o campana de Gauss, en lo que hace relación al sesgo y a la curtosis. Adicionalmente se toma en consideración al diagrama de caja y bigotes.

Coeficiente de asimetría

Las distribuciones de frecuencia que tienen una representación como la figura de la regla empírica, son distribuciones consideradas como *simétricas*, esto significa que la mitad de la distribución derecha se refleja en la mitad izquierda, tomando como eje de simetría la línea vertical que pasa por la media aritmética.

La asimetría o sesgo se hace presente cuando la distribución carece de simetría, debido a la presencia de valores extremos bien bajos o bien altos. La presencia de estos valores influye en la media aritmética y por lo tanto toma un valor o menor a la mediana o mayor que ella.

La expresión que nos permite calcular el valor del coeficiente de asimetría se lo debe a Karl Pearson, quién desarrolló dos expresiones, una de las más utilizadas es la siguiente expresión:

$$A_s = \frac{3(\bar{x} - Mediana)}{s}$$

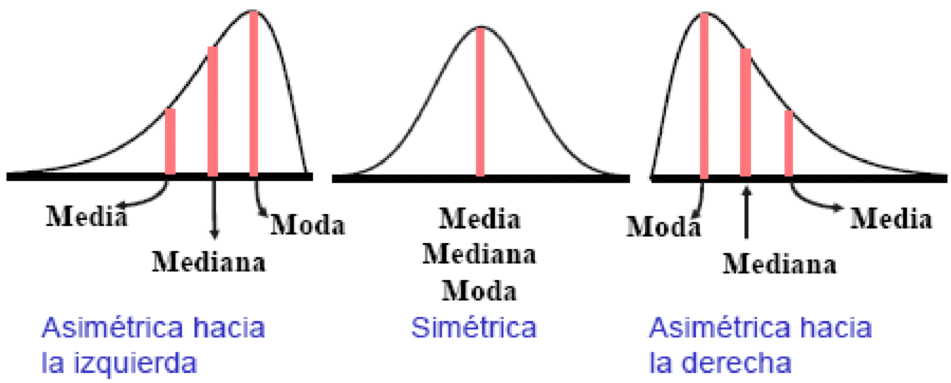
Si la distribución es simétrica, su coeficiente de asimetría es cero, en vista de que la media aritmética es igual que la mediana. Si la media aritmética es mayor que la mediana, entonces la distribución es asimétrica positiva o tiene sesgo positivo; por el contrario si la media es menor que la mediana, entonces la distribución tiene sesgo o asimetría negativa.

Generalmente el coeficiente de asimetría se halla ubicado en el intervalo entre ***- 3*** y ***+ 3***. Si el valor del coeficiente se acerca a los límites indica que la distribución es muy asimétrica, en cambio, si está próximo a cero se trata de una distribución que tiende a ser simétrica

Relación de la media, la mediana y la moda con el sesgo

En la siguiente figura se observa como las medidas de centralización se relacionan entre sí, y con la asimetría de la distribución. Si las tres medidas son iguales la distribución no tiene sesgo; si: **media < mediana < moda**, la distribución es asimétrica negativa y si: **media > mediana > moda**, la distribución tiene sesgo positivo

Figura 13 Relación de la media, la mediana y la moda con el sesgo



Ejemplo 1:

En una muestra de 10 maletas de equipaje, se determinó los siguientes pesos expresados en kilogramos: 25.2, 18.5, 32.7, 28.4, 32.3, 39.7, 23.8, 42.8, 21.5 y 52.3. Determinar el coeficiente de asimetría de estos datos.

Calculamos, la media aritmética, la mediana y la desviación estándar de estos datos, obteniéndose los siguientes resultados.

Media aritmética = 31.72 kg   mediana = 30.35 kg   desviación estándar = 9.57 kg

Entonces el coeficiente de asimetría es:

$$\square\square = \frac{\square(\square - \square\square\square\square\square\square\square)}{\square} = \frac{3(31.72 - 30.35)}{\square.\square\square} = \square.\square\square$$

Este resultado nos indica que los datos se hallan sesgados ligeramente hacia la derecha, por lo tanto tienen asimetría positiva.

Curtosis:

Se encarga de describir el grado de apuntamiento que tiene una distribución, considerando a la distribución normal como referente de comparación, las más altas y que no tienen muchos datos dispersos en las colas, toman el nombre de leptocúrticas, en cambio aquellas que son más bien aplanadas por disponer de datos más dispersos hacia las colas toman el nombre de platicúrticas y aquellas que tienen un apuntamiento como la distribución normal, se denominan mesocúrticas.

La medida que determina este apuntamiento se denomina *curtosis*, la expresión de cálculo que determina el valor de la misma está dado por una expresión que relaciona las desviaciones de cuarto grado de los elementos con la cuarta potencia de la desviación estándar. Nosotros no veremos el cálculo de esta medida por no ser muy utilizada en la cotidianidad. Si el lector está interesado en investigar la forma de cálculo, puede remitirse a cualquier página web relacionada.

En la figura[3] adjunta se observa el grado de apuntamiento de las distribuciones

Figura 14 apuntamiento de las distribuciones

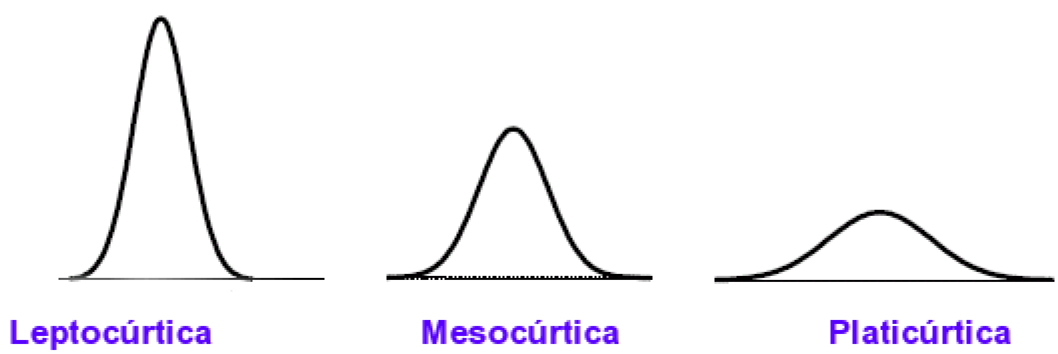


Diagrama de caja y bigotes:

Es un gráfico que también permite observar o describir la forma de un conjunto de datos estadístico. Para su determinación toma en consideración cinco datos:

1. La mediana o cuartil 2
2. El cuartil 1
3. El cuartil 3
4. El máximo (mayor valor observado)
5. El mínimo (menor valor observado)

Trabajo de investigación

En el portafolio entregar un informe de investigación con relación a este tema, haciendo constar Ejemplos de aplicación.

Glosario

Cuartil:	Cada uno de los tres valores que dividen en cuatro partes iguales a un conjunto de datos ordenado
Porcentil o centil:	Cada uno de los 99 valores que dividen en cien
Sesgo:	Presencia de distribuciones que no son simétricas
Curtosis:	Grado de apuntamiento de una distribución
Leptocúrtica:	Distribución alta y con poca dispersión (delgada)
Platocúrtica:	Distribución aplanada y dispersa
Mesocúrtica:	Forma de apuntamiento de la distribución normal

Ejercicios resueltos

Media aritmética para datos no agrupados

La Media Aritmética ($\bar{x}$):

La medida de tendencia central más ampliamente usada es la media aritmética, usualmente abreviada como la media y denotada por $\bar{X}$ (léase como "X barra").

La media aritmética para datos no agrupados

Si se dispone de un conjunto de n números, tales como X1, X2, X3,…,Xn, la media aritmética de este conjunto de datos se define como "la suma de los valores de los ni números , divididos entre n", lo que usando los símbolos explicados anteriormente , puede escribirse como:

$$\bar{X} = \frac{X1, X2, X3, \ldots, Xn,}{n} = \frac{\Sigma X_i}{n}$$

Ejemplo:

Se tienen las edades de cinco estudiantes universitarios de Ier año, a saber: 18, 23, 27, 34 y 25., para calcular la media aritmética (promedio de las edades, se tiene que:

$$\bar{X} = \frac{28+23+27+34+25}{5} = \frac{127}{5} = 25.4 \text{ años}$$

Media aritmética

Lamediaaritméticaeselvalorobtenidoalsumartodoslosdatosydividirelresultadoentreel número total de datos.es el símbolo de la media aritmética.

Ejemplo

Lospesosdeseisamigosson:84,91,72,68,87y78kg.Hallarelpesomedio. Media

aritmética para datos agrupados

Silosdatosvienenagrupadosenunatabladefrecuencias,laexpresióndelamediaes: Ejercicio

de media aritmética

En un test realizado a un grupo de 42 personas se han obtenido las puntuaciones que muestra la tabla. Calcula la puntuación media.

	xi	fi	xi•fi
[10,20)	15	1	15
[20,30)	25	8	200
[30,40)	35	10	350
[40,50)	45	9	405
[50, 60	55	8	440
[60,70)	65	4	260
[70,80)	75	2	150
		42	1820

Mediana

Eselvalorqueocupaellugarcentraldetodoslosdatoscuandoéstosestánordenadosdemenor a mayor.

La mediana se representa por Me.

La mediana se puede hallar sólo para variables cuantitativas. Cálculo

de la mediana

1 Ordenamos los datos de menor a mayor.

2 Silaserietieneunnúmeroimpardemedidaslamedianaeslapuntuación centraldelamisma.

2, 3, 4, 4, 5, 5, 5, 6, 6Me= 5

3 Si la serie tiene un número par de puntuaciones la mediana es la media entre las dos puntuaciones centrales.

7, 8, 9, 10, 11, 12Me= 9.5

Moda

La moda es el valor que tiene mayor frecuencia absoluta. Se

representa por Mo.

Se puede hallar la moda para variables cualitativas y cuantitativas. Hallar la

moda de la distribución:

2, 3, 3, 4, 4, 4, 5, 5 Mo= 4

Si en un grupo hay dos o varias puntuaciones con la misma frecuencia y esa frecuencia es la máxima, la distribución es bimodal o multimodal, es decir, tiene varias modas.

1, 1, 1, 4, 4, 5, 5, 5, 7, 8, 9, 9, 9Mo= 1, 5, 9

Cuando todas las puntuaciones de un grupo tienen la misma frecuencia, no hay moda. 2, 2, 3, 3,

6, 6, 9, 9

Si dos puntuaciones adyacentes tienen la frecuencia máxima, la moda es el promedio de las dos puntuaciones adyacentes.

0, 1, 3, 3, 5, 5, 7, 8Mo = 4

Calculo de la Media Aritmética para datos Agrupados

Propiedades de la media aritmética

1 La suma de las desviaciones de todas las puntuaciones de una distribución respecto a la media de la misma igual a cero.

Las suma de las desviaciones de los números 8, 3, 5, 12, 10 de su media aritmética 7.6 es igual a 0:

8 − 7.6 + 3 − 7.6 + 5 − 7.6 + 12 − 7.6 + 10 − 7.6 =

= 0. 4 − 4.6 − 2.6 + 4. 4 + 2. 4 = 0

2 La media aritmética de los cuadrados de las desviaciones de los valores de la variable con respecto a un número cualquiera se hace mínima cuando dicho número coincide con la media aritmética.

3 Si a todos los valores de la variable se les suma un mismo número, la media

4 Si todos los valores de la variable se multiplican por un mismo número la media aritmética queda multiplicada por dicho número.

Observaciones sobre la media aritmética

1 La media se puede hallar sólo para variables cuantitativas.

2 La media es independiente de las amplitudes de los intervalos.

3 La media es muy sensible a las puntuaciones extremas. Si tenemos una distribución con los siguientes pesos:

65 kg, 69kg , 65 kg, 72 kg, 66 kg, 75 kg, 70 kg, 110 kg.

La media es igual a 74 kg, que es una medida de centralización poco representativa de la distribución.

4 La media no se puede calcular si hay un intervalo con una amplitud indeterminada.

MEDIA ARITMETICA PARA DATOS AGRUPADOS

La media aritmética para datos agrupados

Si los datos se presentan en una tabla de distribución de frecuencias, no es posible conocer los valores individuales de cada una de las observaciones, pero si las categorías en las cuales se hallan. Para poder calcular la media, se supondrá que dentro de cada categoría, las observaciones se distribuyen uniformemente dentro alrededor del punto medio de la clase, por lo tanto puede considerarse que todas las observaciones dentro de la clase ocurren en el punto medio, por lo expuesto la media aritmética para datos agrupados puede definirse de la siguiente manera:

Si en una tabla de distribución de frecuencia, con r clases, los puntos medio son: X1, X2, X3,…,Xn; y las respectivas frecuencias son f1, f2, f3,…, fn, la media aritmética se calcula de la siguiente manera:

$$\overline{X} = \frac{X_1 f_1 + X_2 f_2 + X_3 f_3 + \ldots + X_n f_n}{f_1 + f_2 + f_3 + \ldots + f_n} = \frac{\Sigma(X_i f_i)}{\Sigma f_i} = \frac{\Sigma(X_i f_i)}{N}$$

donde: N = número total de observaciones, por tanto Σfi puede simplificarse y escribirse como N (N= Σfi)

Ejemplo:

Si se toman los datos del Ejemplo resuelto al construir la tabla de distribución de frecuencia de las cuentas por cobrar de **Cabrera's y Asociados** que fueron los siguientes:

Clases 1 2 3 4 5 6

Puntos Medios (Xi) 14,628 29,043 43.458 57,873 72.288 86.703

Frecuencias (fi) 10 4 5 3 3 5

Al calcular la cuenta promedio por cobrar (media aritmética) de estos datos se tiene lo siguiente:

$$\bar{X} = \frac{16.628(10) + 29,043(4) + 43.458(5) + 57.873(3) + 72.288(3) + 88.703(5)}{10 +4 + 5 + 3 + 3 + 5}$$

$$\boldsymbol{N = \Sigma f_i = 30}$$

$$\bar{X} = \frac{146,28 + 116.172 + 217.29 + 173.619 + 216.864 + 433.515}{30}$$

$$\bar{X} = \frac{1303.74}{30} = \boldsymbol{43.458\ balboas}$$

Cálculo de la mediana para datos agrupados

La mediana se encuentra en el intervalo donde la frecuencia acumulada llega hasta la mitad de la suma de las frecuencias absolutas.

Es decir tenemos que buscar el intervalo en el que se encuentre .

Li-1 es el límite inferior de la clase donde se encuentra la mediana es la semisuma de las frecuencias absolutas.

Fi-1 es la frecuencia acumulada anterior a la clase mediana. ai es la

amplitud de la clase.

La mediana es independiente de las amplitudes de los intervalos.

Ejemplo

Calcular la mediana de una distribución estadística que viene dada por la siguiente tabla: fi Fi

[60, 63)	5	5
[63, 66)	18	23
[66, 69)	42	65
[69, 72)	27	92
[72, 75)	8	100
	100	

100/2=50

Clase modal: [66, 69)

Moda en Datos Agrupados

Para determinar la moda de datos agrupados en clases de igual tamaño su cálculo se puede realizar de la siguiente forma**:**

$$Mo = L_i + \frac{\Delta f_i}{\Delta f_i + \Delta f_s} A$$

Donde

$L_i =$ límite inferior o frontera inferior.

$\Delta f_i =$ Exceso de la frecuencia modal sobre la clase modal inferior inmediata.

$\Delta f_s =$ Exceso de la frecuencia modal sobre la clase modal superior inmediata.

$A =$ Anchura o intervalo de la clase modal.

En ocasiones la expresión para el cálculo de la moda suele presentarse de la siguiente forma**:**

$$Mo = L_i + \frac{f_m - f_{(m-1)}}{2f_m - f_{(m-1)} - f_{(m+1)}} A$$

Donde

$f_m =$ Frecuencia de clase modal

$f_{(m-1)} =$ Frecuencia de clase premodal

$f_{(m+1)} =$ Frecuencia de clase posmodal

Aunque la expresión se ve un poco diferente en realidad se trata de una misma ecuación, ya que el exceso de la clase modal inferior se puede determinar como:

$$\Delta f_i = f_m - f_{(m-1)}$$

y el exceso de la clase modal superior se determina como

$$\Delta f_s = f_m - f_{(m-1)}$$

por lo que basta sustituir estos valores en una de ellas para encontrar la otra expresión.

Ejemplo:

Determinar a partir de la tabla presentada, en el Ejemplo de la media, cual es la moda:

Tabla de frecuencias reportadas por la clínica

Clases (Datos en años)	Punto medio de cada clase x_i	Frecuencias de cada clase f_i
$10 \le x < 20$	15	8
$20 \le x < 30$	25	**20**
$30 \le x < 40$	35	14
$40 \le x < 50$	45	8
$50 \le x < 60$	55	2
$60 \le x < 70$	65	2
$70 \le x < 80$	75	1
		55 enfermos atendidos

Identificamos que

$Li = 20;$ $f_m = 20;$ $f_{(m-1)} = 8;$ $f_{(m+1)} = 14;$ $A = 10;$

sustituyendo tenemos

$$Mo = L_i + \frac{f_m - f_{(m-1)}}{2f_m - f_{(m-1)} - f_{(m+1)}} A = 20 + \frac{\square 20 - 8}{2(20) - 8 - 14} = 20.666$$

Pese a que el valor de la moda no pueda constituir un dato real, para el ejercicio, se puede asumir que ese es el parámetro de mayor ocurrencia.

Aportación:

Hola Mi nombre es Pedro Francisco Fuentes Barrientos, soy alumno del ITESM campus Monterrey. Solo escribo para hacerles saber que hay un error en los calculos que publicaron en esta liga: http://dieumsnh.qfb.umich.mx/estadistica/moda.htm , Mientras consultaba su pagina como fuente de estudio note que en la ultima operacion que parece ahi no multiplicaron el valor de la amplitud del intervalo correspondiente a 10. La respuesta verdadera es 26.666666667 en lugar de solo 20.666666...

Medidas de Dispersión para Datos no Agrupados

Las medidas de dispersión nos informan sobre cuánto se alejan del centro los valores de la distribución.

Las medidas de dispersión son:

Rango orecorrido

El rango es la diferencia entre el mayor y el menor de los datos de una distribución estadística. Desviación

media

La desviación respecto a la media es la diferencia entre cada valor de la variable estadística y la media aritmética.

Di = x - x

La desviación media es la media aritmética de los valores absolutos de las desviaciones respecto a la media.

La desviación media se representa por

Ejemplo

Calcular la desviación media de la distribución: 9, 3,

8, 8, 9, 8, 9, 18

Desviación media para datos agrupados

Si los datos vienen agrupados en una tabla de frecuencias, la expresión de la desviación media es:

Ejemplo

Calcular la desviación media de la distribución:

	xi	fi	xi • fi	|x - x|	|x-x|•fi
[10, 15)	12.5	3	37.5	9.286	27.858
[15, 20)	17.5	5	87.5	4.286	21.43
[20, 25)	22.5	7	157.5	0.714	4.998
[25, 30)	27.5	4	110	5.714	22.856
[30, 35)	32.5	2	65	10.174	21.428
		21	457.5		98.57

Varianza

La varianza es la media aritmética del cuadrado de las desviaciones respecto a la media de una distribución estadística.

La varianza se representa por .

Varianza para datos agrupados

Para simplificar el cálculo de la varianza vamos o utilizar las siguientes expresiones que son equivalentes a las anteriores.

Ejercicios de varianza

Calcular la varianza de la distribución: 9, 3,

8, 8, 9, 8, 9, 18

Calcular la varianza de la distribución de la tabla:

	xi	fi	xi • fi	xi2•fi
[10, 20)	15	1	15	225
[20, 30)	25	8	200	5000

[30,40) 35	10	350	12 250
[40, 50) 45	9	405	18 225
[50, 60 55	8	440	24 200
[60,70) 65	4	260	16 900
[70, 80) 75	2	150	11 250
	42	1 820	88 050

Propiedades de la varianza

1 La varianza será siempre un valor positivo o cero, en el caso de que las puntuaciones sean iguales. 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:

Si las muestras tienen distinto tamaño:

Observaciones sobre la varianza

1 La varianza, 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 La varianza no viene expresada en las mismas unidades que los datos, ya que las desviaciones están elevadas al cuadrado.

Desviación típica

La desviación típica es la raíz cuadrada de la varianza.

Es decir, la raíz cuadrada de la media de los cuadrados de las puntuaciones de desviación. La desviación típica se representa por σ.

Desviación típica para datos agrupados

Para simplificar el cálculo vamos o utilizar las siguientes expresiones que son equivalentes a las anteriores.

Ejercicios de desviación típica

Calcular la desviación típica de la distribución: 9, 3,

8, 8, 9, 8, 9, 18

Calcular la desviación típica de la distribución de la tabla:

	xi	fi	xi • fi	xi2 • fi
[10, 20)	15	1	15	225
[20, 30)	25	8	200	5000
[30,40)	35	10	350	12 250
[40, 50)	45	9	405	18 225
[50, 60)	55	8	440	24 200
[60,70)	65	4	260	16 900
[70, 80)	75	2	150	11 250
		42	1 820	88 050

Propiedades de la desviación típica

1 La desviación típica será siempre un valor positivo o cero, en el caso de que las puntuaciones sean iguales.

2 Si a todos los valores de la variable se les suma un número la desviación típica no varía.

3 Si todos los valores de la variable se multiplican por un número la desviación típica queda multiplicada por dicho número.

4 Si tenemos varias distribuciones con la misma media y conocemos sus respectivas desviaciones típicas se puede calcular la desviación típica total.

Si todas las muestras tienen el mismo tamaño: Si

todas las muestras tienen el mismo tamaño:

Si las muestras tienen distinto tamaño:

Observaciones sobre la desviación típica

1 La desviación típica, 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.

Cuartiles Centiles

Las Medidas de Posición, también conocidas como Otras Medidas de Dispersión, son otras medidas o métodos que resultan ser más prácticos para precisar ciertas situaciones en las que se busca describir la variación o dispersión en un conjunto de datos.

Introducción

Cuartiles

Los cuantiles son medidas de posición que se determinan mediante un método que determina la ubicación de los valores que dividen un conjunto de observaciones en partes iguales.

Los cuantiles son los valores de la distribución que la dividen en partes iguales, es decir, en intervalos que comprenden el mismo número de valores. Cuando la distribución contiene un número alto de intervalos o de marcas y se requiere obtener un promedio de una parte de ella, se puede dividir la distribución en cuatro, en diez o en cien partes.

Los más usados son los cuartiles, cuando dividen la distribución en cuatro partes; los deciles, cuando dividen la distribución en diez partes y los centiles o percentiles, cuando dividen la distribución en cien partes. Los cuartiles, como los deciles y los percentiles, son en cierta forma una extensión de la mediana.

Para algunos valores *u* , se dan nombres particulares a los cuantiles, Q (u):

u	*Q(u)*
0.5	Mediana
0.25, 0.75	Cuartiles
0.1,...,0.99	Deciles
0.01,...,0.99	Centiles

CUARTILES

Los cuartiles son los tres valores que dividen al conjunto de datos ordenados en cuatro partes porcentualmente iguales.

Hay tres cuartiles denotados usualmente Q1, Q2, Q3. El segundo cuartil es precisamente la mediana. El primer cuartil, es el valor en el cual o por debajo del cual queda un cuarto (25%) de todos los valores de la sucesión (ordenada); el tercer cuartil, es el valor en el cual o por debajo del cual quedan las tres cuartas partes (75%) de los datos.

Datos Agrupados

$$Q_k = L_k + \frac{k\left(\frac{n}{4}\right) - F_k}{f_k} * c$$

Como los cuartiles adquieren su mayor importancia cuando contamos un número grande de datos y tenemos en cuenta que en estos casos generalmente los datos son resumidos en una tabla de frecuencia. La fórmula para el cálculo de los cuartiles cuando se trata de datos agrupados es la siguiente:

k= 1,2,3

Donde:

Lk = Límite real inferior de la clase del cuartil k n =

Número de datos

Fk = Frecuencia acumulada de la clase que antecede a la clase del cuartil k. fk =

Frecuencia de la clase del cuartil k

c = Longitud del intervalo de la clase del cuartil k

Si se desea calcular cada cuartil individualmente, mediante otra fórmula se tiene lo siguiente:

El primer cuartil Q1, es el menor valor que es mayor que una cuarta parte de los datos; es decir, aquel valor de la variable que supera 25% de las observaciones y es superado por el 75% de las observaciones.

Fórmula de Q1, para series de Datos agrupados:

$$Q_1 = l_i + \frac{P - f_{a-1}}{f_1} * I_c \quad P = \frac{n}{4}$$

Donde:

L1 = limite inferior de la clase que lo contiene

P = valor que representa la posición de la medida

f1 = la frecuencia de la clase que contiene la medida solicitada.

Fa-1 = frecuencia acumulada anterior a la que contiene la medida solicitada.

Ic = intervalo de clase

El segundo cuartil Q2, (coincide, es idéntico o similar a la mediana, Q2 = Md), es el menor valor que es mayor que la mitad de los datos, es decir el 50% de las observaciones son mayores que la mediana y el 50% son menores.

Fórmula de Q2, para series de Datos agrupados:

$$Q_1 = l_i + \frac{P - f_{a-1}}{f_1} * I_c \quad P = \frac{2n}{4}$$

Donde:

L1 = limite inferior de la clase que lo contiene

P = valor que representa la posición de la medida

f1 = la frecuencia de la clase que contiene la medida solicitada.

Fa-1 = frecuencia acumulada anterior a la que contiene la medida solicitada.

Ic = intervalo de clase

El tercer cuartil Q3, es el menor valor que es mayor que tres cuartas partes de los datos, es decir aquel valor de la variable que supera al 75% y es superado por el 25% de las observaciones.

Fórmula de Q3, para series de Datos agrupados:

Fórmula de Q3, para series de Datos agrupados:

$$Q_1 = l_i + \frac{P - f_{a-1}}{f_1} * I_c \quad P = \frac{3n}{4}$$

Donde:

L1 = limite inferior de la clase que lo contiene

P = valor que representa la posición de la medida

f1 = la frecuencia de la clase que contiene la medida solicitada.

Fa-1 = frecuencia acumulada anterior a la que contiene la medida solicitada.

Ic = intervalo de clase.

Otra manera de verlo es partir de que todas las medidas no son sino casos particulares del percentil, ya que el primer cuartil es el 25% percentil y el tercer cuartil 75% percentil.

Para Datos No Agrupados

Si se tienen una serie de valores X1, X2, X3 ... Xn, se localiza mediante las siguientes fórmulas:

- El primer cuartil:

Cuando n es par:

$$\frac{1*n}{4}$$

Cuando n es impar:

$$\frac{1(n+1)}{4}$$

Para el tercer cuartil

Cuando n es par:

$$\frac{3*n}{4}$$

Cuando n es impar:

$$\frac{3*n}{4}$$

DECILES

Los deciles son ciertos números que dividen la sucesión de datos ordenados en diez partes porcentualmente iguales. Son los nueve valores que dividen al conjunto de datos ordenados en diez partes iguales, son también un caso particular de los percentiles. Los deciles se denotan D1, D2,..., D9, que se leen primer decil, segundo decil, etc.

Los deciles, al igual que los cuartiles, son ampliamente utilizados para fijar el aprovechamiento académico.

Datos Agrupados

Para datos agrupados los deciles se calculan mediante la fórmula.

$$D_k = L_k + \frac{k\left(\frac{n}{10}\right) - F_k}{f_k} * c$$

k= 1,2,3,... 9

Lk = Límite real inferior de la clase del decil *k* *n* =

Número de datos

Fk = Frecuencia acumulada de la clase que antecede a la clase del decil *k*. *fk* =

Frecuencia de la clase del decil *k*

c = Longitud del intervalo de la clase del decil *k*

Otra fórmula para calcular los deciles:

El cuarto decil, es aquel valor de la variable que supera al 40%, de las observaciones y es superado por el 60% de las observaciones.

$$D_4 = l_i + \frac{P - f_{a-1}}{f_1} * I_c \quad P = \frac{4n}{10}$$

El quinto decil corresponde a la mediana.

$$D_5 = l_i + \frac{P - f_{a-1}}{f_1} * I_c \quad P = \frac{5n}{10}$$

El noveno decil supera al 90% y es superado por el 10% restante.

$$P = \frac{9n}{10}$$

$$D_9 = l_i + \frac{P - f_{a-1}}{f_1} * I_c$$

Donde (para todos):

L1 = limite inferior de la clase que lo contiene

P = valor que representa la posición de la medida

f1 = la frecuencia de la clase que contiene la medida solicitada.

Fa-1 = frecuencia acumulada anterior a la que contiene la medida solicitada.

Ic = intervalo de clase.

Fórmulas Datos No Agrupados

Si se tienen una serie de valores X1, X2, X3 ... Xn, se localiza mediante las siguientes fórmulas: Si se tienen una serie de valores X1, X2, X3 ... Xn, se localiza mediante las siguientes fórmulas:

$\frac{A*n}{10}$ Cuando n es par:

$\frac{A(n+1)}{10}$ Cuando n es impar:

Siendo A el número del decil.

CENTILES O PERCENTILES

Los percentiles son, tal vez, las medidas más utilizadas para propósitos de ubicación o clasificación de las personas cuando atienden características tales como peso, estatura, etc.

Los percentiles son ciertos números que dividen la sucesión de datos ordenados en cien partes porcentualmente iguales. Estos son los 99 valores que dividen en cien partes iguales el conjunto de datos ordenados. Los percentiles (P1, P2,... P99), leídos primer

Datos Agrupados

Cuando los datos están agrupados en una tabla de frecuencias, se calculan mediante la fórmula:

$$P_k = L_k + \frac{k\left(\frac{n}{100}\right) - F_k}{f_k} * c$$

k= 1,2,3,... 99

Donde:

Lk = Límite real inferior de la clase del decil k n =

Número de datos

Fk = Frecuencia acumulada de la clase que antecede a la clase del decil k. fk =

Frecuencia de la clase del decil k

c = Longitud del intervalo de la clase del decil k

Otra forma para calcular los percentiles es:

Primer percentil, que supera al uno por ciento de los valores y es superado por el noventa y nueve por ciento restante.

$$P = \frac{1n}{100}$$

$$P_1 = l_i + \frac{P - f_{a-1}}{f_1} * I_c$$

El 60 percentil, es aquel valor de la variable que supera al 60% de las observaciones y es superado por el 40% de las observaciones.

$$P_{60} = l_i + \frac{P - f_{a-1}}{f_1} * I_c \quad P = \frac{60n}{100}$$

$$P_{99} = l_i + \frac{P - f_{a-1}}{f_1} * I_c \quad P = \frac{99n}{100}$$

El percentil 99 supera 99% de los datos y es superado a su vez por el 1% restante.

Fórmulas Datos No Agrupados

Si se tienen una serie de valores X1, X2, X3 ... Xn, se localiza mediante las siguientes fórmulas: Para los

percentiles, cuando n es par:

$$\frac{A*n}{10}$$

$$\frac{A(n+1)}{100}$$ Cuando n es impar:

Siendo A, el número del percentil.

Es fácil ver que el primer cuartil coincide con el percentil 25; el segundo cuartil con el percentil 50 y el tercer cuartil con el percentil 75.

3. EJEMPLO

Determinación del primer cuartil, el séptimo decil y el 30 percentil, de la siguiente tabla:

Salarios	No. De	fa
(I. De Clases)	Empleados (f1)	
200-299	85	85
300-299	90	175
400-499	120	295
500-599	70	365
600-699	62	427
700-800	36	463

Como son datos agrupados, se utiliza la fórmula

Como son datos agrupados, se utiliza la fórmula

$$P = l_i + \frac{P - f_{a-1}}{f_1} * I_c$$

Siendo,

$P = \frac{n}{4}$ La posición del primer cuartil.

$$P = \frac{7n}{10}$$

La posición del 7 decil.

$$P = \frac{30n}{100}$$

La posición del percentil 30.

