ANTOLOGIA

ESTADISTICA

LICENCIATURA EN PSICOLOGÍA

PRIMER CUATRIMESTRE

Marco Estratégico de Referencia

ANTECEDENTES HISTORICOS

Nuestra Universidad tiene sus antecedentes de formación en el año de 1979 con el inicio de actividades de la normal de educadoras "Edgar Robledo Santiago", que en su momento marcó un nuevo rumbo para la educación de Comitán y del estado de Chiapas. Nuestra escuela fue fundada por el Profesor de Primaria Manuel Albores Salazar con la idea de traer Educación a Comitán, ya que esto representaba una forma de apoyar a muchas familias de la región para que siguieran estudiando.

En el año 1984 inicia actividades el CBTiS Moctezuma Ilhuicamina, que fue el primer bachillerato tecnológico particular del estado de Chiapas, manteniendo con esto la visión en grande de traer Educación a nuestro municipio, esta institución fue creada para que la gente que trabajaba por la mañana tuviera la opción de estudiar por las tarde.

La Maestra Martha Ruth Alcázar Mellanes es la madre de los tres integrantes de la familia Albores Alcázar que se fueron integrando poco a poco a la escuela formada por su padre, el Profesor Manuel Albores Salazar; Víctor Manuel Albores Alcázar en septiembre de 1996 como chofer de transporte escolar, Karla Fabiola Albores Alcázar se integró como Profesora en 1998, Martha Patricia Albores Alcázar en el departamento de finanzas en 1999.

En el año 2002, Víctor Manuel Albores Alcázar formó el Grupo Educativo Albores Alcázar S.C. para darle un nuevo rumbo y sentido empresarial al negocio familiar y en el año 2004 funda la Universidad Del Sureste.

La formación de nuestra Universidad se da principalmente porque en Comitán y en toda la región no existía una verdadera oferta Educativa, por lo que se veía urgente la creación de una institución de Educación superior, pero que estuviera a la altura de las exigencias de los jóvenes que tenían intención de seguir estudiando o de los profesionistas para seguir preparándose a través de estudios de posgrado.

Nuestra Universidad inició sus actividades el 18 de agosto del 2004 en las instalaciones de la 4ª avenida oriente sur no. 24, con la licenciatura en Puericultura, contando con dos grupos de cuarenta alumnos cada uno. En el año 2005 nos trasladamos a nuestras propias instalaciones en la carretera Comitán – Tzimol km. 57 donde actualmente se encuentra el campus Comitán y el Corporativo UDS, este último, es el encargado de estandarizar y controlar todos los procesos operativos y Educativos de los diferentes Campus, Sedes y Centros de Enlace Educativo, así como de crear los diferentes planes estratégicos de expansión de la marca a nivel nacional e internacional.

MISIÓN

Satisfacer la necesidad de Educación que promueva el espíritu emprendedor, aplicando altos estándares de calidad Académica, que propicien el desarrollo de nuestros alumnos, Profesores, colaboradores y la sociedad, a través de la incorporación de tecnologías en el proceso de enseñanza-aprendizaje.

VISIÓN

Ser la mejor oferta académica en cada región de influencia, y a través de nuestra Plataforma Virtual tener una cobertura Global, con un crecimiento sostenible y las ofertas académicas innovadoras con pertinencia para la sociedad.

VALORES

- Disciplina
- Honestidad
- Equidad
- Libertad

ESCUDO

El escudo de la UDS, está constituido por tres líneas curvas que nacen de izquierda a derecha formando los escalones al éxito. En la parte superior está situado un cuadro motivo de la abstracción de la forma de un libro abierto.

ESLOGAN

“Mi Universidad”

ALBORES

Es nuestra mascota, un Jaguar. Su piel es negra y se distingue por ser líder, trabaja en equipo y obtiene lo que desea. El ímpetu, extremo valor y fortaleza son los rasgos que distinguen.

ESTADISTICA

Objetivo de la materia:

Conocer, comprender, analizar e interpretar la estadística y sus aplicaciones en el área de la psicologia

INDICE

PROBABILIDAD Y TEORÍA DE CONJUNTOS

3.1.- Aspectos generales de la probabilidad, (conceptos, tipos de probabilidad, enfoques de probabilidad).

3.2.- Leyes de la probabilidad.

3.3.- Aplicaciones de la probabilidad en psicología.

3.4.- Árboles de probabilidad.

3.5.- Teoremas de Bayes.

3.6.- Teoría de Conjuntos; operaciones aplicadas en la psicología.

UNIDAD IV

DISTRIBUCIONES DE PROBABILIDAD APLICADAS EN LA PSICOLOGÍA

4.1.- Distribución para variables discreta

4.2.- Distribución para variables continúas

4.3-. Muestreo y estimación aplicada al control estadístico de procesos

4.4.-Fundamentos teóricos del muestreo y estimación.

4.4.1.- Distribución de muestreo; características y aplicación en psicología.

4.4.2.- Teorema del límite central.

4.4.3.- Tipos de estimación y características.

4.4.4.- Determinación del tamaño de la muestra.

4.4.5. -Intervalos de confianza aplicados al control estadístico de procesos.

4.4.6.- Gráficas de control y tipos de variación en los procesos.

Unidad I

1.1.- ANTECEDENTES HISTÓRICOS DE LA ESTADÍSTICA

Historia de la estadística

La palabra "estadística" a menudo nos trae a la mente imágenes de números apilados en grandes arreglos y tablas, de volúmenes de cifras relativas a nacimientos, muertes, impuestos, poblaciones, ingresos, deudas, créditos y demás. Al instante de escuchar esa palabra, son estas las imágenes que llegan a nuestra imaginación.

La estadística es mucho más que sólo números apilados y gráficas bonitas. Es una ciencia con tanta antigüedad como la escritura, y es por sí misma auxiliar de todas las ciencias –medicina, ingeniería, sociología, psicología, economía, etcétera–, así como de los gobiernos, mercados y otras actividades humanas.

En la actualidad, la estadística ocupa un lugar de gran importancia en la investigación y en la práctica médica. En los estudios de medicina de cualquier país se incluyen varias asignaturas dedicadas a la estadística; es difícil, por no decir imposible, que un trabajo de investigación sea aceptado por una revista médica sin que sus autores hayan utilizado técnicas y conceptos estadísticos en su planteamiento y en el análisis de los datos.