Entonces,

$\frac{463}{4} = 115.5$ El primer cuartil:

$115.5 - 85 = 30.75$

Li = 300, lc = 100 , fi = 90

$$Q_1 = 300 + \frac{30.75}{90} * 100 = 334$$

El 7 decil:

$$\frac{7(463)}{10} = \frac{3241}{10} = 324.1$$

Posición:

324.1 – 295 = 29.1

Li = 500, fi = 70

$$D_7 = 500 + \frac{29.1}{70} * 100 = 541.57$$

El percentil 30

Posición:

$$\frac{30(463)}{100} = \frac{13890}{100} = 138.9$$

138.9 – 85 = 53.9

fi = 90

$$P_{30} = 300 + \frac{53.9}{90} * 100 = 359.88$$

Estos resultados nos indican que el 25% de los empleados ganan salarios por debajo de $334; que bajo 541.57 gana el 57% de los empleados y sobre $359.88, gana el 70% de los empleados.

MEDIDAS DE FORMA:

ASIMETRÍA

Es una medida de forma de una distribución que permite identificar y describir la manera como los datos tiende a reunirse de acuerdo con la frecuencia con que se hallen dentro de la distribución. Permite identificar las características de la distribución de datos sin necesidad de generar el gráfico.

Tipos de asimetría

La asimetría presenta las siguientes formas:

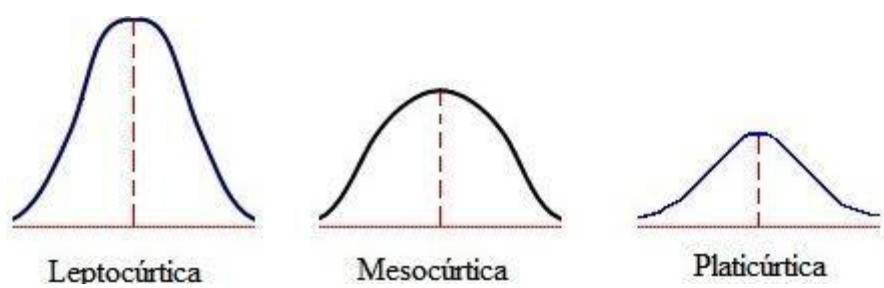
Asimetría Negativa o a la Izquierda.- Se da cuando en una distribución la minoría de los datos está en la parte izquierda de la media. Este tipo de distribución presenta un alargamiento o sesgo hacia la izquierda, es decir, la distribución de los datos tiene a la izquierda una cola más larga que a la derecha. También se dice que una distribución es simétrica a la izquierda o tiene sesgo negativo cuando el valor de la media aritmética es menor que la mediana y éste valor de la mediana a su vez es menor que la moda, en símbolos $\bar{x} < Md < Mo$

Nota: Sesgo es el grado de asimetría de una distribución, es decir, cuánto se aparta de la simetría.

Simétrica.- Se da cuando en una distribución se distribuyen aproximadamente la misma cantidad de los datos a ambos lados de la media aritmética. No tiene alargamiento o sesgo. Se representa por una curva normal en forma de campana llamada campana de Gauss (matemático Alemán 1777- 1855) o también conocida como de Laplace (1749-1827).También se dice que una distribución es simétrica cuando su media aritmética, su mediana y su moda son iguales, en símbolos $\bar{x}$Md=Mo

Asimetría Positiva o a la Derecha.- Se da cuando en una distribución la minoría de los datos está en la parte derecha de la media aritmética. Este tipo de distribución presenta un alargamiento o sesgo hacia la derecha, es decir, la distribución de los datos tiene a la derecha una cola más larga que a la izquierda.

También se dice que una distribución es simétrica a la derecha o tiene sesgo positivo cuando el valor de la media aritmética es mayor que la mediana y éste a valor de la mediana a su vez es mayor que la moda, en símbolos $\bar{x} > Md > Mo$



Medidas de asimetría

Coeficiente de Karl Pearson

$$As = \frac{3(\bar{x} - Md)}{s}$$

Donde:

$\bar{x}$= media aritmética.

Md =Mediana.

s = desviación típica o estándar.

Nota:

El Coeficiente de Pearson varía entre -3 y 3

Si As<0?ladistribuciónseráasimétricanegativa. Si As

= 0 ? la distribución será simétrica.

Si As > 0 ? la distribución será asimétrica positiva.

Medida de Yule Bowley o Medida Cuartílica

$$As = \frac{Q_1 + Q_3 - 2Q_2}{Q_3 - Q_1}$$

Donde:

Q_1= Cuartil uno; Q_2 = Cuartil dos = Mediana; Q_3 = Cuartil tres**. Nota:**

La Medida de Bowley varía entre -1 y 1

Si As<0?ladistribuciónseráasimétricanegativa. Si As

= 0 ? la distribución será simétrica.

Si As > 0 ? la distribución será asimétrica positiva.

Medida de Fisher

Para datos sin agrupar se emplea la siguiente fórmula:

$$As = \frac{\sum(x_i - \bar{x})^3}{n\sigma^3}$$

Para datos agrupados en tablas de frecuencias se emplea la siguiente fórmula:

$$As = \frac{\sum f(x_i - \bar{x})^3}{n\sigma^3}$$

Paradatosagrupadosenintervalosseemplealasiguientefórmula:

Paradatosagrupadosenintervalosseemplealasiguientefórmula:

$$As = \frac{\sum f(xm - \bar{x})^3}{n\sigma^3}$$

Donde:

x_i= cada uno de los valores; n = número de datos; $\bar{x}$= media aritmética; f = frecuencia absoluta

σ^3= cubo de la desviación estándar poblacional; xm = marca de clase

Nota:

Si As < 0 ? Indica que existe presencia de la minoría de datos en la parte izquierda de la media, aunque en algunos casos no necesariamente indicará que la distribución sea asimétrica negativa

Si As = 0 ? la distribución será simétrica

Si As > 0 ? Indica que existe presencia de la minoría de datos en la parte derecha de la media, aunque en algunos casos no necesariamente indicará que la distribución sea asimétrica positiva

Ejemplo ilustrativo:

Calcular el Coeficiente de Pearson, Medida Cuartílica y la Medida de Fisher dada la siguiente distribución: 6, 9, 9, 12, 12, 12, 15 y 17

Solución:

Calculando la media aritmética se obtiene:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{6+9+9+12+12+12+15+17}{8} = \frac{92}{8} = 11{,}5$$

Para calcular los cuartiles se ordena los datos de menor a mayor

6	9	9	12	12	12	15	17
x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8

Calculando el cuartil uno se obtiene:

$$Q_k = X_{\left[\frac{n\cdot k+2}{4}\right]}$$

$$Q_1 = X_{\left[\frac{n+2}{4}\right]} = X_{\left[\frac{8+2}{4}\right]} = X_{\left[\frac{10}{4}\right]} = X_{2{,}5} = \frac{x_2+x_3}{2} = \frac{9+9}{2} = 9$$

Calculando el cuartil dos se obtiene:

$$Q_k = X_{\left[\frac{n\cdot k+2}{4}\right]}$$

$$Q_2 = X_{\left[\frac{n\cdot 2+2}{4}\right]} = X_{\left[\frac{2n+2}{4}\right]} = X_{\left[\frac{2\cdot 8+2}{4}\right]} = X_{\left[\frac{16+2}{4}\right]} = X_{4{,}5} = \frac{x_4+x_5}{2} = \frac{12+12}{2} = 12$$

Calculando el cuartil tres se obtiene:

$$Q_k = X_{\left[\frac{n\cdot k+2}{4}\right]}$$

$$Q_3 = X_{\left[\frac{3n+2}{4}\right]} = X_{\left[\frac{3\cdot 8+2}{4}\right]} = X_{\left[\frac{24+2}{4}\right]} = x_{\frac{26}{4}} = X_{6,5} = \frac{x_6 + x_7}{2} = \frac{12+15}{2} = 13{,}5$$

Calculando la desviación estándar muestral se obtiene:

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

$$s = \sqrt{\frac{(6-11{,}5)^2 + (9-11{,}5)^2 + (9-11{,}5)^2 + (12-11{,}5)^2 + (12-11{,}5)^2 + (12-11{,}5)^2 + (15-11{,}5)^2 + (17-11{,}5)^2}{8-1}}$$

$$s = 3{,}505$$

Calculando el Coeficiente de Pearson se obtiene:

$$As = \frac{3(\bar{x} - Md)}{s} = \frac{3(11{,}5-12)}{3{,}505} = \frac{-1{,}5}{3{,}505} = -0{,}428$$

Calculando la Medida de Bowley se obtiene

$$As = \frac{Q_1 + Q_3 - 2Q_2}{Q_3 - Q_1} = \frac{9 + 13{,}5 - 3\cdot 12}{13{,}5 - 9} = -0{,}333$$

Calculando la desviación estándar poblacional se obtiene:

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}}$$

$$\sigma = \sqrt{\frac{(6-11{,}5)^2 + (9-11{,}5)^2 + (9-11{,}5)^2 + (12-11{,}5)^2 + (12-11{,}5)^2 + (12-11{,}5)^2 + (15-11{,}5)^2 + (17-11{,}5)^2}{8}}$$

$$\sigma = 3{,}279$$

Calculando la Medida de Fisher se obtiene

Datos	$(x_i - \bar{x})^3$
6	-166,375
9	-15,625
9	-15,625
12	0,125
12	0,125

12	0,125
15	42,875
17	166,375
Total	12

$$As = \frac{\sum(x_i - \bar{x})^3}{n\sigma^3} = \frac{12}{8(3{,}279)^3} = 0{,}035$$

Los cálculos en Excel se muestran en la siguiente figura:

	A	B	C	D	E	F	G
1	Datos	$(x_i - \bar{x})^3$					
2	6	-166,375					
3	9	-15,625					
4	9	-15,625					
5	12	0,125					
6	12	0,125					
7	12	0,125					
8	15	42,875					
9	17	166,375					
10	Total	12	=SUMA(B2:B9)				
11	n	8	=CONTAR(A2:A9)				
12	Media aritmética	11,5	=PROMEDIO(A2:A9)				
13	Desviación estándar	3,5050983	=DESVEST.M(A2:A9)				
14	Desviación poblacional	3,2787193	=DESVEST.P(A2:A9)				
15	Cuartil 1	9	=CUARTIL.INC(A2:A9;1)				
16	Cuartil 2	12	=CUARTIL.INC(A2:A9;2)				
17	Cuartil 3	13,5	=CUARTIL.INC(A2:A9;3)+0,25*(A8-A7)				
18	Coeficiente de Pearson						
19–20	$As = \frac{3(\bar{x} - Md)}{s}$	-0,427948	=3*(B12-B16)/B13				
21	Medida de Bowley						
22–23	$As = \frac{Q_1 + Q_3 - 2Q_2}{Q_3 - Q_1}$	-0,333333	=(B15+B17-2*B16)/(B17-B15)				
24	Medida de Fisher						
25–26	$As = \frac{\sum(x_i - \bar{x})^3}{n\sigma^3}$	0,0425577	=B10/(B11*B14^3)				
27	Coeficiente de Asimetría en Excel		0,0530788	=COEFICIENTE.ASIMETRIA(A2:A9)			

Nota: El COEFICIENTE.ASIMETRIA(A2:A9) es un valor que tiene consideraciones semejantes a la Medida de Fisher

2) CURTOSIS O APUNTAMIENTO

La curtosis mide el grado de agudeza o achatamiento de una distribución con relación a la distribución normal, es decir, mide cuán puntiaguda es una distribución.

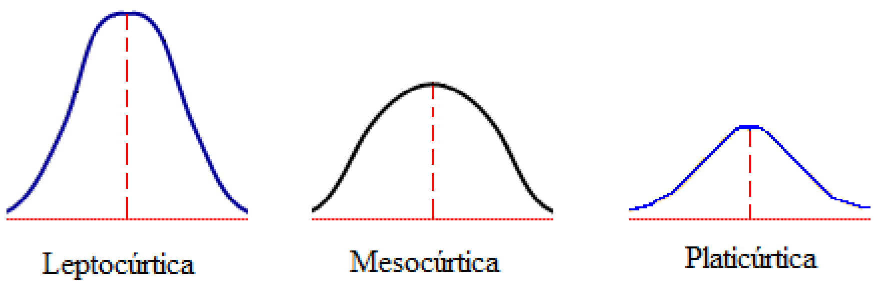
2.1) TIPOS DE CURTOSIS

La curtosis determina el grado de concentración que presentan los valores en la región central de la distribución. Así puedeser:

Leptocúrtica.- Existe una gran concentración.

Mesocúrtica.- Existe una concentración normal.

Platicúrtica.- Existe una baja concentración.



Medidas de curtosis

Medida de Fisher

Para datos sin agrupar se emplea la siguiente fórmula:

$$\alpha = \frac{\sum(x_i - \bar{x})^4}{n\sigma^4}$$

Para datos agrupados en tablas de frecuencias se emplea la siguiente fórmula:

$$\alpha = \frac{\sum f(x_i - \bar{x})^4}{n\sigma^4}$$

Para datos agrupados en intervalos se emplea la siguiente fórmula:

$$\alpha = \frac{\sum f(xm - \bar{x})^4}{n\sigma^4}$$

Donde: x_i=cada uno de los valores; n=número de datos; $\bar{x}$= media aritmética; σ^4 = Cuádruplo de la desviación estándar poblacional; f=frecuencia absoluta; xm=marca de clase

Nota:

Si a < 3 ? la distribución es platicútica

Si a=3 ? la distribución es normal o mesocúrtica Si a >

3 ? la distribución es leptocúrtica

Medida basada en Cuartiles y Percentiles

$$\kappa = \frac{\text{Desviación cuartílica}}{\text{Amplitud cuartílica}} = \frac{\frac{Q_3 - Q_1}{2}}{P_{90} - P_{10}} = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}$$

κ(letra griega minúscula kappa) = Coeficiente percentil de curtosis

Nota:

Si κ< 0,263 ? la distribución es platicúrtica

Si κ=0,263 ? la distribución es normal o mesocúrtica Si κ>

0,263 ? la distribución es leptocúrtica

Esta medida no es muy utilizada.

Ejemplo ilustrativo: Determinar qué tipo de curtosis tiene la siguiente distribución: 6, 9, 9, 12, 12, 12, 15 y 17. Emplear la medida de Fisher y el coeficiente percentil de curtosis.

Solución: Calculando la media aritmética se obtiene

$$\bar{x} = \frac{\sum x_i}{n} = \frac{6+9+9+12+12+12+15+17}{8} = \frac{92}{8} = 11{,}5$$

Calculando la desviación estándar poblacional se obtiene:

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

$$\sigma = \sqrt{\frac{(6-11{,}5)^2 + (9-11{,}5)^2 + (9-11{,}5)^2 + (12-11{,}5)^2 + (12-11{,}5)^2 + (12-11{,}5)^2 + (15-11{,}5)^2 + (17-11{,}5)^2}{8}}$$

$$\sigma = 3{,}279$$

Calculando la Medida de Fisher se obtiene:

Datos	$(x_i - \bar{x})^4$
6	915,0625

9	39,0625
9	39,0625
12	0,0625
12	0,0625
12	0,0625
15	150,0625
17	915,0625
Total	2058,5

$$\alpha = \frac{\sum(x_i - \bar{x})^4}{n\sigma^4} = \frac{2058{,}5}{8 \cdot (3{,}279)^4} = 2{,}23$$

Para calcular los cuartiles y percentiles se ordena los datos de menor a mayor:

6	9	9	12	12	12	15	17
x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8

Calculando el cuartil uno se obtiene:

Calculando el cuartil uno se obtiene:

$$Q_k = X_{\left[\frac{n \cdot k+2}{4}\right]}$$

$$Q_1 = X_{\left[\frac{n+2}{4}\right]} = X_{\left[\frac{8+2}{4}\right]} = X_{\left[\frac{10}{4}\right]} = X_{2,5} = \frac{x_2 + x_3}{2} = \frac{9+9}{2} = 9$$

Calculando el cuartil tres se obtiene:

$$Q_k = X_{\left[\frac{n \cdot k+2}{4}\right]}$$

$$Q_3 = X_{\left[\frac{3n+2}{4}\right]} = X_{\left[\frac{3 \cdot 8+2}{4}\right]} = X_{\left[\frac{24+2}{4}\right]} = x_{\frac{26}{4}} = X_{6,5} = \frac{x_6 + x_7}{2} = \frac{12+15}{2} = 13{,}5$$

Calculando el percentil 90 se tiene:

$$P_k = X_{\left[\frac{n \cdot k+50}{100}\right]}$$

$$P_{90} = X_{\left[\frac{n\cdot 90+50}{100}\right]} = X_{\left[\frac{8\cdot 90+50}{100}\right]} = X_{\left[\frac{770}{100}\right]} = X_{7,7} = \frac{x_7 + x_8}{2} = \frac{15+17}{2} = 16$$

Calculando el percentil 10 se tiene:

$$P_k = X_{\left[\frac{n\cdot k+50}{100}\right]}$$

$$P_{10} = X_{\left[\frac{n\cdot 10+50}{100}\right]} = X_{\left[\frac{8\cdot 10+50}{100}\right]} = X_{\left[\frac{130}{100}\right]} = X_{1,3} = x_1 = 6$$

Calculando el coeficiente percentil de curtosis se obtiene:

$$\kappa = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})} = \frac{13,5 - 9}{2(16 - 6)} = 0,225$$

Como a= 2,23 y $\kappa = 0,225$ la distribución es platicúrtica

Los cálculos en Excel se muestran en la siguiente figura:

	A	B	C	D	E	F
1	Datos	$(x_i - \bar{x})^4$				
2	6	915,06250				
3	9	39,06250				
4	9	39,06250				
5	12	0,06250				
6	12	0,06250				
7	12	0,06250				
8	15	150,06250				
9	17	915,06250				
10	Total	2058,5	=SUMA(B2:B9)			
11	n	8	=CONTAR(A2:A9)			
12	Media aritmética	11,5	=PROMEDIO(A2:A9)			
13	Desviación poblacional	3,2787193	=DESVEST.P(A2:A9)			
14	Cuartil 1	9	=CUARTIL.INC(A2:A9;1)			
15	Cuartil 3	13,5	=CUARTIL.INC(A2:A9;3)+0,25*(A8-A7)			
16	$\alpha = \frac{\sum(x_i - \bar{x})^4}{n\sigma^4}$	2,226609	=B10/(B11*B13^4)			
17						
18	Perceltil 10	7,4	=PERCENTIL.INC(A2:A9;0,1)-0,25*(A3-A2)			
19	Perceltil 90	16,350	=PERCENTIL.INC(A2:A9;0,9)+0,25*(A8-A7)			
20	$\kappa = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}$	0,2500	=(B15-B14)/(2*(B19-B18))			
21						
22						
23	**Curtosis en Excel**	-0,224121	=CURTOSIS(A2:A9)			
24	Valor semejante a la α	2,7758789	=B23+3			

Ejercicios propuestos

EJERCICIO 1

Los miembros de una cooperativa de viviendas tienen las siguientes edades:

4260603860632166565751574445 35

3035475349504938452841474253 32

5438406348333561474155532720 21

4221393934453928543335434848 27

5330295338525427274328634123 58

56 59 60 40 24

Elabore una tabla de frecuencias.

Calculelamediaylladesviacióntípica.

SOLUCIÓN:

Paraelaborarunatabladefrecuenciasescondiciónimprescindibleestablecerunaseriedeclaseso categorías (intervalos) a las que vamos a adjudicar a cada uno de los ochenta miembros de la cooperativa. El investigador puede seguir diferentes criterios en función del objetivo del estudio. Una tabla de frecuencias elaboradaapartirdeestosdatospodríaserlasiguiente:

Edad	**n**
20-29	14
30-39	17
40-49	22
50-59	18
60-69	9
Total	80

Cálculo de la media:

Puede calcularse directamente sumando las edades de todos los miembros de la cooperativa y dividiendo por el total que en este caso es ochenta, el resultado es una media de 43,29. También:

Edad	**x_i**	**n_i**	**$x_i n_i$**
20-29	25	14	350
30-39	35	17	595
40-49	45	22	990
50-59	55	18	990
60-69	65	9	585
Total		**80**	**3510**

$\bar{x} = \frac{3510}{80} = 43{,}875$, por tanto, podemos decir que la media es de casi 44 años.

Cálculo de la desviación típica:

Edad	**x_i**	**n_i**	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 n$
20-29	25	14	-18,875	356,2656	4987,71875
30-39	35	17	-8,875	78,7656	1339,01563
40-49	45	22	1,125	1,2656	27,84375
50-59	55	18	11,125	123,7656	2227,78125
60-69	65	9	21,125	446,2656	4016,39063
Total		**80**			12598,75

$S_{x} = \sqrt{\frac{12598{,}75}{80}} = 12{,}549$

La desviación típica es de 12,5 años

EJERCICIO 2

Explique las similitudes y diferencias de estas distribuciones:

Edad n_	Edadn___
20-29 14	20-29 43
30-39 17	30-39 --
40-49 22	40-49 --
50-59 18	50-59 --
60-69	60-69 37
Total 80	**Total 80**

SOLUCIÓN:

La media y la desviación típica de la primera distribución, ha sido calculada en el primer ejercicio. Calculamos a continuación los mismos estadísticos para la segunda distribución.

Cálculo de la media:

Edad	**x_i**	**n_i**	**$x_i n_i$**
20-29	25	43	1075
30-39	35	-	
40-49	45	-	
50-59	55	-	
60-69	65	37	2405
Total		80	3480

$$\bar{x} = \frac{3480}{80} = 43,5$$

Cálculo de la desviación típica:

Edad	**x_i**	**n_i**	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 n_i$
20-29	25	43	-18,875	356,2656	15319,4219
30-39	35	-	-8,875	78,7656	-
40-49	45	-	1,125	1,2656	-
50-59	55	-	11,125	123,7656	-
60-69	65	37	21,125	446.2656	16511,8281
Total		80			31831,25

$$S_x = \sqrt{\frac{\sum (x_i - \bar{x})^2 n_i}{N}} = \sqrt{\frac{31831,25}{80}} = 19,947$$

La similitud de ambas distribuciones radica fundamentalmente en que tienen la misma amplitud y casi el mismo valor medio. La diferencia es que las frecuencias de la segunda se distribuyen en los intervalos extremos dejando vacíos los del medio. Ello aparece perfectamente reflejado en la desviación típica de 19,9, aproximadamente 20 años. 43 + 20 hacen 63, aproximadamente la mitad del último intervalo, 43 – 20 hacen 23, aproximadamente la mitad del primer intervalo. Recuérdese que la desviación típica es la raíz de la media de las distancias al cuadrado, de cada uno de los elementos de la distribución respecto de la media aritmética.

EJERCICIO 3

En una pregunta del CIS sobre la edad hasta la que consideran convenientes los padres controlar los programas y el tiempo de televisión de los hijos, la media fue de 15,4 años y la desviación típica de 2,11. Teniendo en cuenta que las respuestas se distribuyen aproximadamente como la curva normal y que van de los 7 a los 24 años, calcular:

a)-Cuantos respondieron que la edad debe ser hasta los 13 años b)-

Cuantos dijeron que debe estar entre 14 y 17 años.

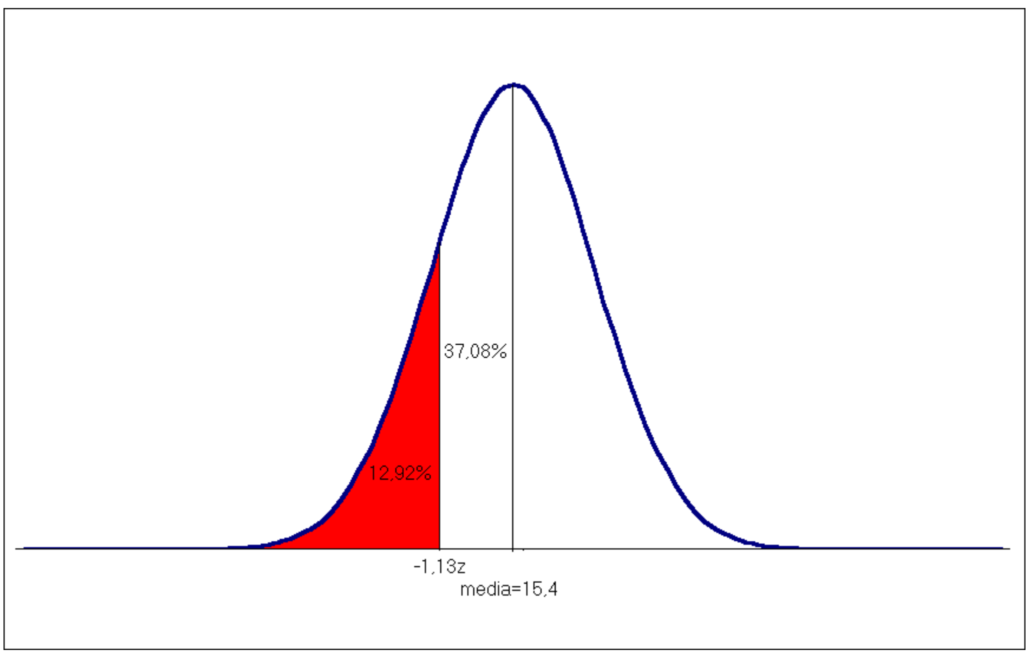
c)- Cuantos respondieron que debe estar por encima de los 19 años

SOLUCIÓN:

a)

$\bar{x} = 15{,}4$

$S_x = 2{,}1$

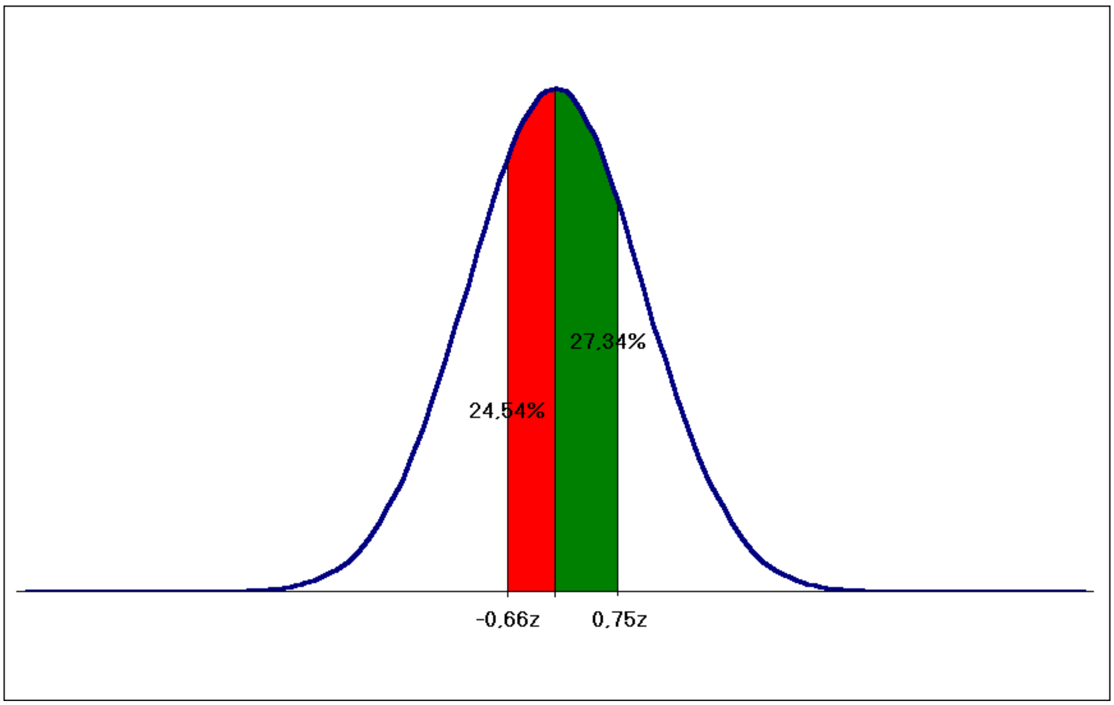


$$z = \frac{x - \bar{x}}{S_x} = \frac{13 - 15{,}4}{2{,}11} = -1{,}13$$

Consultando las tablas de la curva normal comprobamos que entre la media y un desviación típica de 1,13 encontramos un área de 0,3708 que si situaría a la izquierda de la curva por tener signo negativo. Si el área que queremos calcular es el que queda a la izquierda del valor -1,13, es decir, los de menos de 13 años, restamos a 0,5 (que es la superficie de la mitad de la curva) 0,3708 y obtenemos el resultado de 12,92%

0,5-0,3708= 0,1292

b)



$$z_1 = \frac{x - \bar{x}}{S_x} = \frac{14 - 15,4}{2,11} = -0,66$$

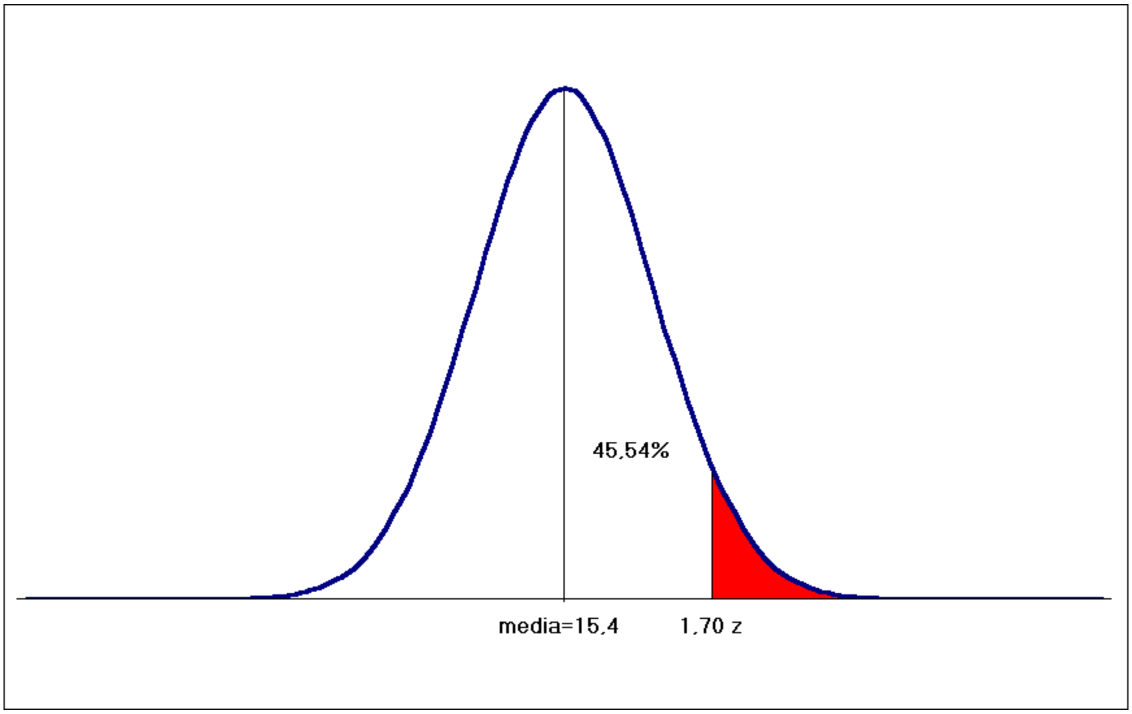
$$z_2 = \frac{x - \bar{x}}{S_x} = \frac{17 - 15,4}{2,11} = 0,75$$

Las áreas correspondientes a estos valores z son 0,2454 y 0,2734 respectivamente.