La estadística que conocemos hoy día debe gran parte de sus logros a los trabajos matemáticos de aquellos hombres que desarrollaron la teoría de las probabilidades, con la cual se adhirió la estadística a las ciencias formales.

Desde los comienzos de la civilización han existido formas sencillas de estadísticas, 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 y otras cosas.

Hacia el año 3000 a. de C. los babilonios utilizaban ya pequeñas tablillas de arcilla para recopilar datos sobre la producción agrícola y los géneros vendidos o cambiados mediante trueque. En el antiguo Egipto, los faraones lograron recopilar, alrededor del año 3050 a. de C., prolijos datos relativos a la población y la riqueza del país; de acuerdo con el historiador griego Heródoto, dicho registro de la riqueza y la población se hizo con el propósito de preparar la construcción de las pirámides. En el mismo Egipto, Ramsés II hizo un censo de las tierras con el objeto de verificar un nuevo reparto.

En el antiguo Israel, la Biblia da referencia, en el libro de los Números, de los datos estadísticos obtenidos en dos recuentos de la población hebrea. El rey David, por otra parte, ordenó a Joab, general del ejército, hacer un censo de Israel con la finalidad de conocer el número de habitantes, y el libro Crónicas describe el bienestar material de las diversas tribus judías.

En China ya había registros numéricos similares con anterioridad al año 2000 a. de C. Los griegos, hacia el año 594 a. de C., efectuaron censos periódicamente con fines tributarios, sociales (división de tierras) y militares (cálculo de recursos y hombres disponibles). La investigación histórica revela que se realizaron 69 censos para calcular los impuestos, determinar los derechos de voto y ponderar la potencia guerrera.

Pero fueron los romanos, maestros de la organización política, quienes mejor supieron emplear los recursos de la estadística. Cada cinco años llevaban a cabo un censo de la térpoblación, y los funcionarios públicos tenían la obligación de anotar nacimientos, defunciones y matrimonios, sin olvidar los recuentos periódicos del ganado y de las riquezas contenidas en las tierras conquistadas. En la época del nacimiento de Cristo sucedía uno de estos empadronamientos de la población bajo la autoridad del Imperio.

Durante los mil años posteriores a la caída del Imperio Romano se hicieron muy pocas operaciones estadísticas, con la notable excepción de las relaciones de tierras pertenecientes a la Iglesia, compiladas por Pipino el Breve y por Carlomagno en los años

758 y 762, respectivamente. En Francia se realizaron algunos censos parciales de siervos durante el siglo IX.

Después de la conquista normanda de Inglaterra en 1066, el rey Guillermo I encargó un censo en el año 1086. La información en él obtenida se recoge en el Domesday Book, o Libro del Gran Catastro, que es un documento acerca de la propiedad, la extensión y el valor de las tierras en Inglaterra. Esta obra fue el primer compendio estadístico de ese país.

Aunque Carlos magno en Francia y Guillermo el Conquistador en Inglaterra trataron de revivir la técnica romana, los métodos estadísticos permanecieron casi olvidados durante la Edad Media. Durante los siglos XV, XVI y XVII, hombres como Leonardo de Vinci, Nicolás Copérnico, Galileo Galilei, William Harvey, Francis Bacon y René Descartes hicieron grandes operaciones con base en el método científico, de tal forma que cuando se crearon los Estados nacionales y surgió como fuerza el comercio internacional, había ya un método capaz de aplicarse a los datos económicos. Debido al temor que Enrique VII tenía de la peste, en el año 1532 empezaron a registrarse en Inglaterra las defunciones causadas por esta enfermedad. En Francia, más o menos por la misma época, la ley exigía a los clérigos registrar los bautismos, fallecimientos y matrimonios.

Durante un brote de peste que apareció a fines del siglo XVI, el gobierno inglés comenzó a publicar estadísticas semanales de los decesos. Esa costumbre continuó muchos años, y en 1632 los llamados Bills of Mortality (Cuentas de Mortalidad) ya contenían datos sobre los nacimientos y fallecimientos por sexo. En 1662, el capitán John Graunt compiló documentos que abarcaban treinta años, mediante los cuales efectuó predicciones sobre el número de personas que morirían de diversas enfermedades, así como de las proporciones de nacimientos de hombres y mujeres que cabía esperar. El trabajo de

Graunt, condensado en su obra Natural and polit ical observations… made upon the Bi lls of Mortality (Observaciones políticas y naturales…hechas a partir de las Cuentas de Mortalidad), fue un esfuerzo de inferencia y teoría estadística.

Alrededor del año 1540, el alemán Sebastián Muster realizó una compilación estadística de los recursos nacionales, que comprendía datos acerca de la organización política, instrucciones sociales, comercio y poderío militar. Durante el siglo XVII se aportaron indicaciones más concretas sobre los métodos de observación y análisis cuantitativo y se ampliaron los campos de la inferencia y la teoría estadística.

Los eruditos del siglo XVII demostraron especial interés por la estadística demográfica como resultado de la especulación sobre si la población aumentaba, disminuía o permanecía estática.

En los tiempos modernos, tales métodos fueron resucitados por algunos reyes que necesitaban conocer las riquezas monetarias y el potencial humano de sus respectivos países. El primer empleo de los datos estadísticos para fines ajenos a la política tuvo lugar en 1691 y estuvo a cargo de Gaspar Neumann, un profesor alemán que vivía en Breslau. Este investigador se propuso destruir la antigua creencia popular de que en los años terminados en 7 moría más gente que en los restantes, y para lograrlo hurgó pacientemente en los archivos parroquiales de la ciudad. Después de revisar miles de partidas de defunción, pudo demostrar que en tales años no fallecían más personas que en los demás. Los procedimientos de Neumann fueron conocidos por el astrónomo inglés Halley, descubridor del cometa que lleva su nombre, quien los aplicó al estudio de la vida humana. Sus cálculos sirvieron de base para las tablas de mortalidad que hoy utilizan todas las compañías de seguros.