Como en este caso nos preguntan por el área comprendida entre las unidades z –0,66 y 0,75 sumaremos ambas con el resultado de del 51,88%

0,2454+0,2734 = 0,518

c)



$$z = \frac{x - \bar{x}}{S} = \frac{19 - 15{,}4}{2{,}11} = 1{,}70$$

El área correspondiente es de 0.4554 y los que están por encima de 1,7 unidades z se obtienen restando de 0,5, el 0,4554 de las tablas.

0,5-0,4554 = 0,0446, es decir el 4,46%.

Ejercicio 4

Calcule el tamaño muestral de una encuesta realizada por CIS sobre la Unión Europea que incluía todas las provincias excepto Ceuta y Melilla. El error teórico era de ±2, con un intervalo de confianza de 95,5% y P=Q en el supuesto de un muestreo aleatorio simple.

SOLUCIÓN

Utilizamos la fórmula para muestras infinitas en la que intervienen los tres factores determinantes del tamaño muestral: la probabilidad con la que queremos trabajar (z), el grado de concentración, dispersión de la población (pq) y el error que estamos dispuestos a asumir.

$$n = \frac{z^2 pq}{e^2} = \frac{2^2 \times 0{,}5 \times 0{,}5}{0{,}02^2} = 2.500$$

Resumen

- **Población:** Colección bien definida de elementos que tienen una característica común que interesa ser analizada en un estudio estadístico
- **Muestra:** Parte de la población que se utiliza para ser estudiada en representación de la población
- **Censo:** Estudio de toda la población
- **Muestreo:** Técnica que permite seleccionar muestras de la población
- **Parámetro:** Medida representativa de una población
- **Estadístico:** Medida representativa de una muestra
- **Frecuencia:** Número de veces que se repite un mismo elemento dentro del estudio
- **Distribución de frecuencias:** Tabla estadística que resume la información de un estudio, a través de la categorización de la variable, conjuntamente con el número de elementos que pertenecen a cada categoría mutuamente excluyente y exhaustiva
- **Frecuencia de clase:** Número de elementos que pertenecen a cada clase o categoría
- **Marca de clase:** Es el punto medio de una categoría formada por intervalos
- **Estadística:** Una ciencia que permite recoger, ordenar presentar, analizar y establecer conclusiones sobre el comportamiento de datos que provienen de variables y ayuda a la toma de decisiones.
- **Estadística descriptiva:** rama que permite a treves de métodos, la organización, resumen, presentación e interpretación de los datos de manera informativa.
- **Estadística inferencial**: conjunto de metodologías que permite conocer características de la población, basándose en una muestra
- **Probabilidades**: rama de la matemática que permite articular as dos ramas de la estadística
- **Tipos de variable**:
  - **Cualitativa**: determina características no numéricas, los datos son medios en escala nominal u ordinal. Ejemplos: genero de las personas, color de producto, calidad de un servicio.
  - **Cuantitativas:** Determinan valores numéricos que poseen los elementos de estudio. Los datos son medidos en escala de intervalo o de razón
  - **Discretas:** pueden asumir solo ciertos valores. Provienen generalmente de conteos. Ejemplos: números de pisos de un edificio, número de títulos que posee una persona.
  - **Continuas**: pueden tomar cualquier valor. Provienen de mediciones. Ejemplos: estatura de una persona, contenido de cereal en un caja, contenido de refresco en una botella, diámetro de tuberías.
- **Gráficos Estadísticos**: son representaciones geométricas de un estudio estadístico que permite resumir la información.
- **Columnas/barras;** sectores: sirven para representar variables cualitativas y cuantitativas discretas.

- **Lineales**: se usan generalmente para estudios de comportamiento o de variables en el tiempo
- **De dispersión**: se utilizan para visualizar la relación entre dos variables.
- **Histogramas, polígonos de frecuencia, ojivas**: son gráficos representativos de variables cuantitativas continuas.

Las medidas estudiadas en esta unidad son de tres categorías: de tendencia central, de variabilidad o dispersión y de forma, las mismas que conforman las medidas estadísticas descriptivas.

- Medidas de centralización o promedios:
  - **Media aritmética**: muy utilizada para variables cuantitativas. Para su determinación se incluyen todos los valores.
  - **Mediana:** se utilizan para variables cuantitativas y datos ordinales. Asociados a esta medida están los cuartiles y percentiles. Para su determinación los datos deben estar ordenados.
  - **Moda**: medida muy representativa de variables cualitativas. Valor o característica que mas aparece
  - **Media geométrica**: medida representativa para tasas de crecimiento.
- Medidas de dispersión:
- **Amplitud de variación** = máximo-mínimo
- **Desviación media**: es el promedio de los valores absolutos con respecto a la media aritmética
- **Desviación Estándar**: raíz cuadrada de la media aritmética de las desviaciones cuadráticas con relación a la media
- **Varianza**: es el cuadrado de la desviación estándar
- **Coeficiente de variación**: es la razón porcentual entre la desviación estándar y la media aritmética.
- **Rango intercuartil**: es la diferencia entre el cuartil tres y el cuartil uno

CAPÍTULO 2

Probabilidades

Objetivos del capitulo:

- Conocer los conceptos básicos de probabilidad.
- Establecer con claridad el enfoque de probabilidad a utilizarse en situaciones reales
- Aplicar con criterio las reglas de probabilidad.
- Utilizar las técnicas de conteo para determinar la cantidad de eventos posibles o favorables.
- Conocer los conceptos básicos de probabilidad.
- Establecer con claridad el enfoque de probabilidad a utilizarse en situaciones reales
- Aplicar con criterio las reglas de probabilidad.
- Utilizar las técnicas de conteo para determinar la cantidad de eventos posibles o favorables.
- Conocer las características de la distribución normal y la normal estándar
- Determinar valores estandarizados para una variable aleatoria continua, que presenta características de normalidad.
- Calcular valores de probabilidad relacionados con intervalos cerrados o abiertos.
- Usar la distribución normal como una aproximación a la distribución binomial

Introducción de probabilidades

En la naturaleza y en la vida cotidiana existen fenómenos que no se pueden determinar su resultado con certeza, más bien se está en el posibilidad de anticipar el mismo con un cierto grado de probabilidad de ocurrencia, ejemplo determinar a qué edad una persona fallecerá, el interés que percibirá un accionista luego de tres años de realizada una inversión. Es lógico que nadie pueda dar una respuesta a estos dos eventos, pero si se establece algún valor, existe una incertidumbre de su validez. Para satisfacer la necesidad de la obtención de resultados en sucesos donde está involucrado el azar, se desarrolló la teoría de probabilidades.

Desde la antigüedad en varias culturas aparecen referencias sobre el estudio de fenómenos aleatorios. El estudio más pormenorizado de las probabilidades aparece en los siglos XVI Y XVII, debido a las preguntas que los jugadores de azar de aquella época realizaron a prestantes estudiosos matemáticos, tales como *Galileo, Pascal* y *Fermat*. *Huygens,* publica el libro *Razonamientos relativos al juego de dados (Ratiociniis in ludo alae),* que se convierte en el primer libro escrito sobre probabilidades de la historia.

Ya en el siglo XVIII, el cálculo de probabilidades se extendió a problemas físicos y de seguros marítimos, constituyéndose como factor principal de desarrollo de la teoría de probabilidades el conjunto de problemas de astronomía y de física que aparecieron relacionados con la verificación empírica de la teoría de Newton. *Pierre Simón Laplace*, introdujo la primera definición explícita de probabilidad y desarrolló la ley normal como modelo para la descripción de la variabilidad de los errores de mediación.

En el siglo XIX, tanto matemáticos como astrónomos continuaron ampliando la teoría probabilística, que permitieron su consolidación como una rama científica. Pese a ello sus aplicaciones se daban mayoritariamente en la física y astronomía. Posteriormente ya en el siglo XX, A. N. Kolmogorov en 1933 estableció una descripción axiomática de la idea de probabilidad, lo que constituyó la base de la moderna teoría, consiguiéndose de esta manera dar aplicación a muchas ciencias y campos de la vida cotidiana. En las últimas décadas, el uso de la teoría de probabilidades, en las ciencias sociales, naturales, cálculos actuariales o en la economía ha crecido enormemente, debido a estas circunstancias su estudio y conocimiento es una necesidad imprescindible e impostergable.

2.1. Conceptos Básicos

Paratenerunamejor comprensióndelostemasdeestaunidad,acontinuaciónvamosaestablecer ciertos conceptosindispensables.

Posibilidad:

Factibilidad de ocurrencia de algún suceso o evento. Probabilidad (p): Es el valor numérico comprendido entre 0 y 1, de que ocurra un evento. Mientras el valor se acerca a 1, existe más probabilidad de que el suceso ocurra, por el contrario si el valor se acerca a cero la probabilidad de ocurrencia es casi nula.

Si el valor de la probabilidad es 1, entonces estamos en presencia de una certeza (evento cierto) y si el valor es cero estamos en una imposibilidad (evento imposible), entonces.

$p_{(certeza)} = 1$

$p_{(imposibilidad)} = 0$ 0

$\leq p_{(evento)} \leq 1$

Espacio Muestral (E)

Es el conjunto de todos los resultados posibles de un experimento.

Ejemplos:

Al lanzar una moneda, el espacio muestral es igual a dos, toda vez que puede caer la moneda con cara hacia arriba o también con sello.

Al extraer una carta de una naipe común, el espacio muestral es 52, pues, es esa cantidad la que tiene el naipe.

Experimento

Proceso bien definido que permite ver uno solo de los resultados posibles del espacio muestral.

Resultado

Evidencia de lo ocurrido en el experimento.

Evento o suceso

Subconjunto del espacio muestral que contiene uno o varios resultados del espacio muestral. Por lo tanto los eventos pueden clasificarse como ***elementales***, si constan de un solo resultado y ***compuestos*** si constan de dos o más resultados.

Ejemplo

Si el experimento consiste en el lanzamiento de un dado podemos definir el evento ***(E1)*** que el dado muestre el tres en su cara superior (evento elemental). También podemos definir otro evento ***(E2)*** que en la cara superior aparezca cualquier número par (2, 4, 6), es un claro ejemplo de evento compuesto, pues consta de tres resultados posibles.

Eventos mutuamente excluyentes

Sucesos que no permiten que otro evento suceda al mismo tiempo. Esto implica que no existe intersección entre ellos.

Eventos complementarios

Se dice que dos sucesos son complementarios si no existe intersección entre ellos y la unión de estos eventos es igual al espacio muestral.

$A \cap B = \Phi$ y $A \cup B = E$, entonces $A^{C} = B$ y $B^{C} = A$

Cualidades o características que deben tener los eventos

Unión

EJERCICIO N° 1

A = { **x/x es una letra del abecedario** }

A = { a, b, c, d, e, f, g, h, i, j, **k, l,** m, **n**}

B = { **k, l, n,** o, p, q, r, s, t, u, v, w, x, y, z }

A U B = { **x/x € A V X€ B**}

A U B = { a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z }

EJERCICIO N° 2

C = { **x/x es un animal** }

C = { **perro,** gato, **conejo,** tigre, puma, oso }

D = { **perro, conejo,** puma, mono, león }

C U D = { **x/x € C V X€ D**}

C U D = { perro, gato, conejo**,** tigre, puma, oso, mono, león }

EJERCICIO N° 3

E = { **x/x es un número** }

E = { 1, 2, 3, 4, 5, 6 }

F = { **3**, **6,** 7, 8**,** 9, 10,11 }

E U F = { **x/x € E V X€ F**}

E U F = { 1, 2, 3, 4, 5, 6, 7, 8**,** 9, 10,11 }

EJERCICIO N° 4

G = { **x/x es un color** }

G = { **amarillo, verde,** celeste, violeta, naranja, rojo}

H = { azul, rosado**, verde, amarillo,** blanco }

G U H = { **x/x € G V X€ H**}

G U H = { amarillo, verde, celeste, violeta, naranja, rojo, azul, rosado**,** blanco }

EJERCICIO N° 5

I = { **x/x es un deporte** }

I = { **patinaje,** futbol, básquet, **vóley** }

J = { **futbol,** futbol, básquet, **vóley, patinaje** }

I U J = { **x/x € I V X€ J**}

I U J = { patinaje, futbol, básquet, vóley }

Interseccion

EJERCICIO N° 1

A= x/x es un color de la bandera del ecuador

A= amarrillo, azul, rojo

B= x/x es un color de la bandera de la ciudad de quito

B= azul, rojo

A ∩ B = {x/x € A € B}

A ∩ B= azul, rojo

EJERCICIO N° 2

A= x/x es un animal domestico

A= perro, gato, gallina, vaca, oveja, caballo

B= x/x es un animal para mascota

B= perro, gato patos, conejo cuy

A ∩ B = {x/x € A € B}

A ∩ B= perro, gato,

EJERCICIO N° 3

A= x/x es un material de oficina

A= esfero, papel bon, perforadora, lápiz, grapadora, cubre hojas

B= x/x es un útil escolar

B= borrador, sacapuntas, pinturas, lápiz, esfero, compás, cubre hojas

A ∩ B = {x/x € A € B}

A ∩ B= lápiz, esfero, cubre hojas

Complemento

EJERCICIO N° 1

U= x/x es un día de la semana

A= miércoles, viernes, domingo

A^ = {x/x € U; €/ A}

A^ = martes, jueves, sábado, domingo

EJERCICION°2

U= x/x es un electrodoméstico de línea blanca

A= licuadora, batidora, microondas, waflera

A^ = {x/x € U; €/ A}

A^ = lavadora, refrigeradora, secadora, cocina

EJERCICIO N° 3

U= x/x es un medio de comunicación

A= periódicos, teléfono, caratas

A^ = {x/x € U; €/ A}

A^ = televisión, radio, internet, revistas

EJERCICIO N° 4

U= x/x es un medio de transporte

A= bus, bicicleta, globo aéreo

A^ = {x/x € U; €/ A}

A^ = avión, helicóptero, tren, barco, canoa

EJERCICION°5

U= x/x es undeporte

A= ciclismo, tenis, natación

A^ = {x**/x € U; €/ A}**

A^ = futtbol, basquett, patinaje. Vóley, moto cross, rafting, paracaidismo

Diferencia

EJERCICIO N° 1

A = { **x/x es una letra del abecedario** }

A = { a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z }

B = { **x/x es una vocal** }

B = { a, e, i, o, u }

A U B = { **x/x € A V B X € B**}

A U B = { b, c, d, f, g, h, j, k, l, m, n, p, q, r, s, t, v, w, x, y, z }

EJERCICIO N° 2

C = { **x/x es un animal doméstico** }

C = { perro, gato, conejo }

D = { **x/x es un animal salvaje** }

D = { puma, mono, león, tigre, oso }

C U D = { **x/x € C V D X € D**}

C U D = { perro, gato, conejo**,** tigre, puma, oso, mono, león }

EJERCICIO N° 3

E = { **x/x es un número par** }

E = { 2, 4, 6, 8, 10, 12, 14, 16, 18, 20 }

F = { **x/x es un número impar** }

F = { 3, 5, 7, 9, 11, 13, 15, 17, 19 }

E U F = { **x/x € E V F X € F**}

E U F = { 1, 2, 3, 4, 5, 6, 7, 8**,** 9, 10,11, 12, 14, 16, 18, 20 }

EJERCICIO N° 4

G = { **x/x es un color cálido** }

G = { amarillo**,** naranja, rojo, azul, blanco }

H = { **x/x es un color frío**}

H = { rosado**,** verde, violeta, }

G U H = { **x/x € G V H X € H**}

G U H = { amarillo, verde, celeste, violeta, naranja, rojo, azul, rosado**,** blanco }

EJERCICIO N° 5

I = { **x/x es un deporte clásico** }

I = { **patinaje,** futbol, básquet, **vóley** }

J = { **x/x es un deporte extremo** }

J = { ciclismo, moto cross, rafting, paracaidismo }

I U J = { **x/x € I V J X € J**}

I U J = { patinaje, futbol, básquet, vóley, ciclismo, moto cross, rafting, paracaidismo }

Ejercicios

1.- A=[x/x es una letra del equipo rey de copas de ecuador]

A=[L,I,G, A]

AUA={L,I,G,A}

3.- B=[x/x es una letra del nombre del presidente del Ecuador]

B=[R,A,F,E,L]

U=[x/x es una letra del abecedario]

U=[A,B,C,D,E,F,G,H,I,J,K,L,M,N,Ñ,O,P,Q,R,S,T,U,V,W,X,Y,Z]

B’ = [B,C,D,E,H,I,J,K,M,N,Ñ,O,P,Q,S,T,U,V,W,X,Y,Z]

(B’)’= [R,A,F,E,L]

5.- U=[x/x es un poder del estado del ecuador]

U= [EJECUTIVO, LEGISLATIVO, JUDICIAL Y JUSTICIA INDIGENA, TRANSPARENCIA Y CONTROL SOCIAL Y ELECTORAL]

C=[x/x es un poder del estado en el que es encargado el presidente] C=

[EJECUTIVO]

C’= [LEGISLATIVO, JUDICIAL Y JUSTICIA INDIGENA, TRANSPARENCIA Y CONTROL SOCIAL Y ELECTORAL]

E∩E’ = Ø

7.- U=[x/x es una provincia del ecuador]

U=[esmeraldas, Manabí, santa Elena, guayas, santo domingo de los sachilas, los ríos, el oro, Carchi, Imbabura, pichincha, Cotopaxi, bolívar, Tungurahua, Chimborazo, cañar, Azuay, Loja, Sucumbíos, napo, francisco de Orellana, Pastaza, morona Santiago, Zamora Chinchipe, Galápagos]

D=[x/x es una provincia verde del ecuador] D=

[ESMERALDAS]

D∩U= [ESMERALDAS]

9.- G ∩ Ø =

G = X/X Es un color de la bandera del Ecuador

G= amarillo, azul y rojo

Ø = Ø

G ∩ Ø = Ø

11.- H ∩ I = I ∩ H

H = X/X Es un idioma oficial del Ecuador

H = Español, quichua, shuar

I = X/X Es un idioma oficial del Perú

I = Español, Quechua, Aymará e Inglés

H ∩ I = I ∩ H

Español = Español

13.-(J ∩ S) = (S ∩ J)

J = X/X Es una letra de la ciudad blanca del Ecuador J=
I,B,R,A

D= [x/x es una letra de la provincia verde del ecuador]

D= [E,S,M,R,A,L,D]

R,A = R,A

15.-E ∩ (S∩G) =(E∩S) ∩ G

E = X/X Es un color representativo del equipo liga de quito del Ecuador

E= B,L,A,N,C,O

G= X/X Es un color representativo del equipo liga de quito del Ecuador

G= A,M,R,I,L,O

S= X/X Es una letra de las vocales

S= A,E,I,O,U

E ∩ (S∩G) =(E∩S) ∩ G

(A,O) = (A,O)

INTERSECCIÓN

1. A=(x/x ∈ es un partido político no representado en la asamblea) (PCE) B=(x/x es un partido político) (P C M L E)

A∩B= (P C E)

2. A=(x/x ∈ es una cuenta de patrimonio) (Capital, Social) B=(x/x es una cuenta de patrimonio neto) (Capital, Pagado) A∩B= (Capital)

3. A=(x/x ∈ es un componente de la computadora) (monitor, bocinas, teclado) B=(x/x es un aparato de la computadora) (ratón, modem, bocinas)

A∩B= (bocinas)

4. A=(x/x ∈ es un dispositivo de almacenamiento) (CD, DVD, disquete)

B=(x/x ∈ es un dispositivo de información) (disquete, DVD, FLASH) A∩B=

(disquete)

5. A=(x/x ∈ es una parte del cuerpo humano) (cabeza, antebrazo, cara, cuello, hombro) B=(x/x ∈ es una parte del ser humano) (abdomen, pierna, rodilla, antebrazo)

A∩B= (antebrazo)

DIFERENCIA

A-B=(X/X € A; X̸ € B)

1.- A= (X/X Es una letra de mi primer nombre) A=

(l,o,r,e,n,a)

B=(X/X Es una letra de mi segundo nombre) B= (

m,a,r,i,t,z,a)

A-B=(l,o,e,n)

2.- A= (X/X Es una letra del abecedario)

A= (a,b,c,d,e,f,g,h,i,j,k,l,ll,m,n,ñ,o,p,q,r,s,t,u,v,w,x,y,z) B= (

X/X Es una vocal abierta)

B= (a,e,o)

A-B=(b,c,d,f,g,h,i,j,k,l,ll,m,n,ñ,p q,r,s,t,u,v,w,x,y,z)

3.- A=(X/X Es una letra de la ciudad blanca del Ecuador) A=(

i,b,a,r,r,a)

B=(X/X Es una letra de la cuidad capital del Ecuador)

B=(q,u,i,t,o)

A-B=(i,b,a,r)

4.- A=(X/X Es una letra del primer día de la semana) A=(

l,u,n,e,s)

B=(X/X Es una letra del último día de la semana)

B=(d,om,i,n,g,o)

A-B=(l,u,e,s)

5.- A=(X/X Es una letra del color del cielo)

A=(a,z,u,l)

B=(X/X Es una letra del segundo color del semáforo)

B=(a,m,a,r,i,l,l,o)

A-B=(z,u)

Complemento:

1 U=[X/X es un abecedario alfabético]

A=[a,b,c,ch,d,e,f,g,h,i,j,k,l,ll,m,n,ñ,o,p,q,r,s,t,u,v,w,x,y,z]

A=[vocales a,e,i,o,u]

A^c=[b,c,ch,d,f,g,h,j,k,l,ll,m,n,ñ,p,q,r,s,t,v,w,x,y,z]

2 A=[X/X es un números 1-20]

U=[1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20]

A=[2,4,6,8,10,12,14,16,18,20]

A^c=[1,3,5,7,9,11,13,15,17,19]

3 A= [X/X es un todos los colores]

U=[amarillo, azul, rojo, verde claro, verde oscuro, naranja, plomo, celeste, blanco, morado, negro, rosado,]

A=[amarillo, azul, rojo]

A^c= verde claro, verde oscuro, naranja, plomo, celeste, blanco, morado, negro, rosado]

4 A=[X/X es un mes que se celebra navidad]

U=[enero, febrero ,marzo ,abril, mayo ,junio, julio, agosto, septiembre, octubre, noviembre, diciembre]

A=[diciembre]

A^c=[enero, febrero ,marzo, abril, mayo, junio, julio, agosto, septiembre, octubre, noviembre]

5 A=[X/X es un día de la semana en que el presidente del ecuador da un enlace nacional]

U=[lunes, martes, miércoles, jueves, viernes, sábado, domingo] A=[sábado]

A^c=[lunes, martes, miércoles, jueves, viernes, domingo]

UNION

1.-

A=(X/XesunZpares$^+$>11^≤20) A=(12,14,16,18,20) B=(X/X esunR>1^<12) B=(2,3,4,5,6,7,8,9,10,11)

AUB=(2,3,4,5,6,7,8,9,10,11,12,14,16,18,20)

2.-

A=(X/Xesunavocalabierta) A=(a,e,o)

B=(X/Xesunavocal) B=(a,e,i,o,u)

AUB=(a,e,i,o,u)

3.-

A=(X/Xesunmultiplode3>5^≤20) A=(6,9,12,15,18)

B=(X/XesunZimpares$^-$>2^<15) B=(3,5,7,9,11,13)

AUB=(3,5,6,7,9,11,12,13,15,18)

4.-

A=(X/XesunaletradelnombredelpresidentedelarepublicadelEcuadoractual)
A=(R,AF,A,E,L)

B=(X/XesUnaletradelacapitaldelEcuador) B=(Q,u,i,t,o)

AUB=(R,A,F,Q,E,O,U,I,L,T)

5.-

A=(X/Xescolordelabanderadelecuador) A=(amarillo,azul,rojo)

B=(X/XesUncolordelabanderadeEEUU) B=(azul,rojo yblanco)

AUB=(amarillo,azul,rojo,blanco)

2.2. Técnicas de conteo

Cuando el evento es sencillo, la determinación tanto de la cantidad de resultados posibles que pueden ocurrir dentro del experimento, como la de resultados favorables a la ocurrencia del evento, es muy fácil; pues podemos conocer con toda seguridad estos valores sin recurrir a ninguna técnica especial. Si por ejemplo, el experimento consiste en lanzar un dado, sabemos que los resultados posibles son seis, pues son 6 las caras que tiene el dado, y que aparezca cada una de ellas es un evento equiprobable, si el evento es define como que salga un número impar, los resultados favorables son tres (1, 3, 5).

Por el contrario si el experimento consiste en una secuencia de etapas o la selección de varios elementos al mismo tiempo, la determinación de la cantidad de resultados favorables como posibles, debemos utilizar ciertas técnicas que nos permitan calcular estos valores. Ejemplo: si en un ánfora existen 10 esferas de diferentes colores (4 blancas, 3 rojas, 3 azules), queremos conocer de cuántas formas posibles podemos seleccionar tres esferas al mismo tiempo y calcular la probabilidad de que al realizar tal selección las esferas escogidas sean 2 de color blanco y una esfera roja; vemos que los resultados ya no pueden ser determinados sin realizar algunos cálculos previos. Para este tipo de casos es necesario contar con los conocimientos que constan en el análisis combinatorio.

En primer lugar procedamos a definir el factorial (***!***) de un número entero positivo (***n***), como el producto continuo de “n” factores desde (n) hasta (1).

$n! = n.(n-1).(n-2)\ldots 3.2.1$

6! = 6*5*4*3*2*1 = 720

Si consideramos un conjunto finito compuesto por ***n*** elementos diferentes: $\{a1, a2, a3, \ldots an\}$. Se desea conformar una colección de ***r*** elementos, donde ***r ≤ n***. El número de estos subconjuntos depende si los elementos irán ordenados o no. Las selecciones formadas por elementos ordenados se denominan ***variaciones*** y las que no toman en consideración el orden de los elementos se llaman ***combinaciones***.

Variación.

Toma el nombre de variación cada uno de los arreglos ordenados de “r” elementos tomados de otro conjunto de “n” elementos, donde r ≤ n, de manera que cada uno de estos arreglos difieren entre sí en algún elemento o en el orden de ubicación.

El número de variaciones de “r” elementos que se pueden obtener a partir de un conjunto de “n” elementos, es denotado por: V_r^n o $V_{(n,\,r)}$ y es igual a:

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

$$V_{(n,\,r)} = n(n-1)\,(n-2)\ldots\,(n-r+1)$$

Permutaciones

Es cada uno de los arreglos ordenados realizados con todos los elementos “n” de un conjunto determinado, sin considerar repetición de los mismos.

El número de permutaciones de “n” elementos es igual a:

$$P_n = n!$$

Combinaciones

Se denomina combinación a cada uno de los subconjuntos formados por “r” elementos seleccionados de un conjunto de “n” elementos, donde $r \leq n$, sin tomar en consideración el orden en el que se hallan ubicados los elementos.

El total de combinaciones de “r” elementos que se puede generar de un conjunto de “n” elementos, se denota por C_r^n o $C_{(n,\,r)}$, es igual a:

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

Ejemplo:

Determinar el número de variaciones y combinaciones de tres elementos que se pueden formar con el conjunto {a, b, c, d} y calcular el número de permutaciones de los cuatro elementos.

Solución: Se tiene que $\boldsymbol{n} = \boldsymbol{4}$ y $\boldsymbol{r} = \boldsymbol{3}$, entonces se puede formar:

$$V_{(n.r)} = \frac{n!}{(n-r)!} = \frac{4!}{(4-3)!} = 24 \text{ variaciones de 3 elementos}$$

$$C_{(n.r)} = \frac{n!}{r!(n-r)!} = \frac{4!}{3!(4-3)!} = 4 \text{ variaciones de 3 elementos}$$

$P_n = P_4 = 4! = 24$ permutaciones de 4 elementos

1.- Un cuerpo directivo de la cámara de industria del calzado para el año 2017 deberá estar integrado por 6 miembros desde ya han sido propuestos de 7 hombres y 5 mujeres para ser electos: Determinar

a) Figuran 4 hombres y 3 mujeres

b) Se requiere que haya lo mínimo 3 hombres y 2 mujeres

c) Nótese que no se ha tomado en cuenta todos los elementos y que el orden carece importancia.

a) $$C\,7.4 = \frac{7!}{(7-4)!\,4!} = \frac{7!}{3!\,4!} = \frac{7.6.5.4}{3!\,4!} = \frac{210}{6} = 35$$

$$C5.3 = \frac{5!}{(5-3)!3!} = \frac{5.4.3.}{2!\,3!} = \frac{20}{2} = 10 \qquad 35.\,10 = 350$$

b) $$C5.2 = \frac{5!}{(5-2)!2!} = \frac{5.4.3.2}{3!\,2!} = \frac{20}{2} = 10$$

$$C7.5 = \frac{7!}{(7-5)!5!} = \frac{7.6.5}{2!\ 5!} = \frac{42}{2} = \underline{21}$$

$$C5.2 = C7.5 = 10.21 = 210$$

2.- la empresa la "Salada S.A" se ha visto en la necesidad de ser liquidada el contador ha calificado en 8 grupos de deudas distintas no existe un sistema periodo para atender las deudas, con la salvedad de los trabajadores.

Grupo A debe ser indemnizado antes que nadie los accionistas del grupo 7 recibirán su dinero después de que todos los demás grupos hayan recibido su parte.

¿Cuántas secuencias diferentes pueden hacerse frente a las adversidades de la empresa?

6.5.4.3.2.1 = 720

P6 = 6! = 6.5.4.3.2.1 = 720

3.- Si un librero tiene 4 libros de Ingles 5 de Algebra 7 de Historia de cuantas maneras se puede coger 7 de ellos en cada uno de los siguientes casos

a) Debe ser de 3 de cada materia

b) Debe ser al menos de 3 de Historia y 3 de Algebra

c) Debe ser como mínimo 6 de Historia

a) $C5.3 = \frac{5!}{2!\ 3!} = \frac{5.4.3}{2!\ 3!} = \frac{20}{2!} = 10$

$$C4.3 = \frac{4!}{1!\,3!} = \frac{4!\ \cancel{3!}}{1!\ \cancel{3!}} = 4$$

$$C7.3 = \frac{7!}{4!\ 3!} = 35$$

b) $C7.5\ C5.3 = \frac{7!}{2!\ 5!} . \frac{5!}{2!\ 3!} = 21.\ 10 =$ **210**

$C7.3\ C5.3 = \frac{7!}{4!\ 3!} . \frac{5!}{3!} = 35.\ 1 =$ **35**

$C7.3\ C5.3 = \frac{7!}{4!\ 5!} . \frac{5!}{2!\ 3!} = \frac{4!}{1!\ 3!} = 35.\ 10.\ 4 =$ **1400**

$C7.4\ C5.4 = \frac{7!}{4!\ 5!} = \frac{5!}{1!\ 4!} = 35.\ 5 =$ **175**

$C7.4.C5.3.C4.2 = \frac{7!}{3!\ 4!} . \frac{5!}{2!\ 3!} . \frac{4!}{2!\ 2!} = 35.\ 10.\ 6 =$ **2100**

$C7.3.\ C5.4.\ C4.2 = \frac{7!}{4!\ 5!} . \frac{5!}{1!!\ 4!} . \frac{4!}{2!\ 2!} = 35.\ 5.\ 6 =$ **1050**

210+35+1400+175+2100+1050 = **4970**

c) $C5.5\ C7.3 = \frac{5!}{1!\ 5!} = 1\ \frac{7!}{4!\ 3!} = 35$

$$C5.5\ C4.3 = \frac{5!}{1!\ 5!} \cdot \frac{4!}{1!\ 3!} = 1.\ 4 = 4$$

$$C5.5\ \ C7.2\ \ C4.2 = \frac{5!}{1!\ 5!} \cdot \frac{7!}{5!\ 2!} \cdot \frac{4!}{2!\ 2!} = 1.\ 21.\ 6 = \frac{126}{165}$$

d) $$C7.7 = \frac{7!}{7!} = 1$$

$$C7.6\ \ C4.2 = \frac{7!}{1!\ 6!} \cdot \frac{4!}{2!\ 2!} = 7.\ 6 = 42$$

$$C7.6\ \ C5.2 = \frac{7!}{1!\ 7!} = \frac{5!}{3!\ 2!} = 7.\ 10 = 70$$

1+42+70 = 113

4.- Un estudiante debe responder 11 de 14 preguntas en las cuantas maneras debe

a) escoger las diez preguntas

b) Una de las 3 primeras es obligatoria y la otra no debe contestar

c) Si tiene que contestar 4 de las 6 primeras

a) $$C14.11 = \frac{14!}{3!\ 4!} = 364$$

b) C12. 10 = 66

$$C12.10 = \frac{66}{132}$$

C) $C6.4 = \frac{6!}{2!\,4!} = 15$

$C9.8 = \frac{9!}{1!\,8!} = \frac{9.\,8!}{1!\,8!}\ \cancel{= 9}$

d) $C6.4\ C\,9.8 = \frac{6!}{2!\,4!} . \frac{9!}{1!\,8!} = 15.\ 9 = 135$

$C6.5\ C9.7 = \frac{6!}{1!\,5!} . \frac{9!}{2!\,7!} = 6.\ 36 = 216$

$C6.6\ C\,9.6 = \frac{6!}{6!} . \frac{9!}{3!\,6!} = 1.\ 84 = 84$

435

Arreglos múltiples o multiplicación de opciones: Si un experimento consta de una sucesión de etapas, donde en cada una de ellas pueden ocurrir de diferente manera, el total de arreglos que se puede lograr es igual al producto del número de formas en que puede ocurrir cada etapa. Lo mismo sucede si se deben formar arreglos con cada elemento de diferentes conjuntos, el número total de arreglos diferentes es igual al producto del número de elementos de cada conjunto.