Godofredo Achenwall, profesor de la Universidad de Gotinga, acuñó en 1760 la palabra estadística, que extrajo del término italiano statista (estadista). Creía, y con sobrada razón, que los datos de la nueva ciencia serían el aliado más eficaz del gobernante consciente. La raíz remota de la palabra se halla en el término latino s t a t u s, que significa “estado” o“situación”. Esta etimología aumenta el valor intrínseco de la palabra por cuanto que la estadística revela el sentido cuantitativo de las más variadas situaciones.

Uno de los primeros trabajos sobre las probabilidades corresponde al matemático italiano del siglo XVI Girolano Cardano, aunque fue publicado 86 años después de su fallecimiento. En el siglo XVII encontramos correspondencia relativa a la probabilidad en los juegos de azar entre los matemáticos franceses Blaise Pascal y Pierre de Fermat, fundamentos sobre los que Chri stian Huygens, físico, matemát ico y astrónomo danés, publicaría un libro en 1656. Durante ese mismo siglo y principios del XVIII, matemáticos como Bernoulli, Maseres, Lagrange y Laplace desarrollaron la teoría de probabilidades. No obstante, durante cierto tiempo la teoría de las probabilidades limitó su aplicación a los juegos de azar, y no fue sino hasta el siglo siguiente que comenzó a aplicarse a los grandes problemas científicos.

Durante el siglo XVIII empieza el auge de la estadística descriptiva en asuntos sociales y económicos, y es a finales de ese siglo y comienzos del XIX cuando se comienzan a asentar verdaderamente las bases teóricas de la teoría de probabilidades con los trabajos de Joseph Louis Lagrange y Pierre Simon de Laplace, del brillantísimo y ubicuo matemático y astrónomo alemán Carl Friedrich Gauss, y de Simeón-Denis Poisson. Previamente, cabe destacar el descubrimiento de la distribución normal por Abraham de Moivre, distribución que será posteriormente “redescubierta” por Gauss y Poisson.

Jacques Quételect es quien aplica la estadística a las ciencias sociales. Interpretó la teoría de la probabilidad para su uso en esas ciencias y aplicó el principio de promedios y de la variabi l idad a los fenómenos sociales. Quételect fue el primero en efectuar la aplicación práctica de todo el método estadístico entonces conocido a las diversas ramas de la ciencia.

En el periodo de 1800 a 1820 se desarrollaron dos conceptos matemáticos fundamentales para la teoría estadística: la teoría de los errores de observación, aportada por Laplace y Gauss, y la teoría de los mínimos cuadrados, realizada por Laplace, Gauss y Legendre. A finales del siglo XIX, Sir Francis Galton ideó el método conocido como c o r r e l a c i ó n, que tenía por objeto medir la influencia relativa de los factores sobre las variables. De aquí partió el desarrollo del coeficiente de correlación creado por Karl Pearson y otros cultivadores de la ciencia biométrica, tales como J. Pease Norton, R. H. Hooker y G.
Udny Yule, que efectuaron amplios estudios sobre la medida de las relaciones.

Una vez sentadas las bases de la teoría de probabilidades, podemos situar el nacimiento de la estadística moderna y su empleo en el análisis de experimentos en los trabajos de

Francis Galton y Kurt Pearson. Este último publicó en 1892 el libro The Grammar of Science (La gramática de la ciencia), un clásico en la filosofía de la ciencia, y fue él quien ideó el conocido test de Chi -cuadrado. El hijo de Pearson, Egon, y el matemát ico nacido en Polonia Jerzy Neyman pueden considerarse los fundadores de las pruebas modernas de contraste de hipótesis.

Pero es sin lugar a dudas Ronald Arnold Fisher la figura más influyente de la estadística, pues la situó como una poderosa herramienta para la planeación y análisis de experimentos. Contemporáneo de Pearson, desarrolló el análisis de varianza y fue pionero en el desarrollo de numerosas técnicas de análisis multivariante y en la introducción del método de máxima verosimilitud para la estimación de parámetros. Su libro Statistical Methods for Research Workers (Métodos estadísticos para los investigadores), publicado en 1925, ha sido probablemente el libro de estadística más utilizado a lo largo de muchos años.

Mientras tanto, en Rusia, una activa y fructífera escuela de matemáticas y estadística aportó asimismo –como no podía ser de otro modo– su considerable influencia. Desde finales del siglo XVIII y comienzos del XIX cabe destacar las figuras de Pafnuty Chebichev y Andrei Harkov, y posteriormente las de Alexander Khinchin y Andrey Kolmogorov.

En el siglo XIX, con la generalización del método científico para estudiar todos los fenómenos de las ciencias naturales y sociales, los investigadores vieron la necesidad de reducir la información a valores numéricos para evitar la ambigüedad de las descripciones verbales.

En nuestros días, la estadística se ha convertido en un método efectivo para describir con exactitud los valores de los datos económicos, políticos, sociales, psicológicos, biológicos y físicos, y sirve como herramienta para relacionar y analizar dichos datos. El trabajo del experto estadístico no consiste ya sólo en reunir y tabular los datos, sino sobre todo en interpretar esa información.

El desarrollo de la teoría de la probabilidad ha aumentado el alcance de las aplicaciones de la estadística. Muchos conjuntos de datos se pueden estudiar con gran exactitud utilizando determinadas distribuciones probabilísticas. La probabilidad es útil para comprobar la fiabilidad de las inferencias estadísticas y para predecir el tipo y la cantidad de datos necesarios en un determinado estudio estadístico.

1.2 SUCESOS DE INTERÉS EN EL DESARROLLO DE LA ESTADÍSTICA

A continuación se presenta una relación cronológica de diferentes sucesos que nos permiten tener una idea general de la evolución de la estadística.

Dos hechos contradictorios en la historia de la estadística

La estadística y el nazismo

Tal y como quedó dicho, R. A. Fisher constituye una figura capital en el desarrollo de la estadística moderna, y se puede incluso decir que es quizás la más importante e influyente; sin embargo, también existen zonas de sombra en su importante trabajo. A raíz de los descubrimientos de Charles Darwin sobre el mecanismo hereditario de evolución de las especies, surgió una nueva teoría científica (?) denominada eugenesia, término acuñado por Francis Galton en 1883, quien era por cierto sobrino de Darwin y "descubridor" de las huellas digitales.

Podríamos definir la eugenesia como la ciencia que estudia cómo mejorar la raza humana, proporcionando los mecanismos para que las características que se consideran como mejores se desarrollen más rápidamente que las inadecuadas. Se trata por tanto de dirigir de forma controlada la selección natural. En cuanto escuchamos esta definición, enseguida nos viene a la mente el nazismo y sus teorías de superioridad de la raza aria, limpieza étnica y demás.

Desgraciadamente, no sólo muchos matemáticos sino también un gran número de científicos de otras especialidades fueron defensores de las teorías eugenésicas. La lista de los científicos que, al menos inicialmente, prestaron su apoyo a dicha teoría es lamentablemente muy grande. Entre los estadísticos hallamos a Galton, a Pearson y sobre todo a Fisher.

En 1933, el gobierno alemán, presidido por Hitler, promulgó la ley de esterilización eugenésica, que puede considerar se ya como el antecedente de los exterminios perpetrados en los campos de concentración y de las atrocidades cometidas en nombre de una supuesta experimentación médica en dichos campos.

Aunque en 1930 Huxley, Haldane, Hogben, Jennings y otros biólogos renombrados comenzaron a reaccionar en contra de lo descabellado de muchas ideas propugnadas por la eugenesia, ya era demasiado tarde puesto que dichas ideas habían logrado difusión e importancia, y no sólo en los regímenes fascistas europeos: un importante biólogo americano, Charles Davenport, financiado por la Carnegie Foundation, creó el Eugenics Record Office en 1910, y miles de americanos llenaron un "registro de rasgos familiares", que era una especie de pedi-greefamiliar. Y fumar, ¿produce cáncer?

Hacia 1920 se observó un gran incremento de los fallecimientos debidos al cáncer pulmonar. Aunque había trabajos previos sobre la posible relación entre el hábito de fumar y el cáncer de pulmón, como los de Lombard y Doering (1928) y Müller (1939), no será sino hasta la década de los cincuenta –con los trabajos de Wynder y Graham (1950) y sobre todo de Doll y Hill (1952 y 1959) – que la cuestión cobrará verdadero interés e incluso propiciará agrios debates en la opinión pública. Este último trabajo, publicado en el British Medical Journal, es un estudio de casos controles, donde los casos eran los pacientes que habían ingresado en ciertos hospitales con diagnóstico de cáncer de pulmón, mientras que los controles eran pacientes cuyo ingreso se debía a otras causas. A ambos tipos de pacientes se le interrogaba sobre sus hábitos de fumar tabaco, de inhalar otros gases y otros posibles agentes etiológicos. Las encuestas fueron efectuadas por personal "ciego", en el sentido de que desconocía el propósito del trabajo. El resultado fue que los casos y los controles tenían una exposición similar a todos los posibles factores de riesgo, salvo el tabaco, con los siguientes resultados:

Si efectuamos los cálculos, el odds ratio es de 9.1, y dado que las tasas de cáncer de pulmón en la población son bajas, puede interpretarse como un riesgo relativo de padecer cáncer de pulmón de los fumadores f rente a los no fumadores. El resultado es estadísticamente significativo, con un nivel de confianza inferior a 0.001.

En 1954, Doll y Hill comenzaron un estudio prospectivo, de cohortes, en el que se efectuaba un seguimiento de médicos británicos y se estudiaba la posible asociación entre las tasas de mortalidad y el hábito de fumar tabaco, que corroboró no sólo los resultados anteriores sino también una mortalidad más rápida debida también a ot ras causas – fundamentalmente enfermedades coronarias– entre los fumadores.

A medida que la evidencia se fue acumulando, tanto Berkson como Neyman fueron cambiando de opinión, aunque Fisher permaneció irreductible en su posición. Otro gran estadístico, Jerome Cornfield, y cinco expertos más del Nacional Cancer Institute, de la American Cancer Sociaty y del Sloan-Kettering Institute, escribieron un artículo en 1959 en el que se revisaban los diferentes trabajos publicados al respecto, así como las objeciones que habían sido planteadas tanto por Fisher como por Berkson y Neyman y el propio Tobacco Institute, demostrando la abrumadora evidencia a favor de la tesis de que el hábito de fumar es una causa importante del aumento en la incidencia de cáncer de pulmón.

Bosquejo Histórico:

La estadística fue fundada por el londinense John Graunt, "un mercader de mercería", en un pequeño libro "Natural and political Observations made upon the Bells of Mortality". Este libro fue el primer intento para interpretar fenómenos biológicos de masa y de la conducta social: a partir de datos numéricos escribir las cifras brutas de nacimientos y defunciones en Londres, de 1604 a 1661. El opúrculo de Graunt apareció en 1662. Treinta años más tarde, la Royal Society publicó en su "Philosophical Transactions" un artículo sobre tasas de mortalidad escrito por el eminente astrónomo Edmund Halley.

Ambas publicaciones constituyen la base de todo trabajo posterior sobre esperanza de vida, indispensable para la solvencia de las compañías de seguros de vida.

John Graunt nació en 1620 en Berchin Lane, Londres, bajo el signo de las siete estrellas, donde su padre tenía una tienda y el hogar. Aprendió pronto el oficio de vendedor de mercería y prosperó en el negocio. El éxito le dio la posibilidad de dedicarse a ocupaciones más amplias que las de la venta de artículos de mercería. Aubrey lo describe como "una persona muy ingeniosa y estudiosa... se levantaba muy temprano para sus estudios antes de abrir la tienda". Se hizo amigo de Sir William Petty, más tarde autor de un libro sobre la nueva ciencia de la aritmética política, y probablemente discutió con él las ideas expresadas en sus "Obervations".

Las tablas de mortalidad, que atrajeron la atención de Graunt, eran publicadas semanalmente por la compañía de Sacristanes parroquiales y contenían el número de muertes acaecidas en cada parroquia, sus causas y también un "Recuento de todos los entierros y bautizos habidos en la semana" en las cuales anotaban el número de nacimientos de acuerdo a los que acudían al bautismo y lo mismo sucedía cuando presentaban sus defunciones (en las parroquias se llevaba el control).