Total de arreglos = $n_1.n_2.n_3...n_k$

Ejemplo:

Una distribuidora de autos tiene para un mismo auto, estas diferentes alternativas, tipo de combustible que utiliza (2) diesel o gasolina, color de la carrocería (5) blanco, negro, azul, plateado y verde; opciones de la tapicería (3), tipo de trasmisión (2). ¿Cuál es el número de alternativas de que dispone un cliente para elegir?

Solución: Tenemos que $n_1 = 2$; $n_2 = 5$; $n_3 = 3$; $n_4 = 2$ Total de alternativas = 2(5) (3) (2) = 60 diferentes opciones

Los resultados obtenidos mediante alguna de las técnicas antes señaladas, pueden ser representados en forma gráfica, mediante el uso de los diagramas de árbol, vamos a ejemplificar el resultado obtenido en el ejemplo 12.1, en lo que hace referencia al número de variaciones de tres elementos.

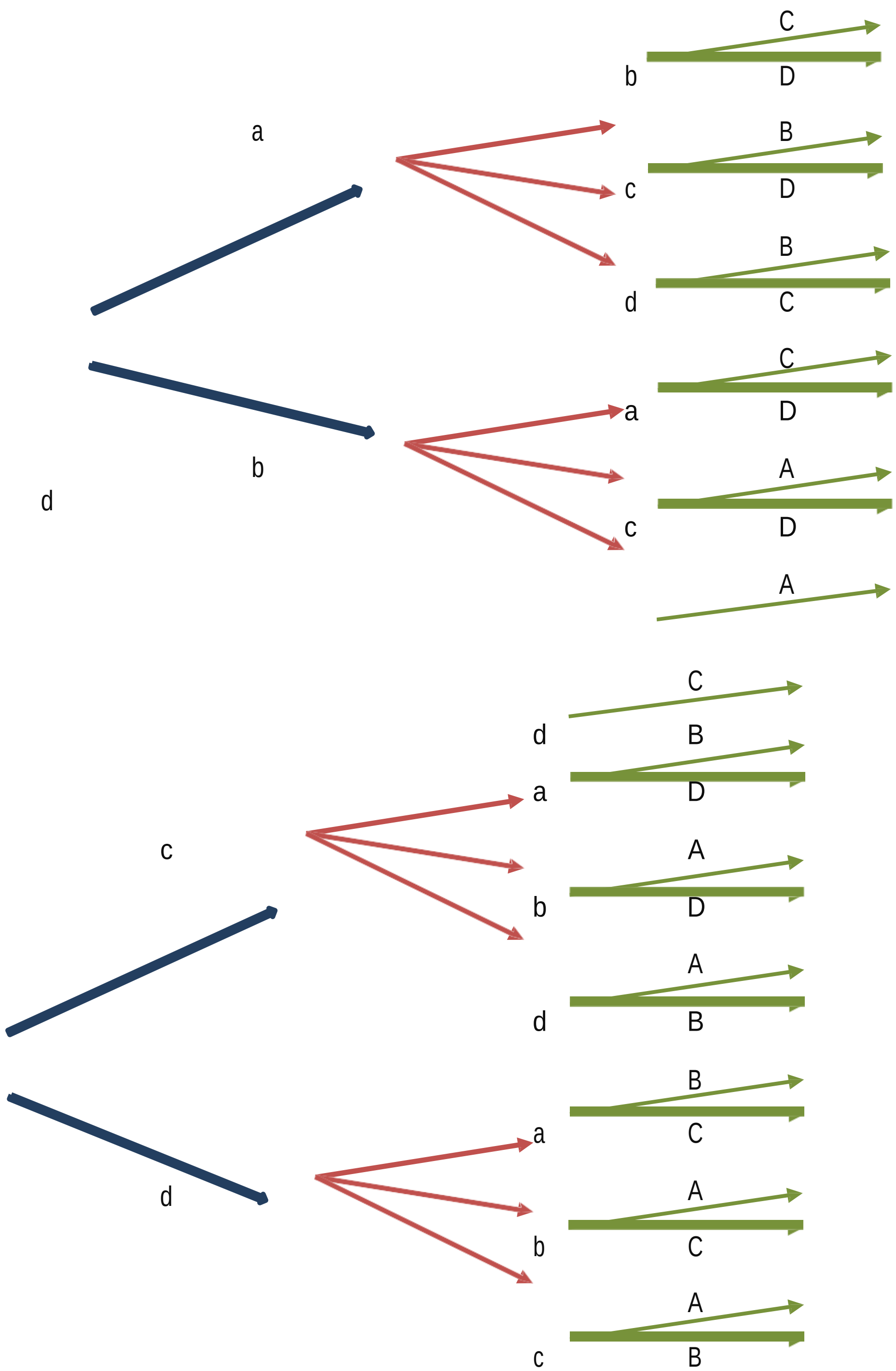


Diagrama de árbol para la variación de 3 elementos de 4 posibles.

Ejemplo

Una empresa dispone de 15 ejecutivos, se desea seleccionar una muestra de 3 ejecutivos para que asistan a una entrevista. ¿Cuántas muestra diferentes de tamaño 3 se pueden obtener del grupo total de 15 ejecutivos?

¿Cuántos números de 6 cifras se pueden formar, si: No se

permite repetición de cifras.

Las cifras pueden repetirse

Trace un diagrama de árbol, para el siguiente caso, lanzamiento de 3 monedas equilibradas.

¿De cuántas maneras diferentes se puede llenar un examen compuesto de 10 preguntas de opción múltiple, si cada pregunta ofrece tres alternativas de respuesta, suponiendo que el examen es contestado de forma aleatoriacompletamente.

Análisis Combinatorio:

Ejemplo de permutación y combinación

1.-El cuerpo directivo de la Cámara de Industria para el año 2017 deberá estar integrado por 6 miembros desde ya han sido propuesto 7 hombres y 5 mujeres para ser electos deberá.

Si se desea que el grupo Directivo figuren 4 hombres y 3 mujeres. Si se

requiere que haya como mínimo 3 hombres 2 mujeres.

Nótese que no se toma en cuenta todos los elementos y que el orden carece de importancia.

Solución:

a) C 7.4 C 5.3 = 35* 10= 350

$$C_{7.4} = \frac{7!}{(7-4)!4!} = \frac{7!}{3!*4!} = \frac{7*6*5*4!}{3*2*1\ 4!} = \frac{210}{6} = 35$$

C 5.3= 5! = 5! = 5*4* 3! =20 = 10

(5-3)!3!   2!*3!   2*1 3!   2

b) C7.5   C 5.2 = 21* 10= 210

C 7.5= 7! = 7! = 7*6* 5! = 42 = 21

(7-5)!5!   2!*5!   2*1 5!   2

C 5.2= 5! = 5! = 5*4* 3! =20 = 10

(5-2)!2!   3!*2!   2*1 3!   2

Ejemplo:

2.- La empresa La Salada S.A. se ha visto en la necesidad de ser liquidada al contado, ha clasificado sus deudas en 8 grupos distintos:

a) No existe un sistema de prioridades para atender las deudas

b) Con la seguridad de los trabajadores de grupo “A” deben ser indemnizados antes que nadie y los accionistas del grupo “G” recibirán su dinero después de que todos los demás grupos haya recibido su parte. Cuantas secuencias diferentes pueden seguirse para ser frente a todas las deudas de la empresa.

P!=6! = 6 * 5 * 4 * 3 * 2 * 1= 720

Ejemplo

3.- Si un librero tiene 4 libros de Ingles, 5 de Algebra y 7 de Historia, de cuantas maneras se pueden escoger 6 de ellos, en cada uno de los siguientes casos:

Solución:

a) C5.2   C4.3   C 7.2

10   4   35 = 1400

$$C\,5.2 = \frac{5!}{(5-2)!\,2!} = \frac{5!}{3!*2!} = \frac{5*4*3!}{2*1\,3!} = \frac{20}{2} = 10$$

$$C\,4.3 = \frac{4!}{(4-3)!\,3!} = \frac{4!}{2!*3!} = \frac{4*3!}{2*1\,3!} = \frac{4}{1} = 4$$

$$C\,7.3 = \frac{7!}{(7-3)!\,3!} = \frac{7!}{4!*3!} = \frac{7*6*5*4!}{3*2*1\,4!} = \frac{210}{6} = 35$$

b) $C\,7.5 \quad C\,5.3 = 21 * 10 = 210$

$$C\,7.5 = \frac{7!}{(7-5)!\,5!} = \frac{7!}{2!*5!} = \frac{7*6*5!}{2*1\,5!} = \frac{42}{2} = 21$$

$$C\,5.2 = \frac{5!}{(5-2)!\,2!} = \frac{5!}{3!*2!} = \frac{5*4*3!}{2*1\,3!} = \frac{20}{2} = 10$$

$C\,7.3 \quad C\,5.5 = 35 * 1 = 35$

$$C\,7.3 = \frac{7!}{(7-3)!\,3!} = \frac{7!}{4!*3!} = \frac{7*6*5*4!}{3*2*1\,4!} = \frac{210}{6} = 35$$

$$C\,5.5 = \frac{5!}{(5-5)!\,5!} = \frac{5!}{5!} = 1$$

$C\,7.3 \quad C\,5.3\ C\,4.3 = 35 * 10 * 4 = 1400$

$$C7.3 = \frac{7!}{(7-3)!\,3!} = \frac{7!}{4!*3!} = \frac{7*6*5*4!}{3*2*1\,4!} = \frac{210}{6} = 35$$

$$C\,5.3 = \frac{5!}{(5-3)!\,3!} = \frac{5!}{2!*3!} = \frac{5*4*\,3!}{2*1\,3!} = \frac{20}{2} = 10$$

$$C\,4.3 = \frac{4!}{(4-3)!\,3!} = \frac{4!}{2!*3!} = \frac{4*\,3!}{2*1\,3!} = \frac{4}{1} = 4$$

C7.4 C 5.4 = 35 * 5 = 175

$$C7.4 = \frac{7!}{(7-4)!\,4!} = \frac{7!}{3!*4!} = \frac{7*6*5*4!}{3*2*1\,4!} = \frac{210}{6} = 35$$

$$C\,5.4 = \frac{5!}{(5-4)!\,4!} = \frac{5!}{1!*4!} = \frac{5*\,4!}{1*\,4!} = \frac{5}{1} = 5$$

C7.4 C 5.3 C 4.2 = 35 * 10 *6 = 2100

$$C7.4 = \frac{7!}{(7-4)!\,4!} = \frac{7!}{3!*4!} = \frac{7*6*5*4!}{3*2*1\,4!} = \frac{210}{6} = 35$$

$$C\,5.3 = \frac{5!}{(5-3)!\,3!} = \frac{5!}{2!*3!} = \frac{5*4*\,3!}{2*1\,3!} = \frac{20}{2} = 10$$

$$C\,4.2 = \frac{4!}{(4-2)!\,2!} = \frac{4!}{2!*2!} = \frac{4*3*\,2!}{2*1\,2!} = \frac{12}{2} = 6$$

C7.3 C 5.4 C 4.2 = 35 * 5 *6 = 1050

$C7.3 = \frac{7!}{(7-3)!3!} = \frac{7!}{4!*3!} = \frac{7*6*5*4!}{3*2*1\ 4!} = \frac{210}{6} = 35$

$C\ 5.4 = \frac{5!}{(5-4)!4!} = \frac{5!}{1!*4!} = \frac{5*\ 4!}{1*\ 4!} = \frac{5}{1} = 5$

$C\ 4.2 = \frac{4!}{(4-2)!2!} = \frac{4!}{2!*2!} = \frac{4*3*\ 2!}{2*1\ 2!} = \frac{12}{2} = 6$

210 + 35 + 1400 + 175 + 2100 + 1050 = 4970

c) Deben ser 5 de algebra

C 5.5 C 7.3 = 1*35 =35

C 5.5 C 4.3 = 1*4 =35

$C5.5 \quad C\ 7.2\ C\ 4.2 = 1*21*6 = \underline{126}$

165

d) Deben ser como mínimo 6 de historia. C

7.7 = 1

C 7.6 C 4. 2 = 7 * 6 = 42

C 7.6 C 5.2 = 7 * 10= 70

Mínimo 6 de historia 1+ 42 + 70 = 113

Ejemplo

4.-Un estudiante debe resolver 11 preguntas de 14 de cuantas maneras puede escoger las 11 preguntas bajo las siguientes condiciones.

a) escoger las 11 preguntas

b) si 2 de las 3 primeras es obligatoria y la otra no debe contestarse

c) si debe contestar exactamente 4 de las 6 primeras

d) si contesta al menos 4 de las 6 primeras

SOLUCIÓN:

a) $C14.\ 11 = \frac{14!}{(14-11)!\,11!} = \frac{14!}{3!*11!} = \frac{14*13*12*11!}{3*\ 11!} = 364$

b) $C12.\ 10 = \frac{12!}{(12-10)!\,10!} = \frac{12!}{2!*10!} = \frac{12*11*10!}{2*1\ 10!} = \frac{132}{2} = 66$

$C12.\ 10 = \frac{12!}{(12-10)!\,10!} = \frac{12!}{2!*10!} = \frac{12*11*10!}{2*1\ 10!} = \frac{132}{2} = 66$

Si se escoge una u otra 66+66= 132

c) $C6.4 = \frac{6!}{(6-4)!\,4!} = \frac{6!}{2!*4!} = \frac{6*5*4!}{3*2*1\ 4!} = \frac{30}{2} = 15$ para las cuatro primeras

$C9.8 = \frac{9!}{(9-8)!\,8!} = \frac{9!}{1!*8!} = \frac{9*8!}{1*8!} = 9$ para las 8 restantes

Si contestan 4 de las 6 primeras

15 * 9 = 135

d) Se tiene que contestar por lo menos 4 de las 6 primeras

C6.4 C 9.8 = 15*9 = 135

Si contesta 5 de la 6

150

C6.5 C 9.7 = 6*36 = 216

Si contesta las 6

C6.6 C 9.6 = 84*1 = 84

Tiene que contestar por lo menos 4 de las 6 primeras

135 + 216 + 84 = 435

PERMUTACIONES:

A = {l, u, i, s} 4! = 24

TABULACION:

A = {(l, u, i, s) ;(l, u, s, i) ;(l, i, u, s) ;(l, i, s, u) ;(l, s, u, i) ;(l, s, i, u) ;(u, l, i, s) ;(u, l, s, i) ;(u, i, l, s) ;(u, i, s, l) ;(u, s, l, i) ;(u, s, i, l);

(i, l, u,s);(i, l, s,u);(i, u, l, s);(i, u, s, l);(i, s, l, u);(i, s, u, l));(s, l, u, i) ;(s, l, i, u);(s, u, l, i) ;(s, u, i, l); (s, i, u, l); (s, i, l, u) }

Conjunto A total 24 elementos

1.-El cuerpo directivo de la Cámara de Industria para el año 2017 deberá estar integrado por 6 miembros desde ya han sido propuesto 7 hombres y 5 mujeres para ser electos deberá.

Si se desea que el grupo Directivo figuren 4 hombres y 3 mujeres Si se

requiere que haya como mínimo 3 hombres 2 mujeres

Nótese que no se toma en cuenta todos los elementos y que el orden carece de importancia.

SOLUCIÓN:

a) Si se desea que el grupo Directivo figuren 4 hombres y 3 mujeres

C5 3 = $\underline{\quad 5! \quad}$ = $\underline{5.4.3}$ = **10**

(5-3)!3!   2! 3!

$$C_4^7 = \frac{7!}{(7-4)!4!} = \frac{7.6.5.4}{3!\,4!} = \mathbf{70}$$

C5 3 . C7 4 = 10 . 70= **700**

Si se requiere que haya como mínimo 3 hombres 2 mujeres

$$C_2^5 = \frac{5!}{(5-2)!2!} = \frac{5.4.3}{3!\,2!} = \mathbf{10}$$

$$C_5^7 = \frac{7!}{(7-5)!5!} = \frac{7.6.5}{2!\,5!} = \mathbf{21}$$

C5 2 . C7 5 = 10 . 21= **210**

2.-La empresa La Salada S.A. se ha visto en la necesidad de ser liquidada al contado, ha clasificado sus deudas en 8 grupos distintos:

a) No existe un sistema de prioridades para atender las deudas

b) Con la seguridad de los trabajadores de grupo “A” deben ser indemnizados antes que nadie y los accionistas del grupo “G” recibirán su dinero después de que todos los demás grupos haya recibido su parte. Cuantas secuencias diferentes pueden seguirse para ser frente a todas las deudas de la empresa.

SOLUCIÓN:

Pn=n!

6!= 6.5.4.3.2.1= 720

R= 720

3.- Si un librero tiene 4 libros de inglés y 5 de algebra y 6 de historia

De cuantas maneras se puede escoger 7 de ellos en uno de los siguientes casos. Deben ser 3 de cada materia

Deben ser al menos 3 de historia y 3 de algebra

Deben ser 5 de algebra

Deben ser como mínimo 6 de historia

SOLUCIÓN:

Deben ser 3 de cada materia

C4.3	C	5.3	C7.3			
4	.	2.5	**.**	35	**=**	**350**

Deben ser al menos 3 de historia y 3 de algebra

C7.5	.	C5.3	.		=	262.5
C7.3	.	C5.5	.		=	35
C7.3	.	C5.3	.	C4.3	=	350
C7.4	.	C5.4	.		=	175
C7.4	.	C5.3	.		=	87.5
C7.3	.	C5.4	.	C4.1	=	700
+	**35+**	**350+**	**175+**	**87.5+**	**700 =**	**1610**

Deben ser 5 de algebra

CA.A **.** C7.3 = 1.35 = **35**

CA.A	.	C4.3	=1.4	= **4**
CA.A	.	C7.2	.C4.2	= 1.126 = **126**
				R= 165

Deben ser como mínimo 6 de historia

C7.7					**= 1**
C7.7	.	C4.2	=	21.6	**=126**
C7.6	.	C5.2	=	7.10	**=70**
				R=	**197**

4.-Un estudiante debe resolver 11 preguntas de 14 de cuantas maneras puede escoger las 11 preguntas bajo las siguientes condiciones.

SOLUCIÓN:

a) escoger las 11 preguntas

b) si 2 de las 3 primeras es obligatoria y la otra no debe contestarse

c) si debe contestar exactamente 4 de las 6 primeras

d) si contesta al menos 4 de las 6 primeras

a) escoger las 11 preguntas

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

$$C14,11= \frac{14!}{(14-11)!\ 11!} \qquad C14,11= \frac{14*13*12*\cancel{11!}}{3!\ \cancel{11!}} = C14,11= \frac{2184}{6} = \mathbf{364}$$

b) si 2 de las 3 primeras es obligatoria y la otra no debe contestarse

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

$$C11,9= \frac{11!}{(11-9)!\ 9!} \qquad C11,9= \frac{11*10*\cancel{9!}}{2!\ \cancel{9!}} = \quad C11,9= \frac{110}{2} = \underline{\underline{\mathbf{55}}}$$

Si escoge 2 de las 3 primeras 55 + 55 **<u>= 110</u>**

c) si debe contestar exactamente 4 de las 6 primeras

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

$$C6,4= \frac{6!}{(6-4)!\ 4!} \qquad C6,4= \frac{6*5*\cancel{4!}}{2!\ \cancel{4!}} = \quad C6,4= \frac{30}{2} = \underline{\underline{\mathbf{15}}}$$

Para las 4 primeras

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

$$C8,7= \frac{8!}{(8-7)!\ 7!} \qquad C8,7= \frac{8*\cancel{7!}}{1!\ \cancel{7!}} = \quad C8,7= \frac{8}{1} = \underline{\underline{\mathbf{8}}}$$

Para 7 restantes

Si contesta 4 de las 6 primeras 15*8 = **120**

d) si contesta al menos 4 de las 6 primeras

Si contesta 4 de las 6 primeras

C6,4 * C 8,7 = **<u>120</u>**

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

$$C6,5=\frac{6!}{(6-5)!\,5!}\qquad C6,5=\frac{6*\cancel{5!}}{1!\,\cancel{5!}}=\qquad C6,5=\frac{6}{1}=\underline{\underline{\mathbf{6}}}$$

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

$$C8,6=\frac{8!}{(8-6)!\,6!}\qquad C8,6=\frac{8*7*\cancel{6!}}{2!\,\cancel{6!}}=\qquad C8,6=\frac{56}{2}=\underline{\underline{\mathbf{28}}}$$

Si contesta 5 de las 6 primeras 6*28 =**168**

$$C6,6=\frac{\cancel{6!}}{(6-6)!\,\cancel{6!}}$$

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

$$C8,5=\frac{8!}{(8-5)!\,5!}\qquad C8,5=\frac{8*7*6*\cancel{5!}}{3!\,\cancel{5!}}=\qquad C8,5=\frac{336}{6}=\underline{\underline{\mathbf{56}}}$$

Si contesta las 6 120+168+56 **=344**

Ejercicio N° 1

A= { x/x es un Partido Político del Ecuador **}**

A= { ALIANZA PAIS; CND; ID; MCMG; MFE; MIJS; MIL; MITS; MMIN; MNCS; MPC; MPD; MUNO; PRE; PK; PLE; PRIAN; PS; PSC; PSP; R25; RED **}**

Ejercicio N° 2

B= {x/x es un Poder del Estado del Ecuador **}**

B= {Ejecutivo, Legislativo, Electoral, Judicial y Justicia Indígena, Transparencia Judicial **}**

Ejercicio N° 3

C= {x/x es una Norma de la Constitución de la República del Ecuador **}**

C= {Dogmática; De Carácter Orgánica **}**

Ejercicio N° 4

D= { x/x es un Presidente de la República del Ecuador periodo desde el 2008 hasta el 2012 **}**

D= {Ec. Rafael Correa Delgado**}**

Ejercicio N° 5

E= { x/x es un Periodo de Elección Presidencial en el Ecuador **}**

E= { 4 años **}**

Una carta es extraída de un mazo de 13 cartas y consiste en 2 de trébol, 2 de corazones, 2 de espadas, 2 de diamantes, determine el espacio muestral.

EM= {2T, 2C, 2E, 2D}

Una moneda es lanzada 6 veces consecutivas y las caras que muestra son observadas.

CSCSSS; SCSCCC; CSSCCC; SCCSSS; CCSCSC; SSCSCS

CCCCCC; CCCCCS; CCCCSS; CCCSSS; CCSSSS; CSSSSS

SSSSSS; SSSSSC; SSSSCC; SSSCCC; SSCCCC; SCCCCC

EM= CSCSCS; SCSCSC; CCSSCC; SSCCSS; CSSSCC; SCCCSS

CSSSCS; SCCCSC; CSSCCS; SCCSSC; CSCSSC; SCSCCS

SCCSSC; CSSCCS; SSCSCC; CCSCSS; CCCSCS; SSSCSC

Un dado es lanzado 5 veces consecutivas, la suma de las caras son anotadas.

EM={5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30}

Seleccione la sucesión de las 3 letras diferentes de la palabra AMIGO.

AMI, AMG, AMO, AIM, AIG, AIO,

AGM, AGI, AGO, AOM, AOI, AOG

EM= MIG, MIO, MIA, MGA, MGI, MGO

MOI, MOG, MOA, MAI, MAG,MAO

IGO, IGA, IGM, IOA, IOM, IOG, IAM

IAM, IAG, IGO, IMA, IMG, IMO,

GOA, GOM, GOI, GAM, GAI, GAO

GMA, GMI, GMO, GIA, GIO, GIM

OAM, OAI, OAG, OMA, OMI, OMG

OIA, OIM, OIG, OGA, OGM, OGI

Ejercicios propuestos
PERMUTACIONES:

A = {l, u, i, s} 4! = 24

TABULACION:

A = {(l, u, i, s) ;(l, u, s, i) ;(l, i, u, s) ;(l, i, s, u) ;(l, s, u, i) ;(l, s, i, u) ;(u, l, i, s) ;(u, l, s, i) ;(u, i, l, s) ;(u, i, s, l) ;(u, s, l, i) ;(u, s, i, l);

(i, l, u,s);(i, l, s,u);(i, u, l, s);(i, u, s, l);(i, s, l, u);(i, s, u, l));(s, l, u, i) ;(s, l, i, u);(s, u, l, i) ;(s, u, i, l);(s, i, u, l); (s, i, l, u) }

Conjunto A total 24 elementos

1.- El cuerpo directivo de la Cámara de Industria para el año 2017 deberá estar integrado por 6 miembros desde ya han sido propuesto 7 hombres y 5 mujeres para ser electos deberá.

Si se desea que el grupo Directivo figuren 4 hombres y 3 mujeres Si se

requiere que haya como mínimo 3 hombres 2 mujeres

Nótese que no se toma en cuenta todos los elementos y que el orden carece de importancia.

SOLUCIÓN:

a) Si se desea que el grupo Directivo figuren 4 hombres y 3 mujeres

$C5\,3 = \frac{5!}{(5-3)!\,3!} = \frac{5.4.3}{2!\,3!} = \mathbf{10}$

$C7\,4 = \frac{7!}{(7-4)!\,4!} = \frac{7.6.5.4}{3!\,4!} = \mathbf{70}$

C5 3 . C7 4 = 10 . 70= **700**

Si se requiere que haya como mínimo 3 hombres 2 mujeres

$$C5\ 2 = \frac{5!}{(5-2)!2!} = \frac{5.4.3}{3!\ 2!} = \mathbf{10}$$

$$C7\ 5 = \frac{7!}{(7-5)!5!} = \frac{7.6.5}{2!\ 5!} = \mathbf{21}$$

C5 2 . C7 5 = 10 . 21= **210**

2.- La empresa La Salada S.A. se ha visto en la necesidad de ser liquidada al contado, ha clasificado sus deudas en 8 grupos distintos:

a) No existe un sistema de prioridades para atender las deudas

b) Con la seguridad de los trabajadores de grupo "A" deben ser indemnizados antes que nadie y los accionistas del grupo "G" recibirán su dinero después de que todos los demás grupos haya recibido su parte. Cuantas secuencias diferentes pueden seguirse para ser frente a todas las deudas de la empresa.

SOLUCIÓN:

Pn=n!

6!= 6.5.4.3.2.1= 720

R= 720

3.- Si un librero tiene 4 libros de inglés y 5 de algebra y 6 de historia

De cuantas maneras se puede escoger 7 de ellos en uno de los siguientes casos. Deben

ser 3 de cada materia

Deben ser al menos 3 de historia y 3 de algebra

Deben ser 5 de algebra

Deben ser como mínimo 6 de historia

SOLUCIÓN:

Deben ser 3 de cada materia

C4.3	C	5.3	C7.3			
4	.	2.5	.	35	**=**	**350**

Deben ser al menos 3 de historia y 3 de algebra

C7.5	.	C5.3	.		=	262.5
C7.3	.	C5.5	.		=	35
C7.3	.	C5.3	.	C4.3	=	350
C7.4	.	C5.4	.		=	175
C7.4	.	C5.3	.		=	87.5
C7.3	.	C5.4	.	C4.1	=	<u>700</u>
+	**35+**	**350+**	**175+**	**87.5+**	**700 =**	**1610**

Deben ser 5 de algebra

CA.A	.	C7.3	=1.35	=	**35**
CA.A	.	C4.3	=1.4	=	**4**
CA.A	.	C7.2	.C4.2	= 1.126 =	<u>**126**</u>
					R= 165

Deben ser como mínimo 6 de historia

C7.7					**= 1**
C7.7	.	C4.2	=	21.6	**=126**
C7.6	.	C5.2	=	7.10	<u>**=70**</u>
				R=	**197**

4.-Un estudiante debe resolver 11 preguntas de 14 de cuantas maneras puede escoger las 11 preguntas bajo las siguientes condiciones.

SOLUCIÓN:

a) escoger las 11 preguntas

b) si 2 de las 3 primeras es obligatoria y la otra no debe contestarse

c) si debe contestar exactamente 4 de las 6 primeras

d) si contesta al menos 4 de las 6 primeras

a) escoger las 11 preguntas

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

$$C14,11= \frac{14!}{(14-11)!\ 11!} \quad C14,11= \frac{14*13*12*\cancel{11!}}{3!\ \cancel{11!}} = C14,11= \frac{2184}{6} = \mathbf{364}$$

b) si 2 de las 3 primeras es obligatoria y la otra no debe contestarse

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

$$C11,9= \frac{11!}{(11-9)!\ 9!} \quad C11,9= \frac{11*10*\cancel{9!}}{2!\ \cancel{9!}} = C11,9= \frac{110}{2} = \mathbf{55}$$

Si escoge 2 de las 3 primeras 55 + 55 **= 110**

c) si debe contestar exactamente 4 de las 6 primeras

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

$$C6,4= \frac{6!}{(6-4)!\ 4!} \quad C6,4= \frac{6*5*\cancel{4!}}{2!\ \cancel{4!}} = C6,4= \frac{30}{2} = \mathbf{15}$$

Para las 4 primeras

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

$$C8,7 = \frac{8!}{(8-7)!\,7!} \qquad C8,7 = \frac{8 * \cancel{7!}}{1!\,\cancel{7!}} = C8,7 = \frac{8}{1} = \mathbf{8}$$

Para 7 restantes

Sicontesta 4delas6primeras 15*8 = **120**

d) si contesta al menos 4 de las 6 primeras

Sicontesta4delas6primeras

C6,4 * C 8,7 = **120**

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

$$C6,5 = \frac{6!}{(6-5)!\,5!} \qquad C6,5 = \frac{6 * \cancel{5!}}{1!\,\cancel{5!}} = C6,5 = \frac{6}{1} = \mathbf{6}$$

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

$$C8,6 = \frac{8!}{(8-6)!\,6!} \qquad C8,6 = \frac{8 * 7 * \cancel{6!}}{2!\,\cancel{6!}} = C8,6 = \frac{56}{2} = \mathbf{28}$$

Si contesta 5 de las 6 primeras 6*28 =**168**

$$C6,6 = \frac{\cancel{6!}}{(6-6)!\,\cancel{6!}}$$

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

$$C_{8,5}= \frac{8!}{(8-5)!\,5!} \qquad C_{8,5}= \frac{8*7*6*\cancel{5!}}{3!\,\cancel{5!}} = C_{8,5}= \frac{336}{6} = \mathbf{56}$$

Si contesta las 6 120+168+56 **=344**

POLITICA

Ejercicio N° 1

A= { x/x es un Partido Político del Ecuador **}**

A= { ALIANZA PAIS; CND; ID; MCMG; MFE; MIJS; MIL; MITS; MMIN; MNCS; MPC; MPD; MUNO; PRE; PK; PLE; PRIAN; PS; PSC; PSP; R25; RED **}**

Ejercicio N° 2

B= { x/x es un Poder del Estado del Ecuador **}**

B= { Ejecutivo, Legislativo, Electoral, Judicial y Justicia Indígena, Transparencia Judicial **}**

Ejercicio N° 3

C= { x/x es una Norma de la Constitución de la República del Ecuador **}**

C= { Dogmática; De Carácter Orgánica **}**

Ejercicio N° 4

D= { x/x es un Presidente de la República del Ecuador periodo desde el 2008 hasta el 2012 **}**

D= { Ec. Rafael Correa Delgado **}**

Ejercicio N° 5

E= { x/x es un Periodo de Elección Presidencial en el Ecuador **} E= {**

4 años

Una carta es extraída de un mazo de 13 cartas y consiste en 2 de trébol, 2 de corazones, 2 de espadas, 2 de diamantes, determine el espacio muestral.

EM= {2T, 2C, 2E, 2D}

Una moneda es lanzada 6 veces consecutivas y las caras que muestra son observadas.

CSCSSS; SCSCCC; CSSCCC; SCCSSS; CCSCSC; SSCSCS

CCCCCC; CCCCCS; CCCCSS; CCCSSS; CCSSSS; CSSSSS

SSSSSS; SSSSSC; SSSSCC; SSSCCC; SSCCCC; SCCCCC

EM= CSCSCS; SCSCSC; CCSSCC; SSCCSS; CSSSCC; SCCCSS

CSSSCS; SCCCSC; CSSCCS; SCCSSC; CSCSSC; SCSCCS

SCCSSC; CSSCCS; SSCSCC; CCSCSS; CCCSCS; SSSCSC

Un dado es lanzado 5 veces consecutivas, la suma de las caras son anotadas.

EM={5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30}

Seleccione la sucesión de las 3 letras diferentes de la palabra AMIGO.

AMI, AMG, AMO, AIM, AIG, AIO,

AGM, AGI, AGO, AOM, AOI, AOG

EM= **MIG, MIO, MIA, MGA, MGI, MGO**

MOI, MOG, MOA, MAI, MAG,MAO

IGO, IGA, IGM, IOA, IOM, IOG, IAM

IAM, IAG, IGO, IMA, IMG, IMO,

GOA, GOM, GOI, GAM, GAI, GAO

GMA, GMI, GMO, GIA, GIO,GIM

OAM, OAI, OAG, OMA, OMI, OMG

OIA, OIM, OIG, OGA, OGM, OGI

Ejercicios Resueltos
Análisis Combinatorio

EJEMPLO DE PERMUTACION Y COMBINACION

Ejemplos

1.-El cuerpo directivo de la Cámara de Industria para el año 2017 deberá estar integrado por 6 miembros desde ya han sido propuesto 7 hombres y 5 mujeres para ser electos deberá.

Si se desea que el grupo Directivo figuren 4 hombres y 3 mujeres. Si se

requiere que haya como mínimo 3 hombres 2 mujeres.

Nótese que no se toma en cuenta todos los elementos y que el orden carece de importancia.

Solución:

a) C 7.4 C 5.3 = 35* 10= 350

$$C7.4 = \frac{7!}{(7-4)!4!} = \frac{7!}{3!*4!} = \frac{7*6*5*4!}{3*2*1\ 4!} = \frac{210}{6} = 35$$

$$C\,5.3 = \frac{5!}{(5-3)!3!} = \frac{5!}{2!*3!} = \frac{5*4*3!}{2*1\;3!} = \frac{20}{2} = 10$$

b) C 7.5   C 5.2 = 21* 10= 210

$$C\,7.5 = \frac{7!}{(7-5)!5!} = \frac{7!}{2!*5!} = \frac{7*6*5!}{2*1\;5!} = \frac{42}{2} = 21$$

$$C\,5.2 = \frac{5!}{(5-2)!2!} = \frac{5!}{3!*2!} = \frac{5*4*3!}{2*1\;3!} = \frac{20}{2} = 10$$

Ejemplo:

2.- La empresa La Salada S.A. se ha visto en la necesidad de ser liquidada al contado, ha clasificado sus deudas en 8 grupos distintos:

a) No existe un sistema de prioridades para atender las deudas

b) Con la seguridad de los trabajadores de grupo “A” deben ser indemnizados antes que nadie y los accionistas del grupo “G” recibirán su dinero después de que todos los demás grupos haya recibido su parte. Cuantas secuencias diferentes pueden seguirse para ser frente a todas las deudas de la empresa.

P!=6! = 6 * 5 * 4 * 3 * 2 * 1= 720

Ejemplo:

3.- Si un librero tiene 4 libros de Ingles, 5 de Algebra y 7 de Historia, de cuantas maneras se pueden escoger 6 de ellos, en cada uno de los siguientes casos:

Solución:

a) C5.2   C4.3   C 7.2

   10   4   35 = 1400

$$C\,5.2=\frac{5!}{(5-2)!2!}=\frac{5!}{3!*2!}=\frac{5*4*3!}{2*1\,3!}=\frac{20}{2}=10$$

$$C\,4.3=\frac{4!}{(4-3)!3!}=\frac{4!}{2!*3!}=\frac{4*3!}{2*1\,3!}=\frac{4}{1}=4$$

$$C\,7.3=\frac{7!}{(7-3)!3!}=\frac{7!}{4!*3!}=\frac{7*6*5*4!}{3*2*1\,4!}=\frac{210}{6}=35$$

b) C7.5   C5.3 = 21 * 10 = 210

$$C\,7.5=\frac{7!}{(7-5)!5!}=\frac{7!}{2!*5!}=\frac{7*6*5!}{2*1\,5!}=\frac{42}{2}=21$$

$$C\,5.2=\frac{5!}{(5-2)!2!}=\frac{5!}{3!*2!}=\frac{5*4*3!}{2*1\,3!}=\frac{20}{2}=10$$

C7.3   C5.5 = 35 * 1 = 35

$$C7.3=\frac{7!}{(7-3)!3!}=\frac{7!}{4!*3!}=\frac{7*6*5*4!}{3*2*1\,4!}=\frac{210}{6}=35$$

$$C5.5=\frac{5!}{(5-5)!5!}=\frac{5!}{5!}=1$$

C7.3 C 5.3 C 4.3 = 35 * 10 * 4 = 1400

$$C7.3 = \frac{7!}{(7-3)!3!} = \frac{7!}{4!*3!} = \frac{7*6*5*4!}{3*2*1\ 4!} = \frac{210}{6} = 35$$

$$C\ 5.3 = \frac{5!}{(5-3)!3!} = \frac{5!}{2!*3!} = \frac{5*4*\ 3!}{2*1\ 3!} = \frac{20}{2} = 10$$

$$C\ 4.3 = \frac{4!}{(4-3)!3!} = \frac{4!}{2!*3!} = \frac{4*\ 3!}{2*1\ 3!} = \frac{4}{1} = 4$$

C7.4 C 5.4 = 35 * 5 = 175

$$C7.4 = \frac{7!}{(7-4)!4!} = \frac{7!}{3!*4!} = \frac{7*6*5*4!}{3*2*1\ 4!} = \frac{210}{6} = 35$$

$$C\ 5.4 = \frac{5!}{(5-4)!4!} = \frac{5!}{1!*4!} = \frac{5*\ 4!}{1*\ 4!} = \frac{5}{1} = 5$$

C7.4 C 5.3 C 4.2 = 35 * 10 *6 = 2100

$$C7.4 = \frac{7!}{(7-4)!4!} = \frac{7!}{3!*4!} = \frac{7*6*5*4!}{3*2*1\ 4!} = \frac{210}{6} = 35$$

$$C\ 5.3 = \frac{5!}{(5-3)!3!} = \frac{5!}{2!*3!} = \frac{5*4*\ 3!}{2*1\ 3!} = \frac{20}{2} = 10$$

$C\,4.2 = \frac{4!}{(4-2)!\,2!} = \frac{4!}{2!*2!} = \frac{4*3*2!}{2*1\,2!} = \frac{12}{2} = 6$

C 7.3 C 5.4 C 4.2 = 35 * 5 *6 = 1050

$C\,7.3 = \frac{7!}{(7-3)!\,3!} = \frac{7!}{4!*3!} = \frac{7*6*5*4!}{3*2*1\,4!} = \frac{210}{6} = 35$

$C\,5.4 = \frac{5!}{(5-4)!\,4!} = \frac{5!}{1!*4!} = \frac{5*4!}{1*4!} = \frac{5}{1} = 5$

$C\,4.2 = \frac{4!}{(4-2)!\,2!} = \frac{4!}{2!*2!} = \frac{4*3*2!}{2*1\,2!} = \frac{12}{2} = 6$

210 + 35 + 1400 + 175 + 2100 + 1050 = 4970

c) Deben ser 5 de algebra

C 5.5 C 7.3 = 1*35 =35

C 5.5 C 4.3 = 1*4 =35

C 5.5 C 7.2 C 4.2 =1*21*6 = <u>126</u>

165

d) Deben ser como mínimo 6 de historia. C

7.7 = 1

C 7.6 C 4. 2 = 7 * 6 = 42

C 7.6 C 5.2 = 7 * 10= 70

Mínimo 6 de historia 1+ 42 + 70 = 113

Ejemplo:

4.- Un estudiante debe resolver 11 preguntas de 14 de cuantas maneras puede escoger las 11 preguntas bajo las siguientes condiciones.

a) escoger las 11 preguntas

b) si 2 de las 3 primeras es obligatoria y la otra no debe contestarse

c) si debe contestar exactamente 4 de las 6 primeras

d) si contesta al menos 4 de las 6 primeras

SOLUCIÓN:

a) $$C14.11 = \frac{14!}{(14-11)!\,11!} = \frac{14!}{3!*11!} = \frac{14*13*12*11!}{3*\ 11!} = 364$$

b) $$C12.10 = \frac{12!}{(12-10)!\,10!} = \frac{12!}{2!*10!} = \frac{12*11*10!}{2*1\ 10!} = \frac{132}{2} = 66$$

$$C12.10 = \frac{12!}{(12-10)!\,10!} = \frac{12!}{2!*10!} = \frac{12*11*10!}{2*1\ 10!} = \frac{132}{2} = 66$$

Si se escoge una u otra 66+66= 132

c) $$C6.4 = \frac{6!}{(6-4)!\,4!} = \frac{6!}{2!*4!} = \frac{6*5*4!}{3*2*1\ 4!} = \frac{30}{2} = 15$$ para las cuatro primeras

$$C9.8 = \frac{9!}{(9-8)!\,8!} = \frac{9!}{1!*8!} = \frac{9*8!}{1*8!} = 9$$ para las 8 restantes

Si contestan 4 de las 6 primeras

15 * 9 = 135

d) Se tiene que contestar por lo menos 4 de las 6 primeras

C6.4 C 9.8 = 15*9 = 135

Si contesta 5 de la 6

C6.5 C 9.7 = 6*36 = 216

Si contesta las 6

C6.6 C 9.6 = 84*1 = 84

Tiene que contestar por lo menos 4 de las 6 primeras

135 + 216 + 84 = 435

2.3. Probabilidades

Para la determinación de la probabilidad de ocurrencia de un evento se recurre a criterios o enfoques de cálculo, los cuales son:

a. Enfoque clásico o probabilidad matemática (a priori)
b. Enfoque de frecuencia relativa o probabilidad estadística (a posteriori)
c. Enfoque subjetivo

Los dos primeros tienen un sustento matemático para el cálculo de la probabilidad, mientras que el tercero, ya su nombre lo indica está basado en el conocimiento que pueda tener una persona sobre la ocurrencia de un evento.

Enfoque clásico: Para el cálculo de la probabilidad de que acontezca un evento de acuerdo a este criterio, no es necesario realizar el experimento, pues podemos determinar de acuerdo a leyes matemáticas o por conocimiento actual de la situación, la cantidad de resultados posibles que pueden ocurrir en el desarrollo del experimento, así como también el número de resultados que son favorables a que ocurra un evento.

La expresión matemática que permite calcular es el cociente entre resultados favorables con relación a los resultados posibles.

$$P(E) = \frac{\text{Número de resultados favorables al evento}}{\text{Número de resultados posibles del experimento}}$$

P(E): Se lee la probabilidad de que ocurra un evento E

Ejemplos.

1. Calcular la probabilidad de que al extraer una carta de un naipe común, ésta sea una carta de diamante.

Solución: Sabemos que un naipe común tiene 52 cartas, cualquiera de ellas puede ser extraída, por lo tanto los resultados posibles son 52. Conocemos además que la opción de diamantes tiene 13 cartas con esa denominación, por lo tanto los resultados favorables a que salga una carta de diamante son 13. Por lo tanto la probabilidad de extraer una carta de diamantes de un naipe común es:

$P(\blacklozenge) = = \frac{13}{52} = 0.25$

2. Un ánfora contiene 8 esferas, de las cuales 5 son rojas y 3 son blancas, calcular la probabilidad de que al extraer dos esferas al mismo tiempo, éstas sean de color rojo.

Solución. Debemos conocer de cuántas formas se pueden seleccionar 2 esferas de un total de 8 (resultados posibles), también es necesario conocer ce cuántas formas posibles se pueden seleccionar 2 esferas rojas de 5 que tienen ese color, para esto recurrimos a las combinaciones.

Resultados posibles: $C_{(n,\, r)} = C_{(8,2)} = \frac{8!}{2!(8-2)!} = 28$

Resultados favorables: $C_{(n,\, r)} = C_{(5,2)} = \frac{5!}{2!(5-2)!} = 10$

$P(2 esferas rojas) = \frac{10}{28} \sim 0.357$

Enfoque de frecuencia relativa: Este criterio de probabilidad es necesario utilizarlo cuando no hay la posibilidad de que matemáticamente no podemos determinar los resultados posibles que pueda tener un determinado experimento o si este experimento fue realizado con anterioridad y conocemos la información que produjo.

Ejemplo.

Si queremos determinar la probabilidad de que ocurra un accidente de tránsito en la intersección de las avenidas 10 de Agosto y Mariana de Jesús en las próximas dos horas, posiblemente no contemos con la información que nos permita calcular la probabilidad deseada, será necesario entonces realizar el experimento para establecer los valores que nos permitan obtener el resultado deseado.

Otro ejemplo podría ser si queremos conocer cuál es la probabilidad de que un jugador de fútbol, especialista en cobrar tiros libres cerca del área rival, en el próximo tiro libre que vaya a ejecutar, convierta en gol esa ejecución; lo más probable es que para determinar la probabilidad debamos recurrir a información estadística que nos puedan proporcionar periodistas deportivos que llevan en cuenta estos datos. En ambos ejemplos nos percatamos que es necesario que el experimento sea realizado o bien en el presente o fue realizado en el pasado inmediato anterior. Por lo tanto podemos establecer que para calcular la probabilidad de ocurrencia de un evento bajo este criterio, debamos recurrir a la siguiente expresión que establece la razón entre los resultados favorables que suceden o sucedieron en el experimento y el número total de observaciones que tuvo el experimento.

$$P(E) = \frac{\textit{Número de veces que ocurrió el evento}}{\textit{Número total de observaciones en el experimento}}$$

Ejemplo:

Una empresa desea conocer cuál es la probabilidad de que el siguiente lote de insumos que va a recibir de un determinado proveedor sea rechazado. En los registros históricos de la empresa se halla determinado que de los 200 lotes anteriormente entregados por el proveedor, 10 de ellos fueron rechazados por no cumplir los estándares de calidad determinados por la empresa. Entonces la probabilidad deseada es.

$$P(\text{rechazo de lote}) = \frac{10}{200} = 0.05$$

Enfoque subjetivo: Este criterio está basado en asignar un valor de probabilidad en función de la impresión, conocimiento o intuición de la persona que desea determinar la probabilidad de que acontezca un evento.

Esta condición hace que el resultado sea una mera suposición o en otros casos, por la acumulación de conocimiento sobre cómo sucede una situación, el valor establecido por una persona pueda tener validez.

Ejemplo cuando médicos oncólogos experimentados entregan una esperanza de vida a un paciente de cáncer.

Ejemplos:

¿Cuál criterio de probabilidad es apropiado utilizar para determinar la probabilidad de que ocurran los siguientes eventos?

El precio del barril de petróleo, hoy termine con un precio más alto

Una computadora producida en una línea de ensamblaje sea defectuosa

Sacar 8 al sumar un lanzamiento de dos dados

Aprobar este módulo de Estadística, sin estudiar.

Ponga dos ejemplos aplicados a los negocios o economía en los cuales se aplicaría cada uno de los tres enfoques de probabilidad

La siguiente tabla indica el número de equipos de sonido vendidos por un distribuidor minorista.

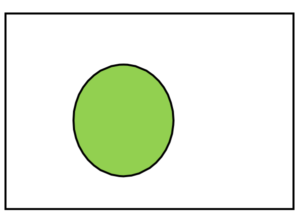
Número de equipos	**Número de días**
0	10
1	35
2	25
3	7
4	3

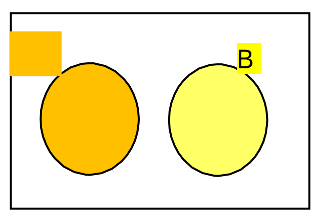
Determinar la probabilidad de que el número de equipos de sonido que se vendan hoy, sea:

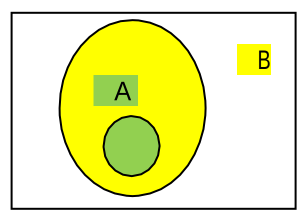
a. 3
b. Menos de 2
c. Más de 1
d. Por lo menos 2

En la siguiente figura se grafica la interpretación que se puede dar a eventos dentro de un espacio muestral, tal como se hace con los conjuntos dentro de un conjunto universo.

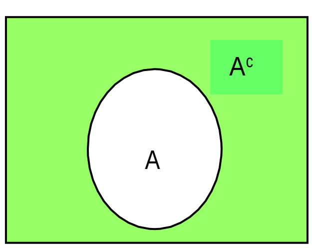
A

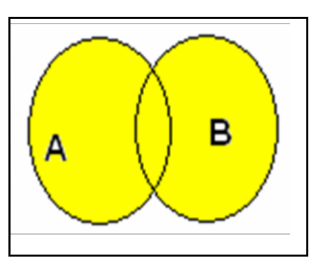


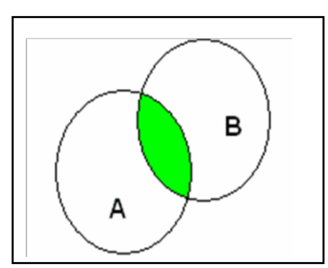




a. Ocurre el evento A b. Eventos mutuamente excluyentes c. Si ocurre A también ocurre B







d. No ocurre A e. Ocurre A u ocurre f. Ocurre A y B (**A** U **B**)

Leyes de las probabilidades

Si es necesario calcular la probabilidad de que ocurran eventos al mismo tiempo, u ocurran un evento u otros, debemos utilizar ciertas reglas que nos permitirán calcular dicha probabilidad. Existen dos leyes fundamentales para este caso de eventos complejos. La regla de la multiplicación y la regla de la adición. En el primer caso nos permite conocer la probabilidad del evento conjunto cuando ocurren 2 o más eventos al mismo tiempo y en el segundo caso cuando sucede un evento u otro.

Reglas de la Multiplicación

Para aplicar la regla especial de la multiplicación es necesario que los eventos sean independientes. Se consideran eventos independientes aquellos que la ocurrencia de uno de ellos no afecta o modifica la probabilidad de que ocurra el siguiente. Si existen dos eventos ***A*** y ***B*** y son independientes, la probabilidad de que sucedan los dos eventos se obtiene de multiplicar las probabilidades individuales de cada uno de ellos, expresado en forma simbólica tenemos:

$P(A \text{ y } B) = P(A \cap B) = P(A)\, P(B)$

Un ejemplo muy claro de la aplicación de esta regla es el lanzamiento de monedas; si se lanzan al aire dos monedas la probabilidad de que la una moneda caiga con cara hacia arriba, en nada modifica o altera la probabilidad de que la otra moneda caiga con cualquiera de las caras hacia arriba. La expresión de cálculo arriba indicada puede ser extensiva para más eventos

Ejemplos:

1. Se lanzan dos dados. ¿Cuál es la probabilidad de que ambos dados caigan con 6 hacia arriba?

Solución: Sea el evento ***A*** que el primer dado caiga con seis y el evento ***B*** que el segundo dado también salga seis. Analizamos y concluimos que los eventos son independientes por cuanto la ocurrencia del un evento no altera la probabilidad de que el otro ocurra; por lo tanto:

$P(A) = P(6) = 1/6$

$P(B) = P(6) = 1/6$

$P(A \text{ y } B) = P(6 \text{ y } 6) = (1/6)\,(1/6) = 1/36$

2. Se extrae dos cartas de un naipe, una a continuación de la otra y el experimento se realiza con sustitución, es decir, la primera carta escogida luego de verificada su denominación, vuelve a ser colocada dentro del naipe y a continuación se extrae la segunda carta. Calcular que las dos cartas escogidas sean ases.

Solución: Como el experimento se realiza con sustitución, los eventos son independientes, por lo tanto las probabilidades de cada evento se multiplican entre sí:

$P(as) = 4/52 = 1/13$

$P(as\ y\ as) = (1/13)\ (1/13) = 1/169$

3. Un almacén tiene registrado que en promedio el 30% de los clientes que entran al local, realizan una compra efectiva, determinar la probabilidad de que los tres primeros clientes que entren realicen una compra.

Solución: Se trata de igual manera de eventos independientes, porque la adquisición o no de un cliente no modifica la intención de compra de otro cliente, por lo tanto multiplicamos las probabilidades de cada evento.

P (1º compre y el 2º compre y el 3º compre)=0.3 (0.3) (0.3)=0.027= 2.7%

Si los eventos son dependientes, es decir; la probabilidad de ocurrencia de cada uno altera o modifica la probabilidad de que los otros ocurran, se utiliza la regla general de multiplicación. Un claro ejemplo de este tipo de eventos se da cuando el experimento se realiza sin sustitución, esto significa, que el elemento seleccionado no vuelve a formar parte del espacio muestral para la siguiente selección. La expresión matemática que simboliza el cálculo de esta probabilidad conjunta es:

$$P(A\ y\ B) = P(A)\ P(B \mid A)$$

La expresión: ***P (B I A)*** se lee la probabilidad de que ocurra B, dado que ocurrió A. El trazo que separa a los dos eventos significa "**dado que**" y no división. A esta condición se le denomina ***Probabilidad condicional***.

La expresión de cálculo también es válida para más de dos eventos

P(A y B y C) = P(A) P (B / A) P(C / A y B)

Ejemplos:

1. Se extraen dos cartas de un naipe, una a continuación de la otra. El experimento se realiza sin sustitución. Calcular la probabilidad de que las dos cartas seleccionadas sean cinco.

Solución: De acuerdo al enunciado del problema el experimento es sin sustitución, por lo tanto los eventos son dependientes, lo cual es evidente, ya que al extraer la primera carta y no ser devuelta al naipe, la probabilidad de extraer otra carta ya no es la misma original.

Primera carta: 5 P (5) = 4/52 =1/13

Segunda carta: 5 P (5 / 5) = 3/51 = 1/17, entonces

$$P\ (5\ y\ 5) = P\ (5)\ P\ (5\ /\ 5) = (1/13)\ (1/17) = 1/221$$

2. Una caja contiene 20 chips, de los cuales se sabe que 4 de ellos no corresponden al mismo modelo, si se sacan tres chips de esta caja uno a continuación de otro, determinar la probabilidad de que los tres sean de modelo distinto.

Solución: La probabilidad de que el primer chip sea de otro modelo es 4/20, la probabilidad de que el segundo sea igual que el primero es 3/19 y la probabilidad de que el tercer chip seleccionado también sea igual a los anteriores es 2/18, por lo tanto, la probabilidad de que los tres chips seleccionados sean de modelo diferente a los otros 16 que están en la caja, es:

Sea el evento ***A*** que el chip sea de modelo distinto, entonces la probabilidad de:

$$P(A\ y\ A\ y\ A) = (4/20)\ (3/19)\ (2/18) = 1/285$$

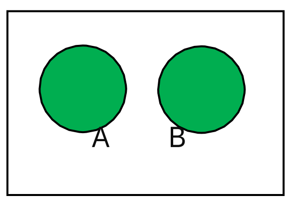
Reglas de la Adición

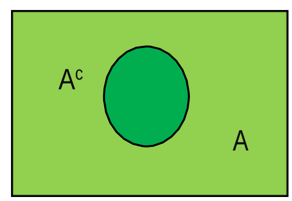
Esta regla nos sirve para calcular la probabilidad de acontezca un evento u otro, con la condición de que los eventos pueden ser mutuamente excluyentes o no mutuamente excluyentes. Se consideran eventos mutuamente excluyentes aquellos eventos que al ocurrir uno de ellos, elimina la posibilidad de ocurrencia de los otros eventos al mismo tiempo. Si los eventos pueden ocurrir al mismo tiempo, se los considera como no mutuamente excluyentes.

Si los eventos son mutuamente excluyentes, se aplica la regla especial de la adición, que indica que la probabilidad de que ocurra un evento ***A*** o un evento ***B*** es igual a la suma de las probabilidades individuales de cada evento.

P(A o B) = P (A U B) = P (A) + P (B)

El diagrama siguiente ilustra de mejor manera esta condición:





La probabilidad de que suceda A o suceda B es la zona sombreada

La probabilidad de que ocurra A o su complemento (A^c) es igual al espacio muestral

La expresión de cálculo de que ocurra el evento A o su complemento determina que es igual a uno.

P(A) + P (A^C) = 1, entonces ***P(A) = 1 – P (A^C)***

Ejemplos:

1. Un experimento consiste en extraer de un ánfora una esfera. Si se conoce que en la misma hay 4 esferas rojas y 3 blancas y 3 azules, determinar la probabilidad de que al extraer esa esfera, sea roja o blanca.

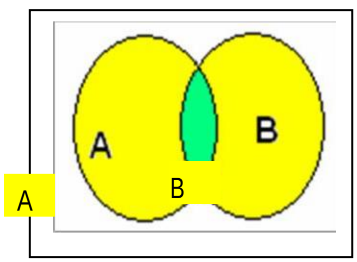
Solución: Definimos el evento A de que la esfera sea roja y el evento B que la esfera sea blanca, entonces la probabilidad de que la esfera extraída sea roja es 4/10 y la probabilidad de extraer una esfera blanca es 3/10. Como son eventos mutuamente excluyentes, ya que al extraer una sola esfera ésta es o roja o blanca o azul. Entonces la probabilidad de que la esfera extraída sea roja o blanca es:

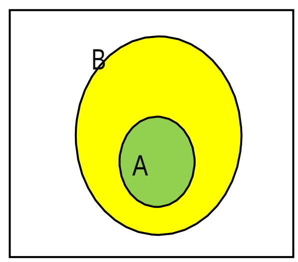
P (roja o blanca) = 4/10 + 3/10 = 7/10 = 0.7

Si los eventos no son mutuamente excluyentes entonces se aplica la regla general de la adición, que indica de que la probabilidad de que ocurra un evento ***A*** o un evento ***B*** es igual a la suma de las probabilidades individuales de cada evento y se le resta la probabilidad conjunta de que ocurran los dos eventos al mismo tiempo. La expresión matemática que nos indica esta condición es:

P(A o B) = P(A U B) = P(A) + P (B) – P(A ∩ B)

Gráficamente podemos advertir que la unión de sucesos no mutuamente excluyentes es lo expresado mediante la fórmula anterior.





Es necesario descontar la probabilidad conjunta ya que está incluida en cada probabilidad individual

La probabilidad de que ocurra A o B es igual a la probabilidad de que ocurra B, por estar conteniendo el evento A en B

Calcular la probabilidad de que al extraer una carta de un naipe común, ésta sea un as o una carta de diamantes.

Solución: Se podría pensar que la respuesta es 17/52 al sumar solo las probabilidades individuales de cada evento. La probabilidad de obtener un as, es 4/52 y la de obtener una carta de diamantes es 13/52. Pero si analizamos hay un resultado que pertenece a los dos eventos (as de diamantes) y está contabilizado en cada probabilidad, por lo tanto debemos descontar esta probabilidad conjunta.

P(as o carta de diamante) = P(as) + P(diamante) – P(as de diamante) P(as o

diamante) = 4/52 + 17/52 – 1/52 = 16/52 ~ 0.308

Observación: Las reglas de la adición también pueden aplicarse a tres o más eventos.

La Regla de Laplace

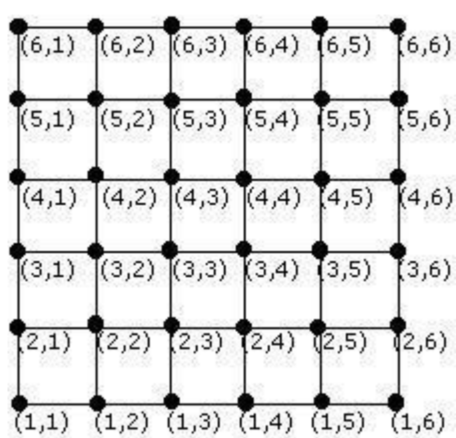
En ciertos casos es posible que se pueda predecir el resultado de un suceso (si va a ocurrir o no). Si se puede predecir, diremos que es un fenómeno determinístico. En caso contrario, se trataría de un evento aleatorio. Por ejemplo, lanzar una moneda al aire constituye un fenómeno de tipo aleatorio, pues en este caso no se puede asegurar si saldrá cara o sello.

La estadística es una rama de las matemáticas que estudia la probabilidad de resultados posibles de un determinado evento o suceso.

Se define espacio muestral como el conjunto de todos los resultados posibles de un experimento aleatorio, y lo designaremos con la letra E. En el caso de la moneda, los resultados posibles son 2: cara o sello, por lo tanto, el espacio muestral de este suceso es E = {cara, sello}.

Al arrojar un dado, existen 6 caras posibles, cada una de las cuales es un resultado. El espacio muestral para este caso es E = {1, 2,3,4,5,6 }

Si se tiran dos dados distintos, su espacio muestral puede ilustrarse mediante el siguiente diagrama de pares ordenados:



En este caso, el espacio muestral está formado por 36 elementos.

Se llama evento a todo subconjunto del espacio muestral. Por ejemplo, si se tiran dos dados, un evento puede ser que la suma de las puntuaciones sea igual a seis. En este caso, solo serán los pares (1,5) (5,1) (2,4) (4,2) y (3,3).

La probabilidad de un evento es un valor que nos permite determinar qué tan posible es que un evento ocurrao no.

La definición clásica de probabilidad, dada por la regla de Laplace, se aplica si todos los resultados posibles de un experimento aleatorio tienen la misma probabilidad o son equiprobables.