Un ejemplo de las observaciones hechas por Graunt en 1632 fueron las siguientes:

Varones 4,994

Bautizados Hembras 4,590

T o t a l 9,584

Varones 4,932

Enterrados Hembras 4,603

T o t a l 9,535

Con estos datos deducía las siguientes observaciones:

a) Hay más varones que hembras
b) Pocos murieron de hambre
c) Hay pocos asesinatos
d) Los lunáticos son pocos

Las "Observations" impresionaron tan favorablemente a Carlos II, que este propuso especialmente a Graunt como socio fundador de la recientemente constituida Royal Society. Para prevenir cualquier posible objeción al hecho de que Graunt era tendero, "su majestad dio este encargo particular a su Sociedad, de que si encontraban algún comerciante más de su estilo, lo admitiesen sin más ceremonia". Graunt fue elegido socio fundador de la Royal Society en 1662.

El mérito de las "Observations" fue inmediatamente reconocido, y fomentó el estudio de las estadísticas de vida en el continente. El libro alcanzó varias ediciones. La quinta, publicada tras la muerte de Graunt fue ampliada por Petty. Los historiadores han discutido largo tiempo la contribución de Petty al trabajo original. Aubrey que era malicioso, sólo dice que Graunt fue "inspirado" por Petty, pero implica mucho más.
Parece indudable que el libro es una obra conjunta.

Desde luego, Graunt escribió la mayor parte, incluidas las aportaciones científicas más valiosas. Petty añadió lo que Thomas Browne llamaría "Elegancia", y así aumentó la popularidad del libro. Sir William Petty era un hombre presuntuoso y algo engreído, incapaz de decidir si patrocinar a Graunt o acreditar su trabajo. No hay pruebas de que alguna vez hubiese entendido la importancia y originalidad de lo que había hecho su amigo.

Graunt fue miembro del consejo común de la ciudad y desempeñó otros cargos, pero al convertirse al catolicismo dejó el comercio y cualquier otra obra pública. Graunt tenía cabeza y talento para el trabajo, y era jocoso y fecundo en su conversación. Graunt murió de ictericia la víspera de Pascua en 1674 y fue enterrado en la iglesia de St. Dunston.

John Arbuthnot

En los trabajos de Graunt y Halley se basó John Arbuthnot en 1970 para probar la existencia de Dios. Su argumento dice: No es posible la suposición de que el sexo está distribuido entre la descendencia humana en una forma puramente casual; debe intervenir una providencia divina que controla las proporciones de los sexos. La demostración de Arbuthnot es el primer ejemplo conocido de inferencia estadística. Anchenwall un economista, acuñó en 1760 la palabra estadística, que deriva del término italiano statista.
La raíz de la palabra procede del latín status que significa estado o situación.

La Ley de los Grandes Números

En el famoso libro de Jacob Bernoulli, Aos Conjectandi, aparece un teorema de importancia cardinal para la Teoría de Probabilidades, comúnmente llamado Teorema de Bernoulli, y también conocido como Ley de los grandes números, nombre que le fue dado por el matemático francés, Simeon Poisson (1781-1840). Este teorema fue el primer intento para deducir medidas estadísticas a partir de probabilidades individuales. El tiempo empleado para escribir este libro no fue perdido, si consideramos la importancia central del resultado. Matemáticos, científicos y filósofos han dedicado más de veinte años examinando y discutiendo el significado exacto del Teorema y su alcance en aplicaciones estadísticas.

El teorema es más sencillo de exponer. De hecho, cuando se ve por primera vez, uno se pregunta cómo Bernoulli pudo preocuparse durante veinte años y cómo ha promovido tantas controversias posteriormente. El hecho es, que es un conjunto de sutilezas y artificios; cuando más lo piensa uno, más complicado lo ve. Bernoulli tuvo un trabajo loco montando el engranaje, lo cual lo distrajo de prever los embrollos lógicos y filosóficos que planteaba. "Si la probabilidad de un suceso es p, y si se hace un número infinito de pruebas, la producción de aciertos es, sin duda p". Aquí, tienen una simple exposición del Teorema de Bernoulli: si la probabilidad de que ocurra un hecho en una prueba única es p, y si se hacen varias pruebas, inmediatamente y en las mismas condiciones, la proporción más probable de que ocurran los hechos en el número total de pruebas es también p; aún más, la probabilidad que la porción en cuestión difiere de p en menos que una cantidad dada, por pequeña que sea, aumenta al mismo tiempo que aumenta el número de pruebas.

Tirando al aire su discreción matemática "un estudioso del sujeto llega a esta definición correcta".

Otra definición más válida: "En un conjunto bastante amplio de "a" elementos es casi seguro que la frecuencia relativa de "b" elementos se aproximará a la probabilidad de un elemento "a" estando "b" dentro de cualquier grado de aproximación deseada". Aquí la frase "casi seguro" ha de entenderse como un medio conveniente para decir que hay una probabilidad tan cercana como queramos a 1.

Como una demostración de la importancia de la Ley de los grandes números en asuntos prácticos es suficiente mencionar los Seguros. Supongamos que la probabilidad de que un hombre de cierta edad y constitución muera en el transcurso de un año es 1/10. si tal individuo decide asegurarse, ésta es la fracción que ha de tener en cuenta y usar cuando tome su decisión. Pero la compañía de seguros que se ofrece a cubrir el riesgo de su muerte en este período tiene en consideración otra probabilidad que se deriva de esta probabilidad. Si hay un gran número de personas de las mismas características, que aseguran sus vidas en esa compañía, hay una probabilidad muy elevada de que la compañía no tenga que pagar a más de, aproximadamente, un décimo de las pólizas. Si, por consiguiente, la compañía carga en cada caso una prima de más de un décimo del total de la póliza, es muy probable que tenga bastante superávit después de pagar todos los derechos, para cubrir los gastos administrativos y distribuir un dividendo a sus accionistas.

Mientras mayor sea el número de personas que se asegura en la compañía, mayor es la probabilidad de que las finanzas de la compañía sean sanas siempre que las primas estén calculadas como acabamos de decir. Esta es la consideración fundamental que distingue el negocio de una compañía de seguros de una apuesta.