En el diagrama anterior, donde se ilustra el espacio muestral para el experimento aleatorio de lanzar dos dados, cada uno de los resultados tiene la misma probabilidad si consideramos que los dados “no están cargados”.

La probabilidad de que un evento A ocurra se anota P(A) y se calcula mediante el cociente:

$$P(A) = \frac{\text{número de casos favorables a A}}{\text{número de casos totales}}$$

Esto es válido, eso sí, cuando en el experimento aleatorio todos sus resultados son equiprobables. Los valores de una probabilidad están entre 0 y 1.

Establece que:

La probabilidad de ocurrencia de un suceso *imposible* es 0.

La probabilidad de ocurrencia de un suceso *seguro* es 1, es decir, P(A) = 1.

Para aplicar la regla de Laplace es necesario que los experimentos den lugar a sucesos equiprobables, es decir, que todos tengan o posean la misma probabilidad.

La probabilidad de que ocurra un suceso se calcula así:

P(A) = Nº de casos favorables / Nº de resultados posibles

Esto significa que, probabilidad es igual al número de casos favorables sobre o dividido el número de resultados total de resultados posibles.

Distribución binomial

La probabilidad de ocurrencia de una combinación específica de eventos independientes y mutuamente excluyentes se determina con la distribución binomial, que es aquella donde hay solo dos posibilidades, tales como masculino/femenino o si/no.

Hay dos resultados posibles mutuamente excluyentes en cada ensayo u observación. La serie

de ensayos u observaciones constituyen eventos independientes.

La probabilidad de éxito permanece constante de ensayo a ensayo, es decir el proceso es estacionario.

Para aplicar esta distribución al calculo de la probabilidad de obtener un número dado de éxitos en una serie de experimentos en un proceso de Bernoulli, se requieren tres valores: el número designado de éxitos (m), el número de ensayos y observaciones (n); y la probabilidad de éxito en cada ensayo (p).

Entonces la probabilidad de que ocurran m éxitos en un experimento de n ensayos es: $P(x = m) = (nCm)(P^m)(1-P)^{n-m}$

Siendo: nCm el número total de combinaciones posibles de m elementos en un conjunto de n elementos.

En otras palabras $P(x = m) = [n!/(m!(n-m)!)](p^m)(1-p)^{n-m}$

Ejemplo:

La entrada al cine por lo menos tendrá un costo de 10 soles (como mínimo podría costar 10 soles o más).

b.− la probabilidad de que aprueben mas de 12 es $P(x > 12)$ es decir, que: $P(x > 12) = P(x = 13)+P(x = 14)+P(x = 15)$ $P(x > 12) = 1{,}47 \cdot 10^{-9} + 3{,}722 \cdot 10^{-11} + 4{,}38 \cdot 10^{-13} = 1{,}507 \cdot 10^{-9}$

La esperanza matemática en una distribución binomial puede expresarse como: $E(x) = np = 15(0{,}15)=2{,}25$

Y la varianza del número esperado de éxitos se puede calcular directamente: $Var(x) = np(1-p)= 15(0{,}15)(1-0{,}15)=1{,}9125$

Tablas de contingencia

Las tablas de contingencia: y probabilidad son usadas en estadística para mostrar la relación entre dos variables generalmente de tipo cualitativo (datos nominales u ordinales) y estas variables desglosadas de acuerdo a la categorización que sea necesaria. Esta relación se establece para un mismo elemento de estudio.

Por ejemplo a una persona se la puede identificar con su género (femenino o masculino) y con el credo religioso (cristiano, judío, budista, etc.).

Con esta condición se genera una tabla donde la columna izquierda sirve para poner la primera variable con su categorización y la fila superior la segunda con su categorización, en la fila inferior y en la columna derecha se indican los totales de cada condición y toman el nombre de frecuencias marginales. En la celda inferior derecha está el gran total de observaciones del estudio.

Ejemplo

Los 500 empleados de King Dynamics, están clasificados como miembros del personal administrativo, trabajadores de línea y personal auxiliar y de acuerdo al género, tal como se muestra en la siguiente tabla de contingencias

CLASIFICACIÓN DE LOS EMPLEADOS

GÉNERO	Personal	Línea	Auxiliar	Total
Hombres	120	150	30	300
Mujeres	50	140	10	200
Total	170	290	40	**500**

4 WEBSTER, Allen. Estadística aplicada a los negocios y economía. Pág. 83. Irwin McGraw - Hill. 3ª edición

Del análisis de esta tabla se puede concluir, por ejemplo que existen 170 empleados que pertenecen al personal administrativo, de los cuales 120 son hombres y 50 mujeres, o que existen en la empresa un total de 200 mujeres repartidas 50 en el personal administrativo, 140 como trabajadoras de línea y 10 como personal auxiliar. Con esta información de la tabla se puede pasar a una tabla de probabilidades, simplemente dividiendo cada valor de las celdas para el total de empleados de la empresa. A continuación consta la respectiva tabla de probabilidades.

CLASIFICACIÓN DE LOS EMPLEADOS

GÉNERO	Personal	Línea	Auxiliar	Total
Hombres	120/500 = 0.24	150/500= 0.30	30/500 = 0.06	300/500 = 0.6
Mujeres	50/500 = 0.10	140/500= 0.28	10/500 = 0.02	200/500 = 0.4
Total	170/500 = 0.34	290/500= 0.58	40/500 = 0.08	**500/500 = 1.0**

Estas tablas son muy útiles para la determinación de la probabilidad de los eventos que están inmersos en ellas.

Resúmen

En esta unidad se abordó maneras de cómo asignar probabilidades, los criterios más usados para esta asignación, las reglas que rigen a las probabilidades, tablas de contingencia y probabilidad entre otros temas. La importancia que tiene este tema dentro de la estadística es fundamental, ya que el uso de muestras para estimar parámetros poblacionales, se realiza con incertidumbre, la reducción de esta incertidumbre se reduce con la ponderación que el caso amerita de acuerdo a una probabilidad asignada.

Se ha determinado también la metodología para establecer los resultados posibles de un experimento, cuando no es posible su determinación directa, para esos casos se recurren a las técnicas de conteo: multiplicación de opciones, variaciones y combinaciones.

Distribuciones Bidimensionales

Es de interés considerar experimentos aleatorios a los cuales se les asignan dos variables aleatorias de entrada, relacionadas o no, que permiten definir las Distribuciones de Probabilidad Bidimensionales o Bivariadas

Variable Aleatoria Bidimensional

Definición: Sea S el espacio muestral asociado a un experimento aleatorio dado. Sean X = X(s) y Y = Y(s) dos funciones que asignan un número real a cada resultado s $\in$ S. Llamaremos a (X,Y) variable aleatoria bidimensional –vab-.

Simbólicamente tenemos

$$\boxed{\begin{array}{lcc} S & \longrightarrow & R \subset \Re^2 \\ \in & & \in \\ s & \longrightarrow & \overbrace{(X[s], Y[s])} \end{array}}$$

Observaciones

1. Interesan más que la naturaleza de las funciones X y Y los valores que asumen así:

$\boxed{(X[s], Y[s]) \equiv (X,Y)}$

2. El recorrido de la v a b (X,Y) es R $\subset \Re^2$

3. P (X[s] $\leq$ a, Y[s] $\leq$ b) $\equiv$ P (X $\leq$ a, Y $\leq$ b)

4. (X,Y) es $\begin{cases} \text{v a b discreta si R es numerable} \\ \text{v a b continua si R es infinito no numerable} \end{cases}$

(X,Y) puede ser vab mixta, o sea, continua y discreta por tramos.

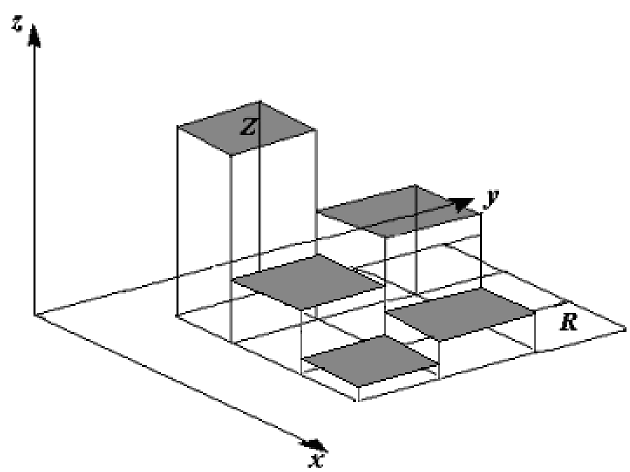
Función De Probabilidad Conjunta Bivariada

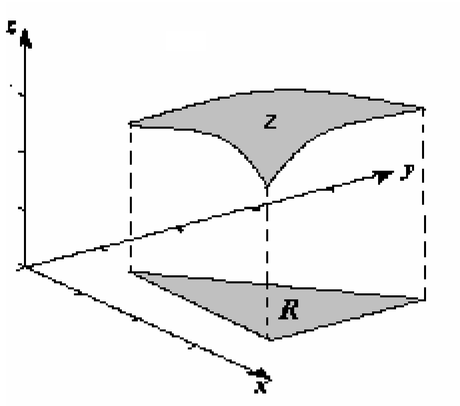
Definición: La función que asocia a cada resultado $(X, Y) \in R \subset \Re^2$ un numero real f(x,y) se denomina función de probabilidad conjunta bivariada si:

1. $f(x,y) \geq 0$ para todo $(x,y) \in \Re^2$

2. $$1 = \begin{cases} \displaystyle\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} f(x_i, y_j) & \text{si } (X,Y) \text{ es discreta} \\ \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy & \text{si } (X,Y) \text{ es continua} \end{cases}$$

Observaciones





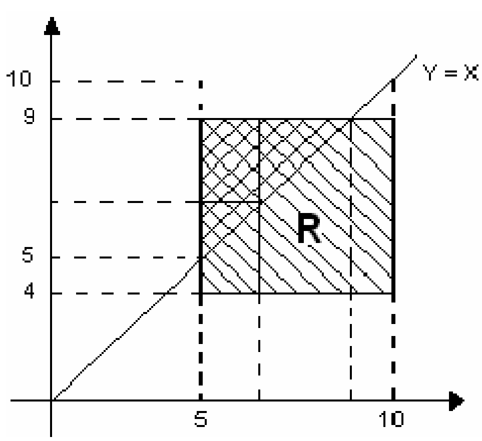
- El volumen acotado por la superficie z = f(x,y) y la región R vale 1

- La proyección de z = f(x,y) sobre el plano x,y es la región dominio $R \subset \Re^2$ donde f(x, y)>0 así es claro que

$$0 \begin{cases} < \mathbf{f(x, y)} & \mathbf{si\ (x, y)} \in \mathbf{R} \\ = \mathbf{f(x, y)} & \mathbf{si\ (x, y)} \notin \mathbf{R} \end{cases}$$

Ejemplo

Si f(x, y) es una fdp divariada definida positivamente para todo (x,y)ЄR=[5,10]X[4,9] en la figura siguiente, Halle c y P($X \leq Y$).



$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dxdy = 1 \text{ entonces } c \int_{4}^{9}\int_{5}^{10} dxdy = 1 \;\Rightarrow\; C = \frac{1}{25}$$

$$P(X \leq Y) = \frac{1}{25}\int_{5}^{9}\int_{5}^{y} dxdy \quad \text{ó} \quad \frac{1}{25}\int_{5}^{9}\int_{x}^{9} dydx$$

$$= \frac{1}{25}\int_{5}^{9}(y-5)\,dy \quad \text{ó} \quad \frac{1}{25}\int_{5}^{9}(9-x)\,dx$$

$$= \frac{1}{25}\left[\frac{y^2}{2} - 5y\right]_5^9 \quad \text{ó} \quad \frac{1}{25}\left[9x - \frac{x^2}{2}\right]_5^9$$

$$P(X \leq Y) = \frac{8}{25}$$

Obteniendo el mismo resultado.

Función De Probabilidad Acumulativa Bidimensional

Sea (X,Y) una v a b con fdp f(x,y), entonces su función de distribución acumulada es:

$\boxed{F(X,Y) = P(X \leq x, Y \leq y)}$

Observación

Asi como $\frac{dF(x)}{dx} = f(x)$ para X vac unidimensional se tiene que $\frac{\partial^2 F(x,y)}{\partial x \partial y} = f(x,y)$ donde quiera que **F(x,y)** es diferenciable.

Distribución Uniforme Bivariada

Decimos que (X,Y) v a b se distribuye uniformemente en R $\subset \Re^2$ si

$$\mathbf{f(x,y)} = \begin{cases} \frac{1}{K} & \text{si } (x, y) \in R \\ 0 & \text{si } (x, y) \notin R \end{cases}$$

Es claro que: K es el número finito de puntos de R si (X, Y) es una v a b discreta ó K es el inverso del área finita de la región R si (X, Y) es una vab continua **Ejercicio**

Sea R = {(x,y) : $0 \leq x \leq 1 \;\Lambda\; x^2 \leq y \leq x$} si **f(x,y)** es una fdp uniforme bidimensional definida en R. Halle f y pruebe que:

a. $\int_0^1 \mathbf{f(x,y)}dy = g(x) = 6(x-x^2), 0 \leq x \leq 1$

b. $\int_0^1 \mathbf{f(x,y)}dx = h(y) = 6(\sqrt{y} - y), 0 \leq y \leq 1$

¿Cómo es la gráfica de **f, g** y **h**?

Función De Probabilidad Marginal

En el ejercicio anterior a g(x) y h(y) se les denomina fdp marginales de las vac unidimensionales X y Y, diremos en general que si f(x,y) es la fdp conjunta de una vac bibimensional (x,y) entonces

g(x) = $\int_{-\infty}^{\infty} f(x, y)\, dy$ y **h(y) =** $\int_{-\infty}^{\infty} f(x, y)\, dx$

son la fdp marginales de las va unidimensionales x y y**.**

Asi P ($c \le x \le d$) = P ($c \le x \le d$, $-\infty \le Y \le \infty$)

$$= \int_c^d \int_{-\infty}^{\infty} f(x, y)\, dydx = \int_c^d g(x)\, dx$$

Ejercicio

Sea $f(x,y)=2(x+y-2xy)$ $0 \le x \le 1$, $0 \le y \le 1$. Halle las fdp marginales de las vac X y Y, y pruebe que son fdp.

En efecto $\int_0^1 g(x)dx = \int_0^1 h(y)dy = 1$

Ejercicio

Sea $f(x,y)=x^2+\frac{xy}{3}$ $0 \le x \le 1$, $0 \le y \le 2$.

Grafique la región R donde f es positiva, demuestre que f es una fdp, pruebe que P ($X + Y \ge 1$) = $\frac{65}{72}$ y halle las fdp marginales de X y Y.

Funciones De Probabilidad Condicional

Sea (X,Y) una v a b c con una fdp conjunta f. Sean g(x) y h(y) las fdp marginales de las v a c X y Y entonces las fdp condicionales de X dado Y = y, y de Y dado X = x

Son

g(x/y) = $\frac{f(x, y)}{h(y)}$, **h(y)** > 0

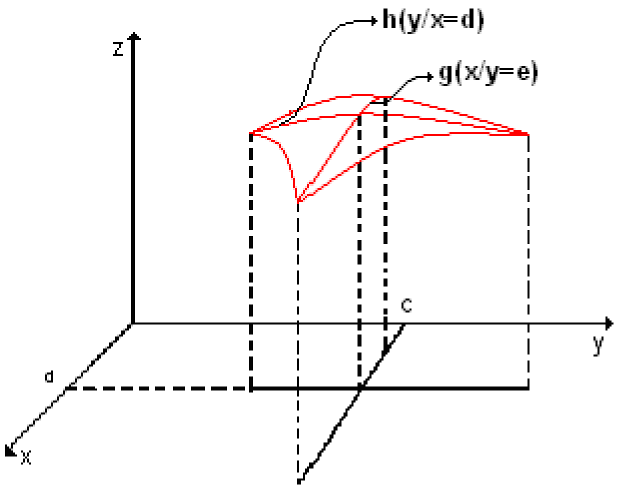
h(y/x) = $\frac{f(x, y)}{g(x)}$, **g(x)** >0

Observe que:

1. $\int_{-\infty}^{\infty} g(x/y)\, dx = \int_{\infty}^{\infty} h(y/x)\, dy = 1$ es

decir g y h marginales son fdp.

2. g(x/y) es la intercepción de f con el plano Y=C. Observe que la gráfica siguiente



Funciones De Probabilidad Bivariada Discreta

Basta, por analogía sustituir, en las definiciones del caso bidimensional continuo las integrales por sumatorias.

Ejemplo

Supóngase una fdp conjunta bivariada que asume valores de probabilidad conjunta según la siguiente tabla:

	X_1=2	X_2=3	X_3=4	X_4=5	X_5=6	
Y_1=2	1	2	3	0	1	7
Y_2=3	4	1	4	2	0	11
Y_3=4	3	2	1	0	1	7
	8	5	8	2	2	25

Podemos calcular algunas probabilidades así:

$$P\,(X = Y) = f(x_1,y_1) + f(x_2, y_2) + f(x_3, y_3) = \frac{3}{25}$$

$$P\,(X \le 3,\ Y \le 3) = \sum_{i=1}^{2}\sum_{j=1}^{2} f(x_i, y_j) = \frac{8}{25}$$

$$P\,(X \le Y) = \sum_{x_i \le y_j} f(x_i, y_j) = \frac{12}{25}$$

$$P\,(X = 3) = P\,(X = 3\,;\ Y = 2,3,4) = \frac{5}{25}$$

Es decir $g(x_i) = P(X = x_i, Y = y_1, y_2, \ldots, y_k)$

$$g(x_i) = \sum_{j=1}^{\infty} f(x_i, y_j)$$

$$h(y_j) = \sum_{i=1}^{\infty} f(x_i, y_j)$$

De nuevo g y h son las fdp marginales de X y Y respectivamente

$$g(x / Y = 3) = \frac{f(x,3)}{h(3)}$$

$$= \frac{1}{11}(4,1,4,2,0)$$

Aquí g (x / 1) es la probabilidad condicional de X dado Y = 1.

Análogamente: $$h(y / X = 4) = \frac{f(4, y)}{g(4)}$$

$$= \frac{1}{8}\begin{bmatrix} 3 \\ 4 \\ 1 \end{bmatrix}$$

Ejercicio

Calcular en la misma tabla g (x / Y = 4) y h (y / X = 5)

Variables Aleatorias Independientes: Sea (X, Y) una vab las siguientes son afirmaciones equivalentes: X es independiente de y **sii f(x, y) = g (x) h(y)** **sii g(x / y) = g(x)** **sii h(y / x) = h(y)** , $\left.\begin{matrix} g(x) \\ h(y) \end{matrix}\right\} > 0$

Observe que

Ejercicio

Asignar en la siguiente tabla probabilidades conjuntas f(x, y) de forma que X y Y sean variables aleatorias independientes

y	X	7	8	9	
3					40
5					60
		40	20		

Ejercicio

Supóngase que f(x, y) = 8xy, para $0 \leq x \leq y \leq 1$ son X y Y variables aleatorias independientes?.

Covarianza

Sea X, Y una v a b con rango $R \subset \Re^2$ y fdp conjunta f(x, y), con E(X) = μ_x y E (Y) = μ_y

Definimos la covarianza de las v a X y Y asi:

C(X, Y) = σ_{xy} = E [(X - μ_x) (Y - μ_y)] o sea

$$\sigma_{xy} = \begin{cases} \sum_R \sum (x-\mu_X)(y-\mu_Y) f(x,y) & \text{si } (X,Y) \text{es} \quad \text{vabd} \\ \int_R \int (x-\mu_X)(y-\mu_Y) f(x,y) & \text{si } (X,Y) \text{es} \quad \text{vabc} \end{cases}$$

Observaciones

1) Las propiedades de operador lineal de la esperanza E permiten una definición alternativa para σ_{xy}, veamos:

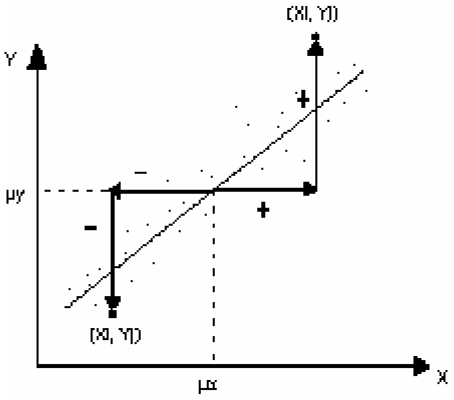
σ_{xy} = E [(X - μ_x) (Y - μ_y)]

= E [XY - Xμ_y - Yμ_x + $\mu_x\mu_y$]

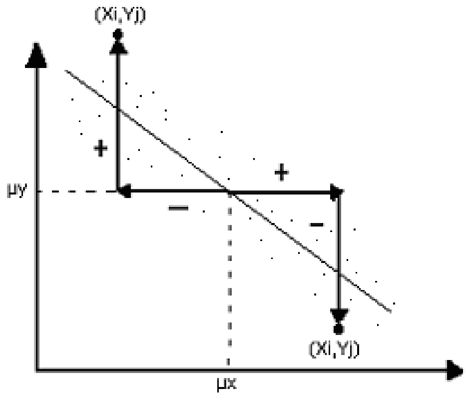
= E (XY) - $\mu_y\mu_x$ - $\mu_x\mu_y$ + $\mu_x\mu_y$, asi:

C (X, Y) = E (XY) - $\mu_x\mu_y$

2) C (X, Y) es la medida de la relacion lineal entre las v a X y Y , en efecto



(1)



(2)

Si C (X, Y) > 0 entonces

$(X - \mu_x) > 0 \;\Delta\; (Y - \mu_y) > 0$ ó $(X - \mu_x) < 0 \;\Delta\; (Y - \mu_y) < 0$. Que

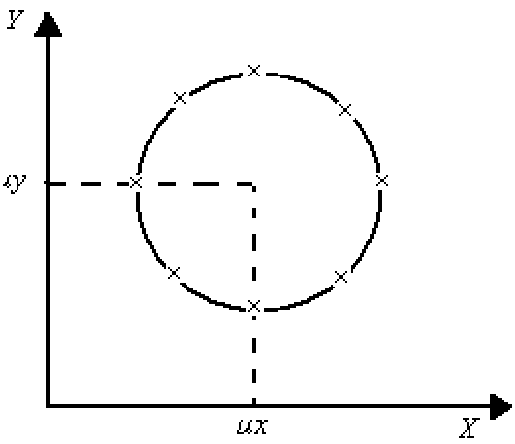
corresponde a la grafica lineal (1) dependiente positiva Puede ser

también, C (X, Y) < 0, o sea

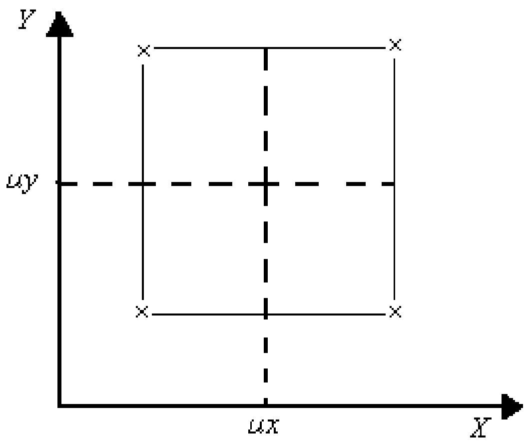
$(X - \mu_x) > 0 \;\Delta\; (Y - \mu_y) < 0$ ó $(X - \mu_x) < 0 \;\Delta\; (Y - \mu_y) < 0$. que

corresponde a la línea dependiente negativa, grafica (2)

También puede ser C (X, Y) = 0 para distribuciones simétricas respecto al punto (μ_x, μ_y) como en las graficas (3) y (4)



(3)



(4)

Correlación

Sean X y Y dos v a de forma que se pueden calcular V(X), V(Y) y C(X, Y). Definimos el coeficiente de correlación de las v a X y Y como

$$\boxed{\textbf{Corr (X, Y)} = \rho_{XY} = \rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}}$$

Como se puede observar en la definición la correlación permite escalar la covarianza en unidades de la desviación estándar de cada variable, precisando la comparación de la linealidad para v a de unidades distintas.

Para a,b,c y d constantes reales, a y c con el mismo signo, se puede probar mediante las propiedades de la esperanza, la varianza y la covarianza que:

a) $\boxed{\textbf{Corr (aX + b, cY + d) = Corr (X, Y}}$**).** Es decir que el coeficiente de correlación es invariante ante un cambio de origen y escala de las variables aleatorias.

b) Para dos v a X y Y cualquiera

$\boxed{\mathbf{-1 \leq Corr(X, Y) \leq 1}}$

es decir que la fuerza de la linealidad entre las v a X y Y se mueve entre -1 y 1 y definimos el siguiente criterio.

Criterio De Correlación Lineal: Diremos que la relación lineal entre las v a X y Y es:

Inexistente si $|\rho_{XY}| = 0$

Débil si $0 < |\rho_{XY}| \leq 0.5$

Moderada si $0.5 < |\rho_{XY}| < 0.8$

Fuerte si $0.8 \leq |\rho_{XY}|$

Observaciones.

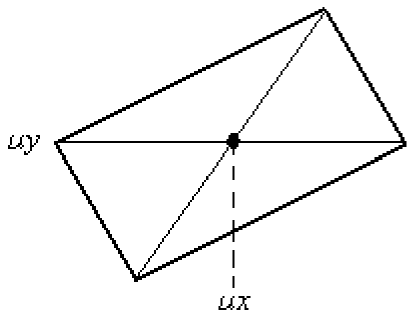
- La relación entre dos variables X y Y no está completamente explicada por Corr(X,Y), solo su relación lineal.
- Si X y Y son independientes entonces la independencia implica que $\rho_{XY} = 0$. Pero $\rho_{XY} = 0$ no implica que X y Y sean independientes $\rho_{XY} = \pm 1$ sii $Y = aX + b$ a y b reales $a \neq 0$, o sea, todos los puntos están sobre la recta.

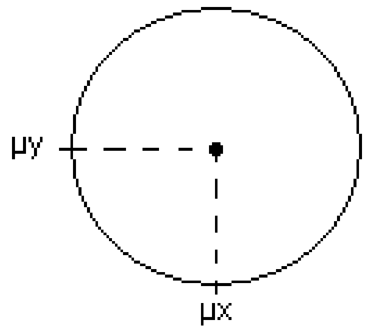
Ejemplo

Sea $f(X, Y) = \frac{1}{4}$ $\quad (X, Y) = \begin{cases} (-4,1) \\ (4,-1) \\ (2,2) \\ (-2,-2) \end{cases}$ y sea $f(x,y) = \frac{1}{12}$, $(x,y) \in Z^2$ $x^2+y^2=25$

Observe que $E(X) = E(Y) = C(X, Y) = 0$. pero X y Y no son independientes. X no se relaciona linealmente con Y

¿Cómo es la simetría respecto a (μ_x, μ_y)?





En estos caso por la simetría $\rho_{XY} = 0$ pero no se puede afirmar que X y Y sean independientes, solo que no están relacionadas linealmente.

Problemas Seleccionados

Una pareja se cita entre las 7 y las 8 de la tarde y llegan a la cita con distribución uniforme en dicho intervalo. Deciden esperarse un máximo de 15 minutos. Calcular la probabilidad de que se encuentren.

La variable bidimensional (x, y) tiene como función de densidad $f(x, y) = e^{-(x+y)}$ en el primer cuadrante; f(x, y)=0 en los otros tres. Si se toman al azar tres puntos en el primer cuadrante calcular la probabilidad de que uno al menos pertenezca al cuadrado $(0 \le x \le 1; 0 \le y \le 1)$.

El tiempo total que un camión permanece en un almacén está definido por una variable aleatoria x. Sea y la variable tiempo de espera en la cola, y z el tiempo de descarga (x = y + z). La distribución conjunta de x e y es:

$$f(x, y) = \begin{cases} \frac{1}{4} e^{-x/2} & 0 \le y \le x < \infty \\ 0 & \text{en otro caso} \end{cases}$$

Se pide:

Calcular el tiempo medio total que permanece un camión en la estación. Calcular

el tiempo medio de descarga.

Calcular el coeficiente de correlación entre el tiempo total y el tiempo de espera en la cola.

Una máquina de empacado automático deposita en cada paquete 81.5g, por término medio, de cierto producto, con σ=8g. El peso medio del paquete vacío es 14.5g, con σ=6g. Ambas distribuciones son normales e independientes. Se pide:

Calcular la distribución del peso de los paquetes llenos.

Escribir la distribución conjunta del peso del paquete y el producto que contiene.

En el problema anterior los paquetes se distribuyen en cajas de 40, cuyo peso medio vacías es 520g, con σ=50g. Calcular:

La distribución del peso de las cajas llenas.

La probabilidad de que un cajón vacío pese menos que 5 paquetes llenos.

Una línea eléctrica se avería cuando la tensión sobrepasa la capacidad de la línea. Si la tensión es N(100, 20) y la capacidad N(140, 10), calcular la probabilidad de avería. Suponiendo que la tensión y la capacidad varían independientemente.

Dada f(x, y)=3x (0<y<x; 0<x<1), obtener las distribuciones marginales y la condicionada f(x/y).

La función de probabilidad de (x, y) es p(x,y)=1/30 para x=0, 1, 2, 3, 4, 5 e y=0, 1, 2, 3, 4; p(x,y)=0 en puntos distintos de los anteriores. Calcular la función de distribución en los puntos de la recta x-2y+2=0.

En un aparato de control actúan dos variables x_1, x_2 independientes, ambas con distribución uniforme, la primera entre 1 y 9, la segunda entre 1 y a. El aparato funciona bien cuando $x_1<4x_2^2$. Calcular el valor de a para que $p(x_1>4x_2^2)\leq 0.01$.

Se toman tres mediciones independientes y_1, y_2, y_3 de la tensión en un circuito con tres aparatos, cuyas varianzas son 1, 2 y 3. SDe forman dos índices del circuito por:

$$z_1 = 3y_1 + 2y_2 + 5y_3$$
$$z_2 = \frac{1}{3}y_1 + \frac{1}{3}y_2 + \frac{1}{3}y_3$$

Calcular el coeficiente de correlación entre z_1 y z_2.

Resumen

Si es necesario calcular la probabilidad de que ocurran eventos al mismo tiempo, u ocurran un evento u otros, debemos utilizar ciertas reglas que nos permitirán calcular dicha probabilidad. Existen dos leyes fundamentales para este caso de eventos complejos. La regla

de la multiplicación y la regla de la adición. En el primer caso nos permite conocer la probabilidad del evento conjunto cuando ocurren 2 o más eventos al mismo tiempo y en el segundo caso cuando sucede un evento u otro.

2.4. Distribucion de probabilidades

En la naturaleza y en la vida cotidiana existen fenómenos que no se pueden determinar su resultado con certeza, más bien se está en el posibilidad de anticipar el mismo con un cierto grado de probabilidad de ocurrencia, ejemplo determinar a qué edad una persona fallecerá, el interés que percibirá un accionista luego de tres años de realizada una inversión. Es lógico que nadie pueda dar una respuesta a estos dos eventos, pero si se establece algún valor, existe una incertidumbre de su validez. Para satisfacer la necesidad de la obtención de resultados en sucesos donde está involucrado el azar, se desarrolló la teoría de probabilidades.

Desde la antigüedad en varias culturas aparecen referencias sobre el estudio de fenómenos aleatorios. El estudio más pormenorizado de las probabilidades aparece en los siglos XVI Y XVII, debido a las preguntas que los jugadores de azar de aquella época realizaron a prestantes estudiosos matemáticos, tales como Galileo, Pascal y Fermat. Huygens, publica el libro

Razonamientos relativos al juego de dados (Ratiociniis in ludo alae), que se convierte en el primer libro escrito sobre probabilidades de la historia.

Ya en el siglo XVIII, el cálculo de probabilidades se extendió a problemas físicos y de seguros marítimos, constituyéndose como factor principal de desarrollo de la teoría de probabilidades el conjunto de problemas de astronomía y de física que aparecieron relacionados con la verificación empírica de la teoría de Newton. Pierre Simón Laplace, introdujo la primera definición explícita de probabilidad y desarrolló la ley normal como modelo para la descripción de la variabilidad de los errores de mediación.

En el siglo XIX, tanto matemáticos como astrónomos continuaron ampliando la teoría probabilística, que permitieron su consolidación como una rama científica. Pese a ello sus aplicaciones se daban mayoritariamente en la física y astronomía. Posteriormente ya en el siglo XX, A. N. Kolmogorov en 1933 estableció una descripción axiomática de la idea de probabilidad, lo que constituyó la base de la moderna teoría, consiguiéndose de esta manera dar aplicación a muchas ciencias y campos de la vida cotidiana. En las últimas décadas, el uso de la teoría de probabilidades, en las ciencias sociales, naturales, cálculos actuariales o en la economía ha crecido enormemente, debido a estas circunstancias su estudio y conocimiento es una necesidad imprescindible e impostergable.

Distribución Hipergeometrica

Distribución de Poisson

2.5. Distribución normal y normal estándar

Introducción

En la unidad anterior se vio tres familias de distribuciones para variable discreta, en esta unidad estudiaremos una distribución diseñada para variables aleatorias continuas, se trata de la distribución normal, distribución que tiene mucha aplicación a numerosas características humanas, así como en variables que se utilizan en negocios y la industria; conforme se avance en el estudio de la estadística se irá verificando la variada aplicación de esta distribución.

El descubrimiento de la distribución normal o llamada también como *curva de errores*, se le asigna por lo general al matemático Karl Gauss (1755 – 1855), el mismo que reconoció que los errores de mediciones iteradas de objetos, están generalmente bajo un mismo patrón, al que lo llamó *curva normal de error*. Por este motivo a esta distribución también se la denomina distribución de Gauss y la curva representativa de la misma como *Campana de Gauss*. En menor medida otros matemáticos también contribuyeron con conocimientos previos a esta distribución, tales como Pierre – Simon de Laplace (1749 – 1827) y Abraham de Moivre (1667 – 1754).

La distribución normal posee las siguientes características:

- Pertenece a una variable continua
- Es una distribución simétrica con relación a la media aritmética, cuyo valor coincide con la mediana y la moda de la distribución
- Es unimodal (una sola moda)
- La curva que representa a la distribución es asintótica con relación al eje horizontal
- El área bajo la curva es 1

- Existe una familia de distribuciones normales, cada una con sus propia media y desviación estándar.

La característica de simetría permite saber con antelación, que, el área que se encuentra bajo la curva se reparte de tal forma que el 50% de la misma se halla por debajo de la media y el otro 50% por encima de la media aritmética. Esta distribución de área proporciona la probabilidad de que la variable de estudio tome valores entre ciertos límites o sea mayor o menor que algún valor determinado. Adicionalmente debemos recordar que las medidas de forma tales como coeficiente de asimetría y la curtosis, toman como referente de comparación a la distribución normal.

Función de densidad de probabilidad de la distribución normal

> **Distribución Normal:** Es la función que define a la distribución normal se caracteriza por tener dos parámetros:
>
> La media aritmética (***μ***) y la desviación estándar (***σ***). Cada par de valores producen una distribución normal diferente.

La expresión de cálculo que establece a una distribución de normal es:

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

Donde:

μ: Media aritmética de la variable de estudio
σ: Desviación estándar de la variable

π: Constante matemática ~3.14159

e: Constantematemática,basedeloslogaritmosnaturales~2.71828…
2.71828…….

Esta condición determina que la distribución normal sea una familia de distribuciones, cada una con sus propios parámetros (media y desviación estándar), lo cual dificulta el cálculo de las áreas (probabilidad) comprendidas entre los límites que se tengan en un determinado problema. Es necesario el uso de cálculo integral para la determinación de las áreas (probabilidades) requeridas, si no disponemos de una computadora a nuestro alcance.

Distribución Normal Estándar: La distribución normal estándar, es la que nos permite reducir cualquier distribución normal al formato de la normal estándar, ya que ésta tiene una media aritmética y desviación estándar únicas, cuyos valores son cero y uno respectivamente.

Esta distribución también se denomina ***distribución z*** y el proceso de transformación o conversión ***estandarización***.

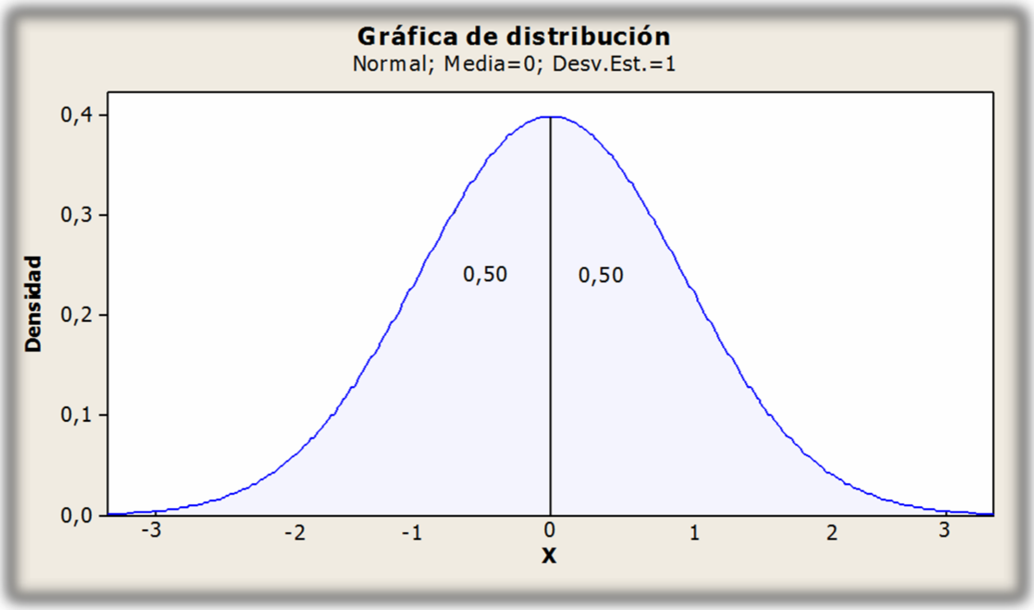
La expresión matemática que nos permitirá realizar la transformación a distribución z es:

$$z = \frac{x - \mu}{\sigma}$$

Donde "x" representa el valor de la variable que se quiere estandarizar, ***μ*** representan la media aritmética y la desviación estándar de la variable. Esta expresión nos lleva a concluir que si la variable toma un valor, igual a la media aritmética del estudio, el valor "z" correspondiente es cero; si la variable toma valores iguales por ejemplo a la media aritmética más una desviación estándar el valor "z" será igual a 1.

En el siguiente gráfico podemos ver la equivalencia entre una distribución normal y la distribución estándar.

Distribución Normal Estándar



El valor z mide la distancia entre la media aritmética y el valor específico de la variable que se estandarizó, medida en unidades de desviación estándar. El conocimiento de este valor permite conocer el área (probabilidad) que se halla ubicada en un intervalo, como ejemplo supongamos que una variable luego del proceso de estandarización, tomó el valor 0.85, entonces podemos determinar las siguientes probabilidades:

a. De que la variable tome valores comprendidos entre la media aritmética y 0.85
b. Que la variable asuma valores mayores que 0.85
c. Que la variable tome valores menores o iguales a 0.85

Para la determinación de estos valores de probabilidad, vamos a recurrir a la tabla n° 1, que consta en el anexo al final del libro, se trata de la tabla de la distribución normal estándar. Ella proporciona el área bajo la curva comprendida entre la media aritmética (0) y el valor de z específico. Para encontrar el valor se debe ubicar en la columna izquierda el valor de z aproximado en unidades y décimas; para el ejemplo, bajamos por esta columna hasta el valor 0.8, se sigue por este renglón hasta ubicarnos en la intersección de la columna titulada con 0.05 (0.85 = 0.8 + 0.05), obteniéndose como lectura final: **0.3023**). Es necesario señalar que valores negativos de z no constan en la tabla, ya que al ser una función simétrica la distribución normal, los valores del área se replican a ambos lados de la media aritmética.

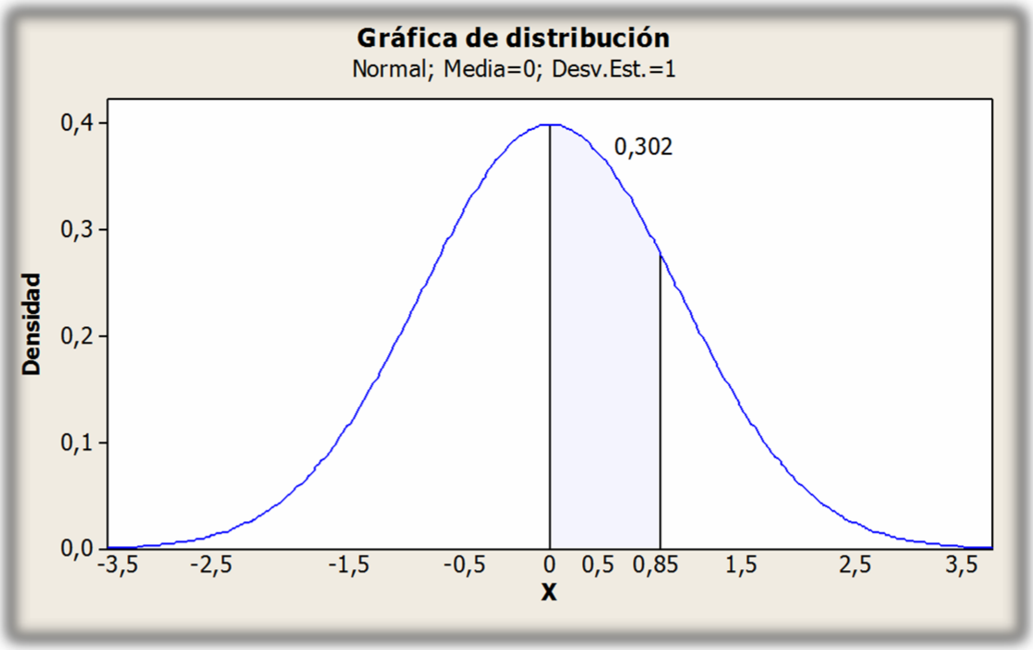
Para una mejor comprensión de los valores de probabilidad que se determinen como resultado de lo que se pide en los literales de la página anterior, vamos a realizar un análisis gráfico de los mismos, esto permite visualizar sin lugar a dudas la validez de los mismos.

De que la variable tome valores comprendidos entre la media aritmética y 0.85

El valor de probabilidad resultante de 0.302 de la gráfica proviene de la siguiente fórmula:

P(0 ≤ Z ≤ 0.85) = 0.302

Gráficamente se tiene:



Observamos en la figura, que la probabilidad de que la variable estandarizada se ubique en el intervalo comprendido entre la media (0) y 0.85 es igual al área subrayada en el gráfico, que representa 0.3023 del área total.

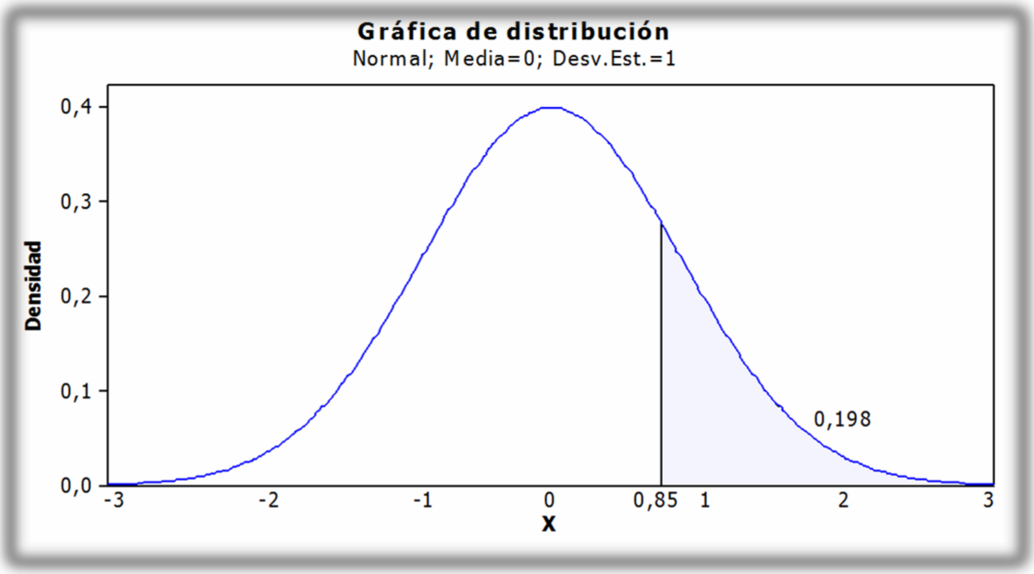
Que la variable asuma valores mayores que 0.85

El valor de probabilidad resultante de 0.198 de la gráfica proviene de la siguiente fórmula:

$$\boldsymbol{P(Z > 0.85) = 0.5 - 0.302 = 0.198}$$

Donde 0.302 es el resultante de la grafica (a)

Gráficamente se tiene:



Advertimos en la figura que antecede que la probabilidad de que la variable estandarizada sea mayor a 0.85 está representada por el área sombreada que se encuentra a la derecha del valor de referencia. Para la determinación de su valor numérico bastará con restar de 0.50 el área que la tabla establece para z = 0.85, esta operación está descrita en el gráfico.

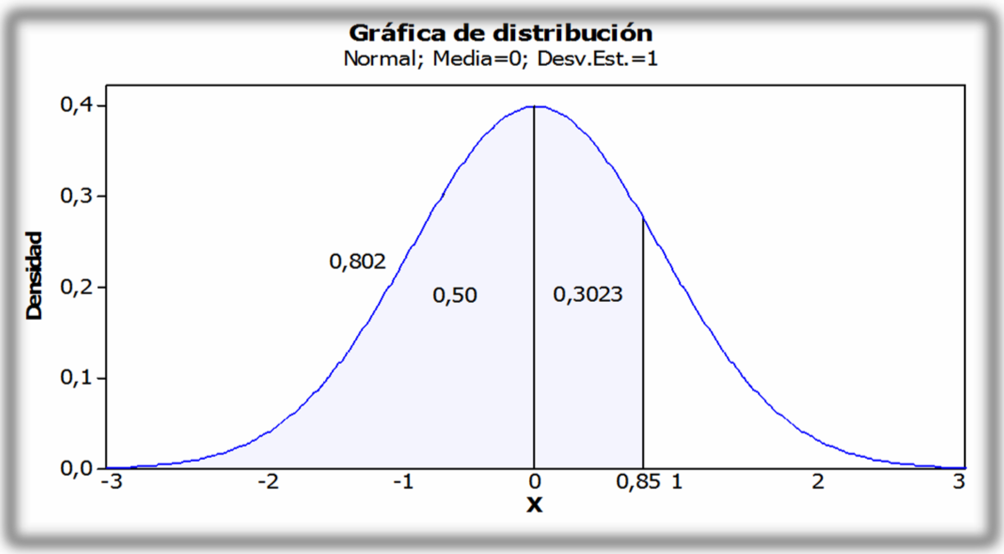
Que la variable tome valores menores o iguales a 0.85

El valor de probabilidad resultante de 0.802 de la gráfica proviene de la siguiente fórmula:

$$P(Z \leq 0.85) = 0.50 + 0.302 = 0.802$$

Donde 0.50 es la media aritmética y 0.302 es el área subyacente en el gráfico.

Gráficamente se tiene:



En este caso la probabilidad de que la variable estandarizada sea menor o igual a 0.85 está representada por el área bajo la curva que se encuentra sombreada. Numéricamente se determina sumando el área entre la media y 0.85 y el área que se encuentra a la izquierda de la media. La operación se aprecia en el gráfico con claridad.

Si se requiere determinar la probabilidad de que la variable estandarizada se encuentre entre dos valores específicos, pueden ocurrir dos posibilidades:

1. Los valores de z son ambos positivos o negativos; y

2. Los valores de z son de diferente signo.

Para el primer caso se determinará el área del valor de z que se halle más alejado de la media y se le restará a ese valor el área de z que se encuentre más cercano a la media. Para el segundo caso se procederá a sumar las dos áreas que se encuentran antes de la media y posterior a ésta. Con el siguiente ejemplo se ilustrará de mejor manera estas dos situaciones.

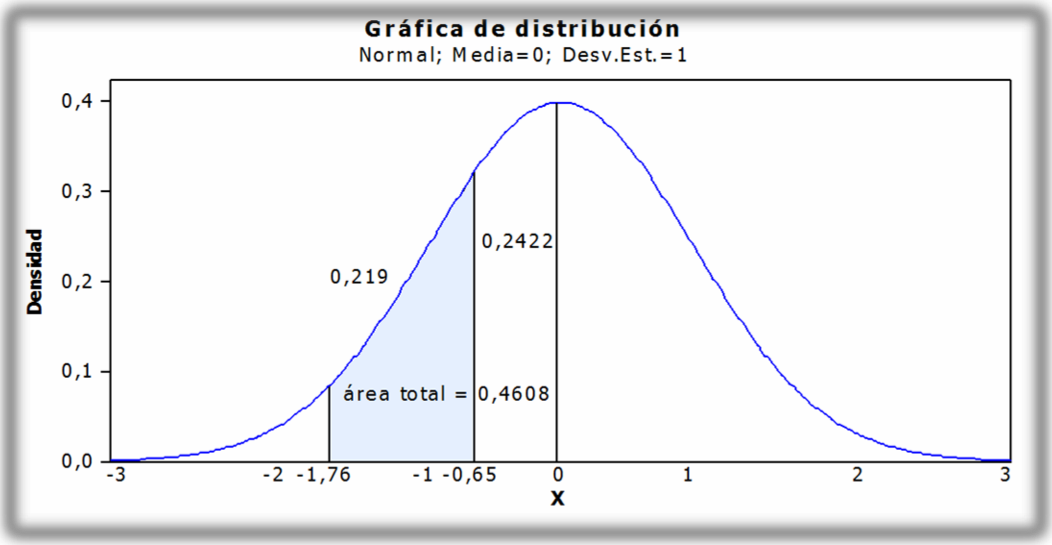
Ejemplo:

a. Determinar la probabilidad de que la variable “z” se halle entre -1.76 y -0.65.

El área para z = -1.76 es 0.4608 y el área para z = -0.65 es 0.2422. Entonces el área entre -1.76 y -0.65 se obtiene restando 0.2422 de 0.4608.

$P(-1.76 \leq z \leq -0.65) = 0.4608 - 0.2422 = 0.219$

Gráficamente se tiene:



b. ¿Cuál es la probabilidad de que la variable "z" se ubique entre -0.73 y 0.84?

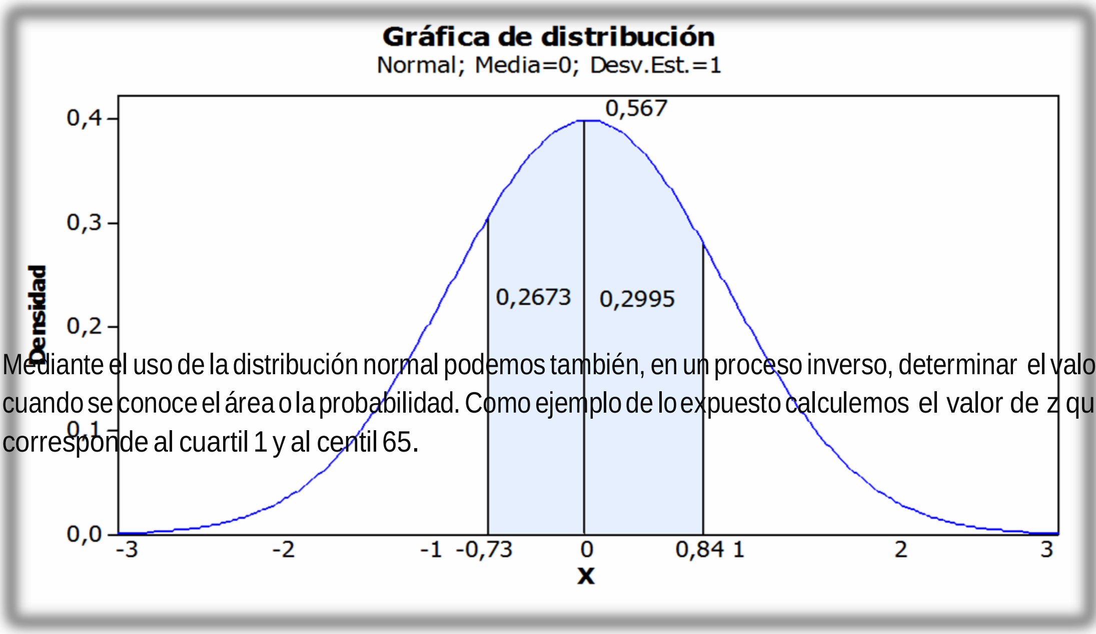
Se busca en la tabla de la distribución normal el área que se halla bajo la curva entre -0.73 y la media y luego el área respectiva entre la media y 0.84; a continuación se suman estas dos áreas.

Si z = -0.73, entonces el área es igual a 0.2673;

Si z = 0.84, entonces el área es igual a 0.2995; por lo tanto la probabilidad total es:

$P(-0.73 \leq z \leq 0.84) = 0.2673 + 0.2995 = 0.567$

Gráficamente se observa lo siguiente:



Mediante el uso de la distribución normal podemos también, en un proceso inverso, determinar el valor "z" cuando se conoce el área o la probabilidad. Como ejemplo de lo expuesto calculemos el valor de z que le corresponde al cuartil 1 y al centil 65.

Para el caso del cuartil 1, sabemos que por la posición este elemento divide a la distribución en dos partes de tal manera que el 25% es menor a él y el 75% es mayor, por lo tanto sabemos adicionalmente que entre dicho cuartil y la mediana (igual a la media) se encuentra el 25% de los datos; entonces este es el valor del área que debo ubicar en el cuerpo de la tabla y así determinar qué valor de z le corresponde. Al ubicarnos en la tabla n° 1 se observa que el valor de área más cercano a 0.25 se encuentra entre 0.2486 y 0.2518, se escoge el valor de z que contenga menos diferencia, en este caso 0.2486, que le corresponde a Z = - 0.67, el signo de este valor es negativo, por cuanto el cuartil 1 se halla antes de la media (0).

Z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
0	0.00	0.00	0.00	0.01	0.01	0.01	0.02	0.02	0.03	0.03
0	0.03	0.04	0.04	0.05	0.05	0.05	0.06	0.06	0.07	0.07
0	0.07	0.08	0.08	0.09	0.09	0.09	0.10	0.10	0.11	0.11
0	0.11	0.12	0.12	0.12	0.13	0.13	0.14	0.14	0.14	0.15
0	0.15	0.15	0.16	0.16	0.17	0.17	0.17	0.18	0.18	0.18
0	0.19	0.19	0.19	0.20	0.20	0.20	0.21	0.21	0.21	0.22
0	0.22	0.22	0.23	0.23	0.23	0.24	**0.24**	0.24	0.25	0.25
0	0.25	0.26	0.26	0.26	0.27	0.27	0.27	0.27	0.28	0.28

Aplicaciones

El uso de la distribución normal estándar se ha generalizado debido a la facilidad que brinda para la determinación de las probabilidades, que, en determinado momento necesitemos calcular en una situación específica.

A continuación vamos a realizar algunos ejercicios de aplicación de esta distribución.

Los gastos mensuales para familias de cuatro integrantes en alimentación, se hallan distribuidos normalmente, con una media de $380 y desviación estándar de $50.

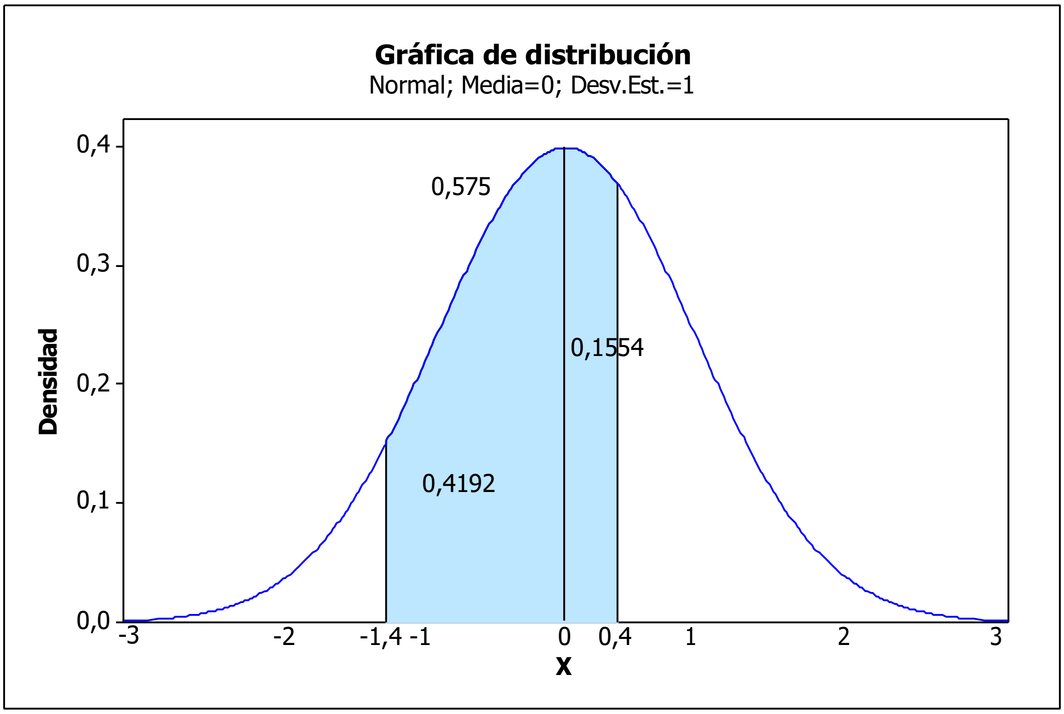
Determinar:

a. ¿Qué porcentaje de estos gastos es menor de $310?
b. ¿Cuál es la probabilidad de que los gastos sean mayores a $400?
c. ¿Qué porcentaje de los gastos está entren 320 y 420 dólares?
d. ¿Cuál es la probabilidad de que los gastos sean menores a $350 o mayores a $430?
e. Los cuartiles 1 y 3.

Para poder determinar las respuestas solicitadas, primero procedemos a estandarizar los gastos indicados, luego con la ayuda de la tabla de la distribución normal estándar encontramos las áreas correspondientes. Para una mejor comprensión del problema realizaremos también un análisis gráfico.

P(x < $310) = 0.50 – 0.4192 = 0.0808; en porcentaje *8.08%*

$\square = \frac{310-380}{50}$ = *- 1.40; entonces el área entre la media y este valor es: 0.4192*



P(x>$400) = 0.50 – 0.1554 = 0.3446

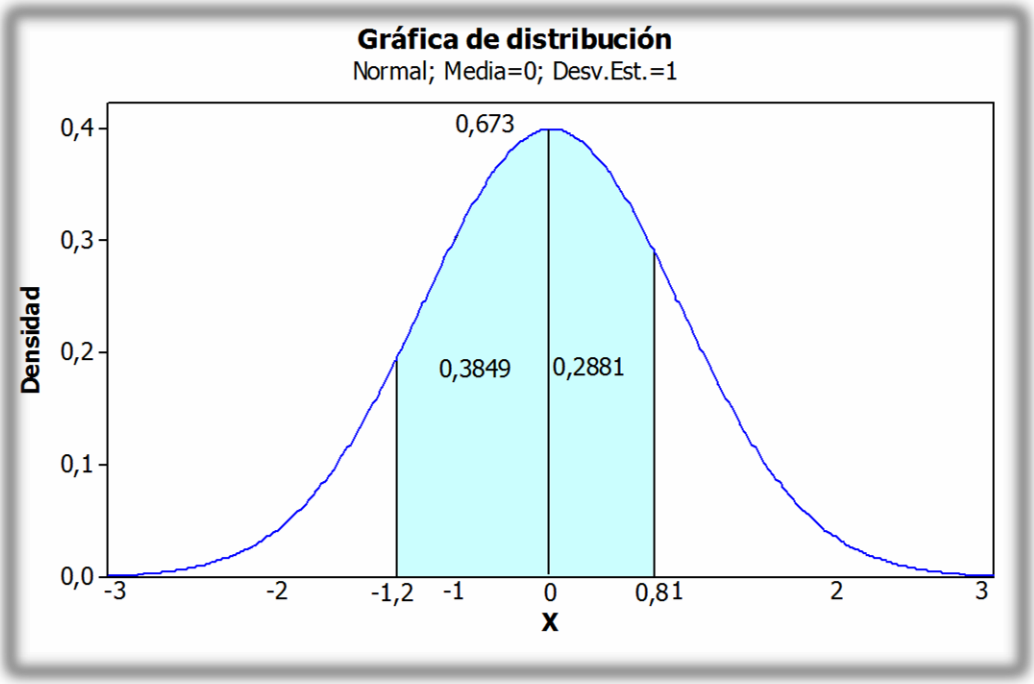
$\square = \frac{400-380}{50} = 0.40; \Rightarrow$ *Area entre la media y 0.40 es: 0.3446*

P($320 ≤ x ≤ $420) = 0.3849 + 0.2881 = 0.673, en porcentaje 67.3%

$$\square = \frac{320-380}{50} = -1.2;\ A = 0.3849$$

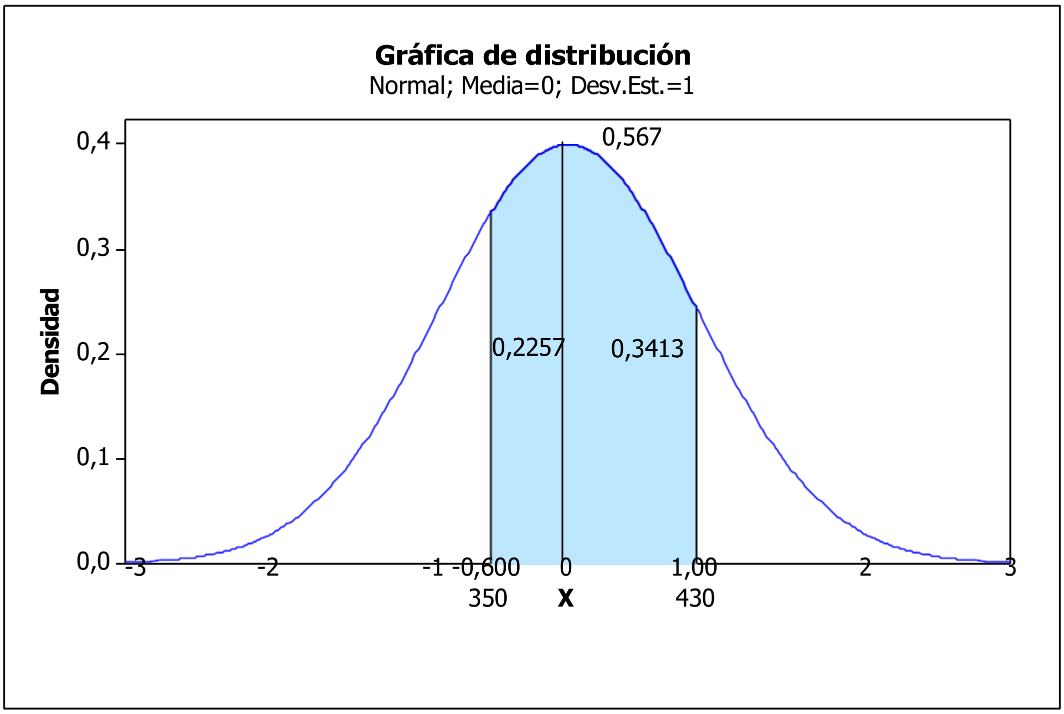
$$\square = \frac{420-380}{50} = 0.80;\ A = 0.2881$$

Entonces sumamos estas dos áreas. El resultado lo multiplicamos por 100, por cuanto nos pide el resultado en porcentaje de 67.3.



d. **P(x<350) o P(x>430) = 0.2743 + 0.1587 = 0.4330**

Se suman las dos áreas por cuanto estamos frente a la probabilidad de eventos mutuamente excluyentes de 0.567.



e. **Cuartil 1 y cuartil 3**

En este caso necesitamos conocer el valor de los gastos que dejan al 25% de los otros gastos por debajo de ese valor (Q1) y el valor de los gastos familiares que hace que el 25 % de los gastos sean mayores que él (Q3). Por lo tanto, conocemos que el área entre el Q1 y la media es 0.25, lo mismo que entre la media y el Q3, con este valor buscamos en la tabla 1 el valor de z que corresponde a estas áreas. Estos valores aproximados son: - 0.67 y 0.67. Luego en la fórmula de estandarización reemplazamos los valores conocidos y despejamos el valor de la variable correspondiente.

$$-0.67 = \frac{x-380}{50} \Rightarrow x = 380 - 0.67(50)$$

$$0.67 = \frac{x-380}{50} \Rightarrow x = 380 + 0.67(50)$$

(Q1)x = 346.5   *(Q3)x = 413.5*

Un breve análisis de estos resultados nos indica que el 50% de todos los gastos familiares de cuatro integrantes en el rubro de alimentación se hallan entre $346.5 y $413.5

Para la utilización de la distribución normal como modelo representativo de variables continuas, será necesario realizar algunas pruebas que nos permitan estar seguros que una variable continua tiene una distribución normal o se aproxima a ella. Básicamente debemos observar si se cumplen las propiedades que tiene el modelo de la distribución normal. En especial debemos tener cuidado de lo siguiente:

1 Si el conjunto de datos es pequeño, realizar diagramas de tallo y hoja, así como el diagrama de caja y bigotes. Si los datos son más numerosos realizar la distribución de frecuencias y con ella graficar el histograma o polígono de frecuencias. A fin de verificar si se cumple la condición de simetría.