Girolamo Cardano

Cardano nació en Pravia en 1501 y murió en 1576. Su vida es una serie de actos incoherentes que pertenecen tanto a la historia de la Matemática como a la de la Astrología y a la de la Patología. Realizó sus primeros estudios en su ciudad natal y luego en la Universidad de Padua, donde alcanzó la Licenciatura en Medicina que ejerció en Sacco y en Milán durante el período de 1524 a 1556. Durante estos años estudió Matemáticas y publicó sus principales obras. Entre estas destaca el Ars Magna, en la cual se presentan raíces negativas de una ecuación, algunos cálculos con números imaginarios y la fórmula de la ecuación cúbica que ha pasado a la historia con el calificativo de Cardámica, aunque ya se sabe que es de Fortaglia, con quien tuvo una de las polémicas más agrias en la historia de las Matemáticas. Se le atribuye la primera discusión sobre "Probabilidad" en su manual para jugadores "Siber De Ludo Aleae" (Manual para tirar dados).

Karl Friedrich Gauss – (1777-1855)

Junto con Arquímedes y Newton, Gauss es uno de los tres grandes de la Matemática. Ellos aportaron conceptos muy útiles en sus distintas ramas tanto en su forma pura como aplicada.

La precocidad de Gauss fue evidente antes de los tres años de edad. Cuando su padre hacía la nómina para pago de los trabajadores, sin darse cuenta que su hijo seguía sus acciones, al terminar, el niño exclamó "Padre el cálculo está equivocado". Al comprobarlo notó que el resultado que le dijo el niño era correcto.

Gauss se hizo notable, ya que a los doce años criticó los fundamentos de la Geometría Euclidiana, a los trece le interesaba la posibilidad de la Geometría NoEuclidiana, a los quince entendió el concepto de convergencia de líneas y probó el binomio de Newton, a los dieciocho inventó el método de los mínimos cuadrados, a los diecinueve, el 30 de marzo de 1796, descubrió la construcción del polígono de 17 lados sólo con regla y compás.

La ley de Gauss de la distribución normal de errores y su curva en forma de campana usada por maestros, estadistas, comerciantes, etcétera, se denomina también curva normal de frecuencias y encuentra sus raíces en la Teoría Matemática de los juegos de azar.

Su lema fue: Pauca. Sed natura, que significa: "Poco, pero maduro".

Johann Von Neumann – (1909-1957)

Epistemólogo austriaco contemporáneo. VON Newsmann llevó a cabo la primera demostración del Teorema Minimax, base fundamental de la Teoría de juegos, que fue propuesto primeramente por Emile Borel en 1921. También fue pionero de la Teoría de Computadoras, habiendo diseñado y construido el llamado MANIAC (analizador matemático, integrador numérico y computador) en el Instituto para estudios avanzados de Pinceton, en 1952. Sus ideas fundamentales sobre la axiomatización de las matemáticas las ha expuesto en varias memorias especialmente en Eine Axiomatiserung der Mengenlebre, Crelle, 1925 y Axiomatiserung der Mengenlebre Math, Zaitrehj 1928.

La situación actual de la Estadística se debe al esfuerzo de grandes matemáticos y científicos. Entre los más famosos se puede mencionar a Laplace, Fermat, Jacques, Bernoulli y Gauss, quienes intervinieron en el primer y más importante estudio de la probabilidad en los siglos XVIII y XIX. El matemático belga Quetelet, los estadísticos escandinavos Charlier y Gram, los ingleses Pearson, Fisher, Galton, también asociaron sus nombres al progreso de esta nueva disciplina, a la que dotaron de bases matemáticas sólidas. Como se observa, la evolución de la Estadística estuvo conformada por una serie de necesidades que condujeron al hombre a su creación.

1.3.- La estadística en las actividades realizadas por los psicologos.

El término estadística se refiere a datos numéricos, tales como promedios, medianas, porcentajes y números índices que ayudan a entender una gran variedad de negocios y situaciones económicas. Sin embargo, el campo de la estadística es mucho más que datos numéricos.

En un sentido amplio, la estadística se define como “el arte y la ciencia de reunir datos, analizarlos, presentarlos e interpretarlos”.

En los negocios y en la economía, la información obtenida al reunir datos, analizarlos, presentarlos e interpretarlos proporciona a directivos, administradores y personas que deben tomar decisiones una mejor comprensión del negocio o entorno económico, permitiéndoles así tomar mejores decisiones con base en mejor información.

El muestreo es imprescindible en la investigación, ya sea ésta de cualquier ciencia aplicada, también ha sido el proceso por medio del cual algunas disciplinas han podido introducir en ellas metodologías y procedimientos para su consolidación como tal, una de las grandes disciplinas beneficiadas es la psicología.

En psicología la estadística es una herramient de gran utilidad para ayudar a comprender major los fenómenos psicológicos. Los datos estadisticos se usan para crear hipótesis, realizer pruebas, generar conclusiones y guiar la toma de decisiones. La estadística tambien ayuda a los investigadores a identificar patrones y relaciones entre los factores de estudio, lo que les ayuda a entender major los problemas y hallarles una solución. Ademas los datos estadisticos se usan para determiner la significancia de los resultados de la investigación lo que ayuda a los investigadores a saber si hay una relación real entre entre los factores estudiados. Por lo tanto, es evidente que la estadística es de gran importancia para la investigación en psicología, ya que ayuda a entender major los fenomenos psicológicos.

La estadística tambien puede ayudar a los psicólogos a identificar patrones de conducta y comportamientos problematicos en los pacientes, lo que les permite diseñar tratamientos que se adapten mejor a las necesidades especificas de los pacientes. Esto puede ser particularmente útil para aquellos pacientes que no responden bien a los tratamientos tradicionales.

En definitiva, comprender como la estadística puede mejorar la psicología es un paso importante para ayudar a los pacientes a obtener los mejores resultados posibles. Al utilizer la estadística para comprender mejor los patrones de comportamiento y los resultados de los tratamientos, los psicólogos pueden mejorar la calidad de vida de sus pacientes. Esto abre la puerta a una major comprension de los problemas de salud mental y a una mayor satisfacción con el tratamiento. Parece que la estadística puede tener un papel muy importante en el futuro de la psicología.

1.4.- APLICACIONES DE LA ESTADÍSTICA

Aunque comúnmente se asocie a estudios demográficos, económicos y sociológicos, gran parte de los logros de la estadística se derivan del interés de los científicos por desarrollar modelos que expliquen el comportamiento de las propiedades de la materia y de los caracteres biológicos. La medicina, la biología, la física y, en definitiva, casi todos los campos de las ciencias emplean instrumentos estadísticos de importancia fundamental para el desarrollo de sus modelos de trabajo.

Campos de aplicación

La estadística es una ciencia de aplicación práctica casi universal en todos los campos científicos:

- En las ciencias naturales: se emplea con profusión en la descripción de modelos termodinámicos complejos (mecánica estadística), en física cuántica, en mecánica de fluidos o en la teoría cinética de los gases, entre otros muchos campos.

- En las ciencias sociales y económicas: es un pilar básico del desarrollo de la demografía y la sociología aplicada.

- En economía: suministra los valores que ayudan a descubrir interrelaciones entre múltiples parámetros macro y microeconómicos.

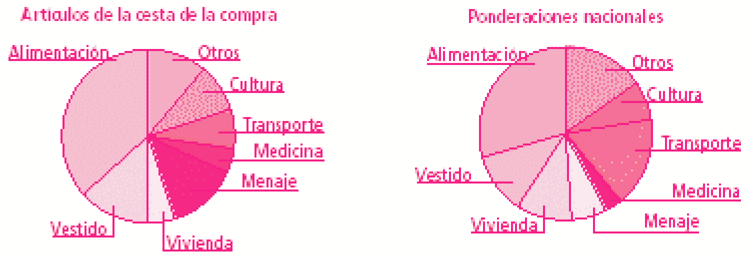
- En las ciencias médicas: permite establecer pautas sobre la evolución de las enfermedades y los enfermos, los índices de mortalidad asociados a procesos morbosos, el grado de eficacia de un medicamento, etcétera.

1.5.- Presentación de datos

Los datos estadísticos se presentan generalmente expresando el valor de la frecuencia absoluta que toman las variables significativas de un estudio, ya correspondan a una población o a una muestra. La frecuencia absoluta de un valor o de una modalidad de una variable estadística es el número de datos observados que presentan ese valor o modalidad. El cociente entre la frecuencia absoluta de un valor o modalidad y el número total de datos es llamado frecuencia relativa. También suelen presentarse los datos en forma de porcentaje (es decir, en forma de razón de denominador 100).

Una razón se obtiene como el cociente entre dos cantidades numéricas comparables. Si el cociente se refiere a dos cantidades que se indican en unidades distintas, la razón recibe el nombre de tasa. Un ejemplo de tasa es la densidad de población, que se define como el número de habitantes por kilómetro cuadrado y que se aplica habitualmente en los estudios demográficos.

Dada una suma de varios sumandos, si el cociente hace referencia a la división numérica entre uno de los sumandos y la suma total, la cantidad expresada se denomina proporción.



Gráficos de sectores que reflejan el Índice de Precios al Consumo (IPC).

Números índices

Otros métodos de presentación de datos utilizados en estadística se basan en el empleo de números índices. Tales números reflejan la evolución que experimenta con el paso del tiempo una variable estadística de interés. Así, se toma como referencia del índice el valor de la variable en un instante dado, de manera que sus valores posteriores se expresan como una razón de cambio con respecto a dicha referencia (a menudo, en forma de porcentaje).

Un ejemplo típico de empleo de números índices es el índice bursátil, cuya definición obedece a criterios diferentes en cada país (índice Dow Jones, en la Bolsa de Nueva York; índice Nikkei, en Tokio, etcétera).

El gráfico de evolución de la Bolsa de Valores constituye un ejemplo de número índice.

Estadísticas comunes

Varios estudios estadísticos comunes que aparecen con frecuencia en los medios de comunicación son los siguientes:

- Encuesta de Población Activa (EPA), elaborada por el Instituto

Nacional de Estadística (INE) con periodicidad trimestral, según recomendaciones de la Organización Internacional del Trabajo (OIT), para obtener y clasificar datos sobre la actividad de la población. Esta encuesta se realiza por muestreo, y los resultados se ordenan por edad, sexo, nivel de estudios, profesión y otros parámetros.

- Índice de Precios al Consumo (IPC), que mide por medios estadísticos la evolución experimentada por los precios de los bienes y servicios consumidos por la población española. Se basa en la Encuesta de Presupuestos Familiares (EPF), y selecciona varios centenares de artículos, clasificados en ocho grupos, que se consideran representativos de la evolución de los precios. Los artículos seleccionados componen lo que se denomina cesta de la compra, considerada en la encuesta.

- Producto Interior Bruto (PIB), que registra la producción nacional de un país en bienes y servicios asociados a procesos considerados productivos.

- Poder adquisitivo, que maneja combinadamente datos del Salario Mínimo Interprofesional (SMI) y el IPC.

¿Qué es la estadística descriptiva?

La estadística emplea métodos descriptivos y de inferencia estadística. Los primeros se ocupan de la recolección, organización, tabulación, presentación y reducción de la información.

Estadística descriptiva

En el caso de la estadística descriptiva se sustituye o reduce el conjunto de datos obtenidos por un pequeño número de valores descriptivos, como pueden ser: el promedio, la mediana, la media geométrica, la varianza, la desviación típica, etc. Estas medidas descriptivas pueden ayudar a brindar las principales propiedades de los datos observados, así como las características clave de los fenómenos bajo investigación.

Por lo general, la información proporcionada por la estadística descriptiva puede ser trasmitida con facilidad y eficacia mediante una variedad de herramientas gráficas, como pueden ser:

Gráficos de tendencia: es un trazo de una característica de interés sobre un periodo, para observar su comportamiento en el tiempo.

Gráfico de dispersión: ayuda al análisis de la relación entre dos variables, representado gráficamente sobre el eje x y el correspondiente valor de la otra sobre el eje y.

Histograma: describe la distribución de los valores de una característica de interés.

Estos métodos gráficos son de mucha utilidad para entender con claridad un fenómeno analizado. La evolución de la inflación, el tipo de cambio, del PBI u otros indicadores macro pueden ser analizados, por ejemplo, con gráficos de tendencia.

Así, la estadística descriptiva constituye un modo relativamente sencillo y eficiente para resumir y caracterizar datos. También ofrece una manera conveniente de presentar la información recopilada.

Este método es potencialmente aplicable a todas las situaciones que involucran el uso de datos. Además de ayudar en el análisis e interpretación de los datos, constituye una valiosa ayuda en el proceso de toma de decisiones.