2 Calcular las medidas de centralización más comunes, como media aritmética, mediana y moda y ver si son iguales o aproximadamente iguales. Además la desviación estándar para verificar si el rango intercuartil es 1.33 veces la desviación estándar. O que la amplitud de variación sea aproximadamente 6 veces la desviación estándar.

3. Verifique si en los intervalos generados en la aplicación de la regla empírica, se hallan al menos el porcentaje de datos que se manifiesta en cada caso.

Aproximación de la distribución normal a la binomial.

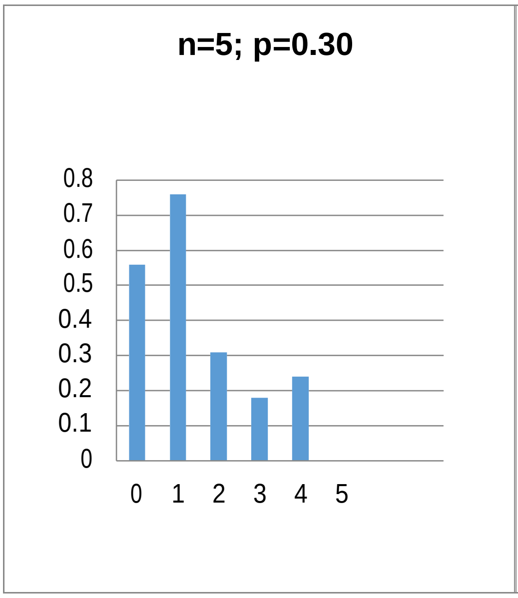
En la unidad anterior estudiamos a las distribuciones de probabilidad para variables discretas, tal el caso de la distribución binomial. En situaciones donde la muestra es muy grande, el generar la distribución de probabilidades es una tarea muy larga, aún se utilice una computadora; también el calcular probabilidades acumuladas se torna en un proceso cansino. Para tales casos es conveniente utilizar la distribución normal como una aproximación de la distribución binomial.

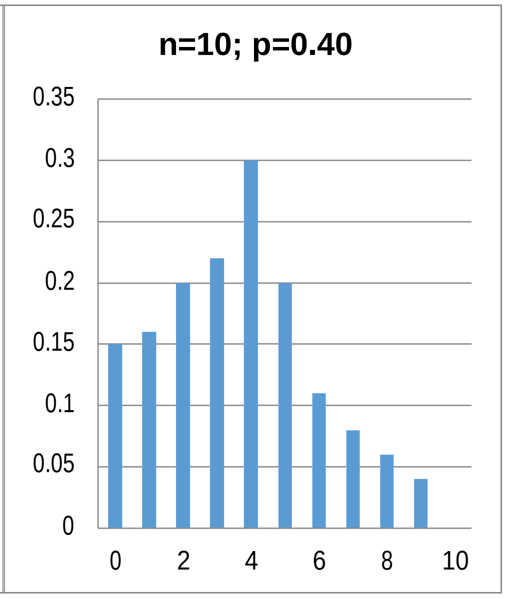
Como ejemplo suponga que un negocio recibe en promedio la visita de 500 clientes diarios y queremos saber cuál es la probabilidad de que en un día específico al menos 150 clientes realicen una compra efectiva, conociendo que la probabilidad de éxito es de 0.25. Si utilizaríamos la distribución binomial, debemos calcular la probabilidad de que el número de éxitos, es decir, que los clientes realicen una compra efectiva sea mayor o igual a 150.

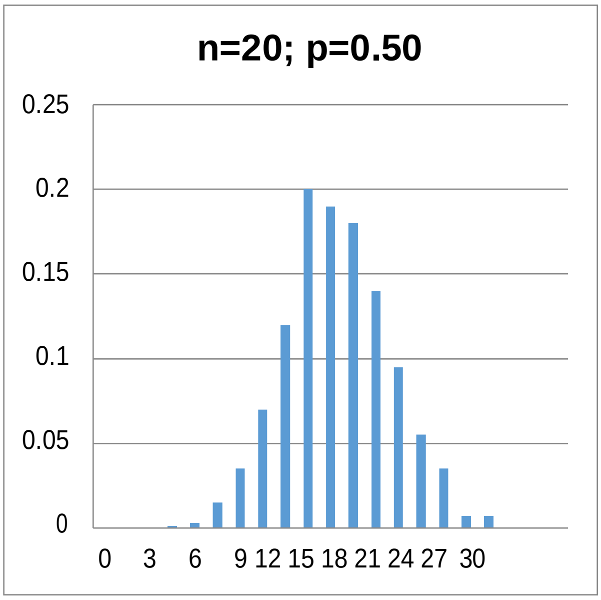
$$P(x \geq 150) = P(150) + P(151) + P(152) + \ldots\ldots + P(500)$$

Entonces observamos que debemos calcular 351 probabilidades para luego sumar estos valores para hallar la probabilidad solicitada.

En la distribución binomial, cuando el número de ensayos o la muestra es grande el gráfico de la distribución de probabilidades tiende a ser un gráfico centrado y simétrico, más aún cuando la probabilidad de éxito se acerca a 0.50; esto lo podemos apreciar en el siguiente gráfico; en el cual se ha variado el número de ensayos (n) y la probabilidad de éxito (p).







Observamos que el primer gráfico es sesgado hacia la izquierda, mientras que el segundo es menos sesgado, hasta llegar al tercero que es simétrico, esto se debe a que el tamaño es mayor y la probabilidad de éxito es 0.5

Para que la aproximación de la distribución normal sea más cercana a la distribución binomial, es necesario comprobar que los productos del número de ensayos por la probabilidad de éxito o la probabilidad de fracaso sea por lo menos 5, es decir, que:

n.p y n(1-p) ≥ 5

Además es lógico que primero deban cumplirse con los requisitos necesarios para que sea considerada la distribución como binomial. Bajo estas circunstancias podemos representar con la distribución normal a la distribución binomial. Como la distribución normal requiere de dos parámetros que son la media aritmética y la desviación estándar, debemos previamente calcular estos valores, para eso recordemos que la media y desviación estándar de la distribución binomial se determinan mediante las siguientes expresiones:

$\mu = n\,(p);\ \sigma = \sqrt{np(1-p)}$

Factor de corrección por continuidad

Como la distribución normal es para variable continua y la distribución binomial para variable discreta es necesario utilizar un factor que establezca que un valor discreto esté representado mediante un pequeño intervalo continuo, esto se consigue sumando y restando al valor entero 0.5, es decir que la variable discreta queda representada por el intervalo: ***x ± 0.50***; lo cual permite calcular el área bajo la curva entre los límites del intervalo [***x – 0.5; x + 0.5***].

Para una mejor comprensión vamos a representar el ejemplo propuesto anteriormente en el que un local comercial, recibe en promedio 500 clientes diarios, y la probabilidad de que realicen una compra efectiva es 0.25, queremos determinar la probabilidad de que al menos 200 clientes en un día en particular realicen una compra efectiva. Calculamos la media y la desviación estándar de la distribución:

$$\mu = 500(0.30) = 150; \qquad \sigma = \sqrt{500(0.30)(0.70)} = 10.25$$

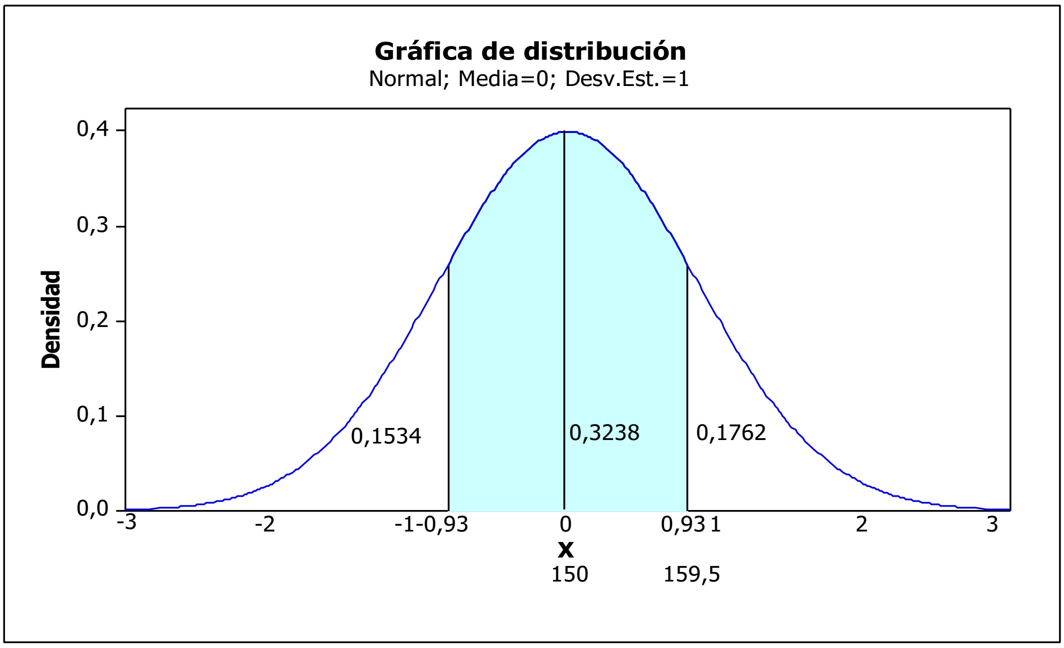
Entonces utilizando la distribución normal, se debe calcular la probabilidad de que x sea al menos 160 clientes, para la obtención del resultado procedemos a aplicar el factor de corrección por continuidad; en este caso 160 – 0.5 = 159.5. Por lo tanto para determinar la P(x ≥ 159.5), estandarizamos el valor del límite.

$$z = \frac{159.5 - 150}{10.25} = 0.93;\ \Rightarrow A = 0.3238$$

Entonces la probabilidad de que al menos 160 clientes realicen una compra efectiva es:

$P(x \geq 159.5) = 0.50 - 0.3238 = 0.1762$

Puesto en porcentaje, representa que el 17.62%, en un día cualquiera, al menos 160 clientes realizarán una compra efectiva. Gráficamente se observa lo siguiente:



Con el mismo ejemplo; si queremos conocer cuál es la probabilidad de que un día cualquiera entre 130 y 140 clientes que visitan el local comercial efectivamente efectúen una compra efectiva, tenemos que buscar el área bajo la curva entre 129.5 y 140.5 (aplicando el factor de corrección a los valores propuestos.

$\square = \frac{129.5-150}{10.25} = -2$ $\square = \frac{140.5-150}{10.25} = -0.93$

$A = 0.4772$ $A = 0.3238;$

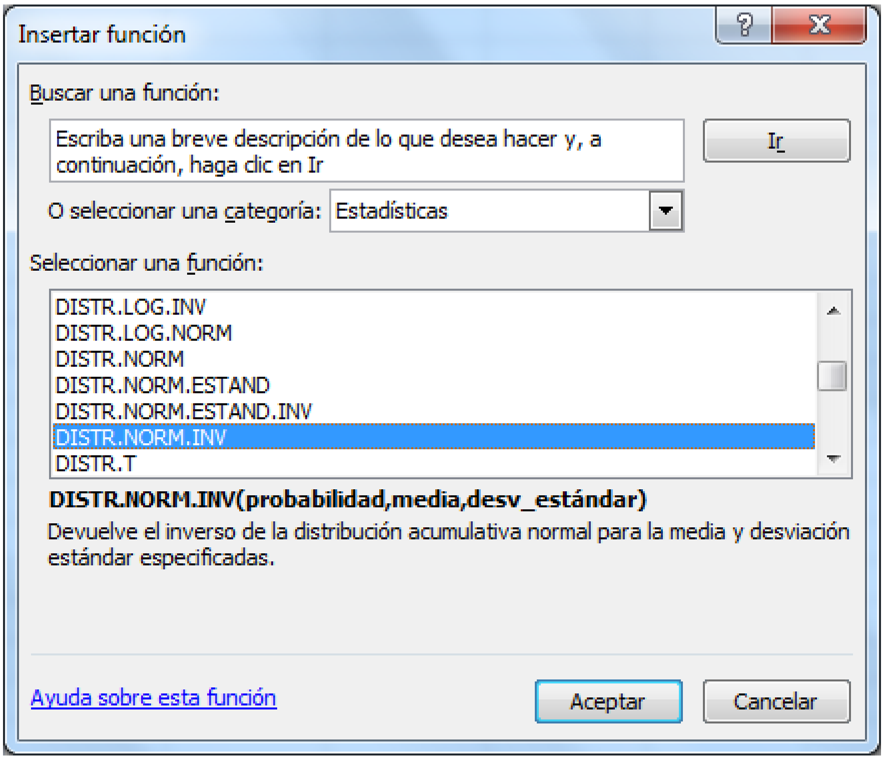
Entonces la probabilidad entre 130 y 140 es: 0.1534

Uso de Excel en la distribución normal

Si se dispone de un ordenador que tenga Excel dentro de su sistema operativo, se pueden realizar todos los cálculos que se han ejecutado en los ejemplos desarrollados anteriormente. Para esto es necesario escoger dentro de las funciones la opción **ESTADÍSTICAS** y se dispone de cuatro de ellas que hacen referencia a la distribución normal, son:

Distribución normal (DISTR.NORM), distribución normal inversa (DISTR.NORM.INV), distribución normal estándar (DISTR.NORM.ESTAND) y distribución normal estándar inversa (DISTR.NORM.ESTAND.INV). Cada una se utiliza para un objetivo específico.

En el caso de la distribución normal, es necesario conocer de antemano la media aritmética y la desviación estándar de la distribución y se utiliza para conocer la probabilidad de que la variable tome un valor específico o menor o mayor que él. Para el primer caso se deberá anotar en el siguiente orden los valores; la variable, la media, la media, la desviación estándar y como valor lógico (0), porque no necesitamos probabilidad acumulada. En la situación de necesitar la probabilidad de que la variable tome valores menores o iguales a uno en particular, se deberá anotar en el valor lógico (1), pues esta alternativa ofrece el área acumulada desde la cola izquierda de la distribución. En la impresión de la pantalla de Excel que consta a continuación se observa las opciones que se despliegan para la distribución normal.



La opción DISTRIBUCIÓN NORMAL INVERSA, se utiliza para conocer el valor que toma la variable cuando se conoce la probabilidad acumulada y la media y desviación de la distribución. Los datos se ingresan en este orden: probabilidad, media aritmética y desviación estándar.

Para el caso de la DISTRIBUCIÓN NORMAL ESTÁNDAR y su inversa se utilizan cuando se conoce z y se requiere la probabilidad; o a la inversa si queremos calcular z conocida la probabilidad acumulada, se utilizará la distribución normal estándar inversa. Para una mejor comprensión de lo expuesto, vamos a realizar algunos cálculos con estas funciones

estadísticas.

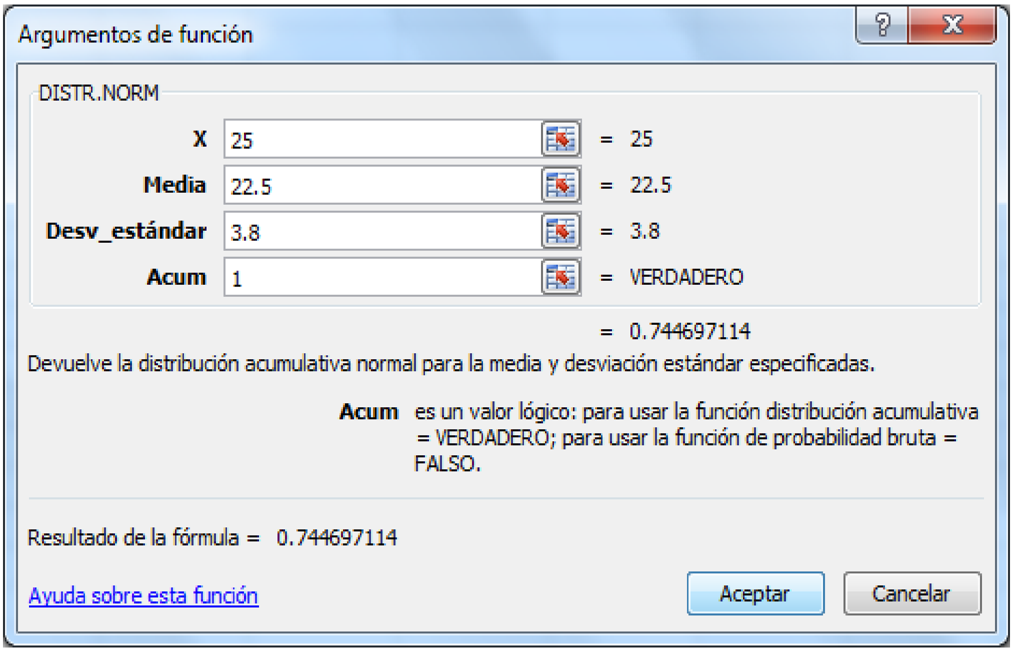
Ejemplos:

1. El peso de los embalajes de viajeros internacionales están distribuidos normalmente con una media de 22.5 kg y desviación estándar 3.8 kg. Se desea conocer:

a. La probabilidad de que un pasajero tenga una maleta cuyo peso sea como máximo 25 kg.

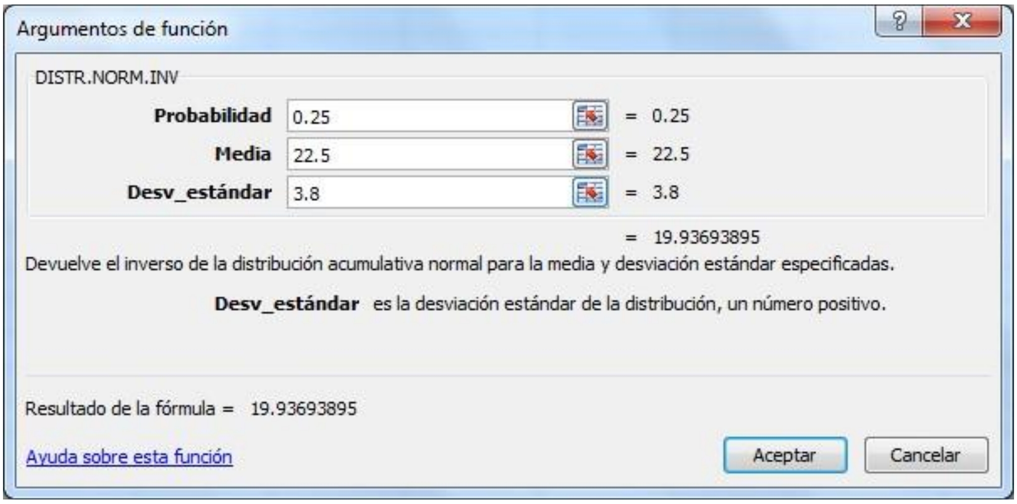
b. ¿Qué peso de los embalajes se puede considerar como el cuartil 1?

Para calcular lo solicitado en la letra a, activamos en la hoja electrónica la función DISTR.NORM y llenamos la información que nos solicita cada uno de los casilleros. Tal como aparece la pantalla siguiente:



El resultado obtenido nos señala que la probabilidad de que la maleta tenga un peso menor o igual a 25 kg es: 0.7447

Para el cálculo de lo solicitado en el siguiente literal, utilizamos la DISTR.NORM.INV, al activar esta función nos pide la probabilidad acumulada desde la cola izquierda, que para el presente caso es 0.25, luego la media (22.5) y finalmente la desviación estándar (3.8).



El resultado obtenido nos indica que aproximadamente 19.94 kg, constituye el cuartil Determinar la

probabilidad de que una variable estandarizada tome valores entre -0.85 y 0.55.

Para obtener este resultado, bastará restar del área acumulada hasta 0.55, el área acumulada hasta-0.85, entonces en la hoja de Excel activamos una celda y escribimos la siguiente fórmula:

=DISTR.NORM.ESTAND(0.55) - DISTR.NORM.ESTAND(0.85), el resultado obtenido es: 0.5112

Ejercicios propuestos

Determinar las siguientes probabilidades de una variable estandarizada.

a) $Z < 0.35$

b) $Z > -0.95$

c) $-1.16 < Z < 1.43$

d) $0.75 < Z < 1.67$

e) $Z < -1.35$ o $Z > 2.08$

f) ¿Qué valor de Z, deja al 80% de los demás valores por debajo de él?

Una población tiene sus elementos distribuidos normalmente, con una media de 18.5 y desviación estándar 3.4

a) Determine el valor de Z asociado a 19.4

b) ¿Qué porcentaje de elementos de la población están entre 16.8 y 18.4?

c) ¿Cuál es la probabilidad de seleccionar aleatoriamente a un elemento cuyo valor sea menor que 8.9?

d) ¿Cuál es el valor de la variable que representa al decil 7?

Una empresa financiera – inmobiliaria, recibe solicitudes de préstamos para la adquisición de casas, cuyos montos están aproximadamente distribuidos normalmente con una media de $55000 y desviación estándar $12500. Suponga que el día de hoy se recibió una solicitud de préstamo, cuál es la probabilidad de que:

a) La cantidad solicitada sea mayor de $70000

b) El valor solicitado se halle en el intervalo entre $50000 y 65000

c) El monto solicitado sea menor a $45000

d) Entre qué valores de las solicitudes de préstamo se halla el 50% centrado de los clientes

Un estudio en relación al grado de satisfacción que muestran los gerentes por el conocimiento de herramientas electrónicas aplicadas a la administración, demostró que el 65% de ellos utilizaban frecuentemente estas herramientas. Se realizó una encuesta a 100 gerentes. Con esta información determinar la probabilidad de que:

a) Al menos 75% de los clientes utilicen las herramientas referidas anteriormente.

b) Solo el 53% utiliza las herramientas señaladas anteriormente

"Un contratista afirma que puede renovar la cocina y un comedor de 200 pies cuadrados en 40 horas, más o menos 5 horas (media y desviación estándar respectivas). El trabajo incluye, plomería, instalaciones eléctricas, gabinetes, piso, pintura e instalación de accesorios nuevos. Suponga por experiencia que, los tiempos para terminar proyectos semejantes tienen distribución normal con media y desviación estándar como las que se indican.

a) ¿cuál es la probabilidad de que el proyecto termine en menos de 35 horas?

b) ¿cuál es la probabilidad de que el proyecto termine de 28 a 32 horas?

c) ¿cuál es la probabilidad de que el proyecto termine en 35 a 48 horas?

d) ¿cuántas horas adicionales requieren el 10% de estos proyectos?

e) Determine el eje medio para el tiempo de terminación.

f) Determine el rango intercuartil del tiempo de terminación.

g) ¿cuáles serían sus respuestas de (a) a (f) si la desviación estándar fuera horas?"

Resumen

En esta unidad se ha tratado sobre una distribución de probabilidades para variable continua, que es muy utilizada en estadística, se trata de la distribución normal y normal estándar.

1.- La distribución normal tiene las siguientes características principales:

1.1.- La media, la mediana y la moda son iguales.

1.2.- El gráfico de la función de densidad es acampanado y simétrico con relación a la media.

1.3.- Es asintótica con relación al eje horizontal, lo que significa que se aproxima a él pero nunca lo interseca.

1.4.- Es una función paramétrica que tiene dos parámetros: media aritmética y desviación estándar.

1.5.- Es una familia de distribuciones, en la cual cada variable tiene su propia media y desviación estándar.

2.- La distribución normal estándar es un tipo especial de distribución normal, que cumple con todas las características anteriores, pero que posee adicionalmente estas características.

2.1.- La media aritmética es cero (0) y la desviación estándar uno (1)

2.2.- La distribución normal estándar puede representar a cualquier distribución normal, mediante la transformación de ejes, mediante la siguiente fórmula:

$$\square = \frac{\square - \square}{\sigma}$$

2.3.- La estandarización de una variable permite medir la distancia expresada en unidades de desviación estándar desde la variable hasta la media aritmética.

3.- La distribución normal estándar puede representar a la distribución binomial, si se cumple que n(p) o n(1-p) > 5, además de que el proceso debe cumplir los requisitos necesarios para ser catalogado como distribución binomial. Se utiliza un factor de corrección por continuidad que consiste en sumar o restar al valor discreto ±0.50

4.- La operación de aplicar el factor de corrección por continuidad, sirve para compensar o darle continuidad a un valor discreto.

Autoevaluación

Complete las palabras que faltan, para que las proposiciones siguientes, sean comprensibles.

La probabilidad es un valor___________comprendido entre________________y uno, asignado a la posibilidad de ocurrencia de un evento.

Experimento es un_____________que conduce a la observación de un________________posible.

La regla especial de la adición, indica que para calcular la probabilidad de dos o más eventos__________________, se deben sumar las probabilidades individuales de cada evento.

Se denominan eventos_____________aquellos sucesos que la ocurrencia de uno de ellos_______________la probabilidad de que ocurran los otros eventos.

Indicar si las siguientes proposiciones son verdaderas o falsas.

5. Evento es un conjunto formado por uno o más resultados posibles dentro de un experimento. ( )

6. El total de formas posibles que se puede seleccionar a dos personas aleatoriamente de un grupo de ocho es 16 ( )

7. Para la determinación del número de variaciones que se pueden generar con 3 elementos de un total de 10, el orden en el que se ubican los elementos es irrelevante ( )

8.- La probabilidad de que acontezcan varios eventos que están dentro de un espacio muestral es mayor que uno. ( )

9.-Con los valores que aparecen en la siguiente tabla de contingencias, termine de completarla y calcule las probabilidades que se le pide.

		VARIABLE 1			**TOTAL**
		A	**B**	**C**	
	D	15	25	10	
	E		28	16	60
	F	8	7		20
TOTAL		39	60		130

9. a P (F) =

9. b P(A o B) =

9. c P(C y E) =

9. d P (B y F) =

9. e P(C / E) =

9. f P (F / C) =

Llenar los espacios en blanco, para que las proposiciones siguientes tengan sentido.

La distribución de probabilidades normal es representativa de variables……………………………………………………………………

La curva que representa a la distribución normal es ………………y……………………………………………………………

La distribución normal estándar tiene media aritmética igual a ……………………………………….. y desviación estándar igual a ……………………………………………….......................................

La distribución normal se aproxima a la distribución binomial cuando……… y.. es mayor que ....

Determinar si las siguientes proposiciones son verdaderas (V) o falsas (F)

5. El área bajo la curva de la distribución normal estándar que se halla a la derecha de – 1, es – 0.1587. ( )

6. El eje medio de la distribución normal es igual a la media aritmética de la distribución. ( )

7. El factor de corrección por continuidad que se aplica cuando se utiliza la distribución estándar como una aproximación a la distribución binomial, no se lo utiliza si los cálculos son realizados en la hoja electrónica de Excel. ( )

8. Los cuartiles 1 y 3 de una distribución normal se hallan a una distancia de la media aritmética que no es la misma para los dos. ( )

Complete las palabras que faltan, para que las proposiciones siguientes, sean comprensibles.

La probabilidad es un valor___________comprendido entre________________y uno, asignadoa la posibilidad de ocurrencia de un evento.

Experimento es un_____________que conduce a la observación deun_______________posible.

La regla especial de la adición, indica que para calcular la probabilidad de dos o más eventos_ _________________, se deben sumar las probabilidades individuales de cada evento.

Se denominan eventos_____________aquellos sucesos que la ocurrencia de uno de ellos_ _______________la probabilidad de que ocurran los otros eventos.

Indicar si las siguientes proposiciones son verdaderas o falsas.

Evento es un conjunto formado por uno o más resultados posibles dentro de un experimento. ( )

El total de formas posibles que se puede seleccionar a dos personas aleatoriamente de un grupo de ocho es 16 ( )

Para la determinación del número de variaciones que se pueden generar con 3 elementos de un total de 10, el orden en el que se ubican los elementos es irrelevante ( )

La probabilidad de que acontezcan varios eventos que están dentro de un espacio muestral es mayor que uno. ( )

RESPUESTAS A LOS EJERCICIOS PROPUESTOS.

1. a) 0.6368
   b) 0.8289
   c) 0.8006
   d) 0.1792

e) 0.1073

b) 17.97%

c) 0.0024

d) 20.28

b) 0.4436

c) 0.2119

b) 0.0036

5. a) 0.1587

b) 0.0466

c) 0.7865

d) 46.4 horas

e) 40.0 horas

f) 6.7 horas

g) si la desviación estándar es 10 horas a)

0.3085

b) 0.0968

c) 0.4796

d) 52.8 horas

e) 40.0 horas

f) 13.4 horas

Bibliografía

NewboldPaul;CarlsonWillaL;ThorneBetty(2008)EstadísticaparaAdministraciónySexta Edición, Prentice Hall. (España)

Levin,R;Rubin,D(2010)EstadísticaparaIaAdministraciónyEconomía.Séptimaedición. Prentice Hall(Mexivo).

Mason; Lind; Marchal (2013). Estadística para administración y economía. Decima Edición, Alfaomega. (Colombia).

AndersonD;SweeneyD;WiliamsT.(2008).EstadisticaparaIaAdministr4acionyeconomía. Octava primera EDICION, McGraw Hill Interamericana. (México)