Aplicaciones de la estadística descriptiva

La estadística descriptiva es aplicable en casi todas las áreas donde se recopilan datos cuantitativos. Puede brindar información acerca de productos, procesos o diversos aspectos del sistema de gestión de la calidad, como también en el ámbito de la dirección y organización de personas, la logística, etc. Algunos ejemplos de dichas aplicaciones son los siguientes:

•Resumen de las mediciones principales de las características de un producto.

•Describir el comportamiento de algún parámetro del proceso, como puede ser la temperatura de un horno.

•Caracterizar el tiempo de entrega o el tiempo de respuesta en el sector de los servicios.

•Procesar datos relacionados con muestras a clientes, tales como la satisfacción o insatisfacción del cliente.

•Ilustrar la medición de los datos, tales como los datos de calibración del equipo.

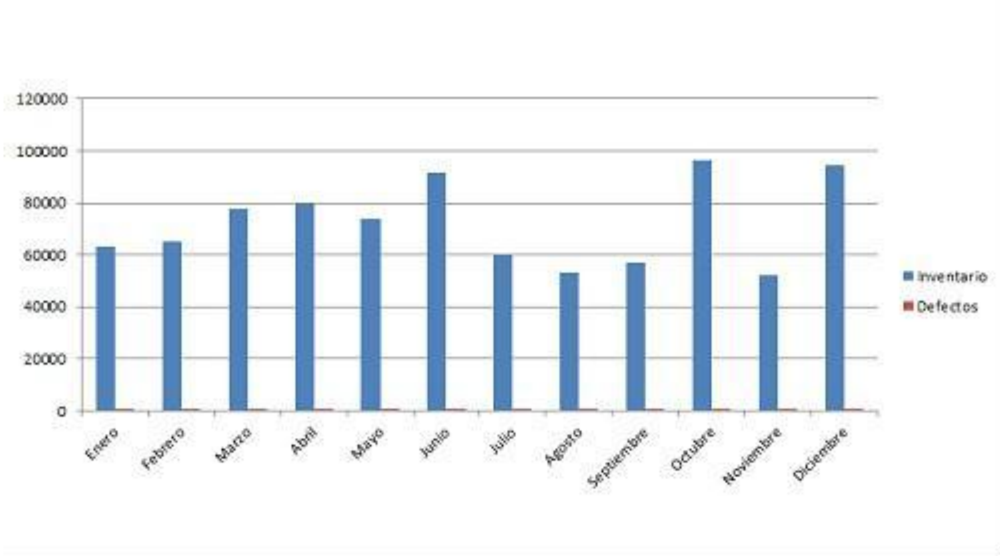
•Visualizar el resultado del desempeño de un producto en un periodo mediante un gráfico de tendencia

Tipos de Gráficas.

Los tipos de gráficas son muy variados y se pueden describir a continuación:

Gráfica de Columna

Los gráficos de columna sirven para exhibir las modificaciones que, en un tramo de tiempo, han sufrido determinados datos, comparándolos entre diversos elementos. Por lo general, la organización horizontal se corresponde con las categorías, y verticalmente se ubican los valores; para así resaltar la variación que se ha producido al pasar el tiempo.



Gráfica de columnas en perspectiva 3D

Una gráfica de columnas en perspectiva 3D se utiliza para establecer comparaciones entre puntos de datos colocados en dos ejes.

Gráfica de Cono, cilindro y pirámide

Las distintas gráficas de datos, dispuestas en forma de cono, cilindro y pirámide, son aquellas capaces de mejorar la presentación de gráficos de columnas y barras 3D, mostrando y comparando datos de la misma manera.

Existe simplemente la diferencia de que estos tipos de gráfico muestran sus datos en diversas formas: cónicas, cilíndricas y piramidales; en vez de en forma de rectángulos horizontales.

Gráfica de Barra

Los gráficos de barra son aquellos que revelan cotejos entre elementos individuales. En este tipo de gráficas, las categorías se muestran organizadas de manera vertical; mientras que los valores se ordenan horizontalmente.

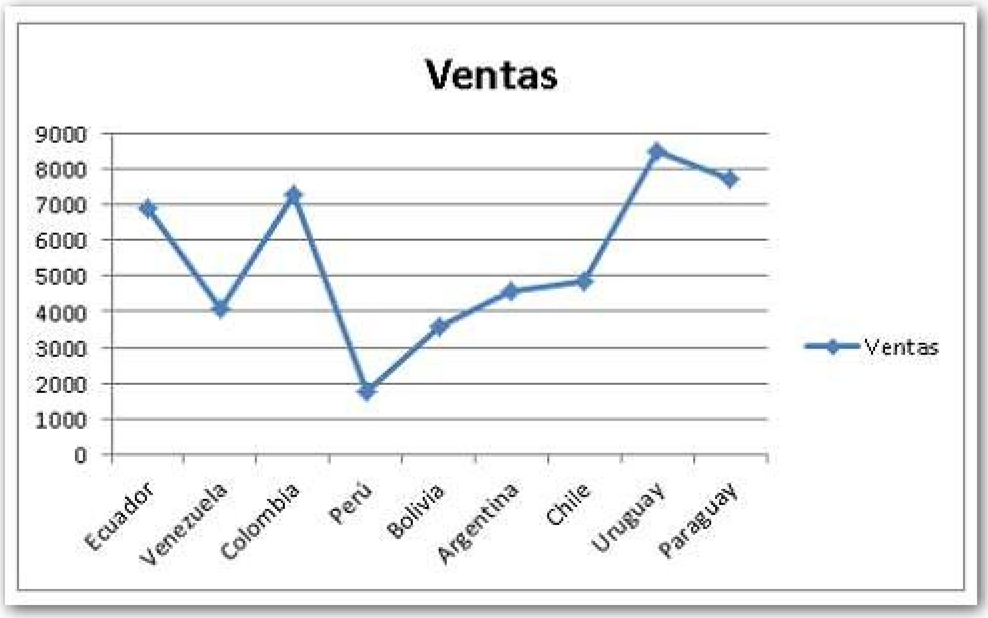
Todo esto, con el propósito de poder concentrarnos en comparar los valores y poner menos firmeza en el tiempo transcurrido.

Gráfica de barras apiladas

Los gráficos de barras apiladas son los que muestran la relación de los elementos individuales con el todo.

Gráfica de Línea

Los gráficos de líneas son aquellos que muestran las predisposiciones existentes en los datos a intervalos exactos.



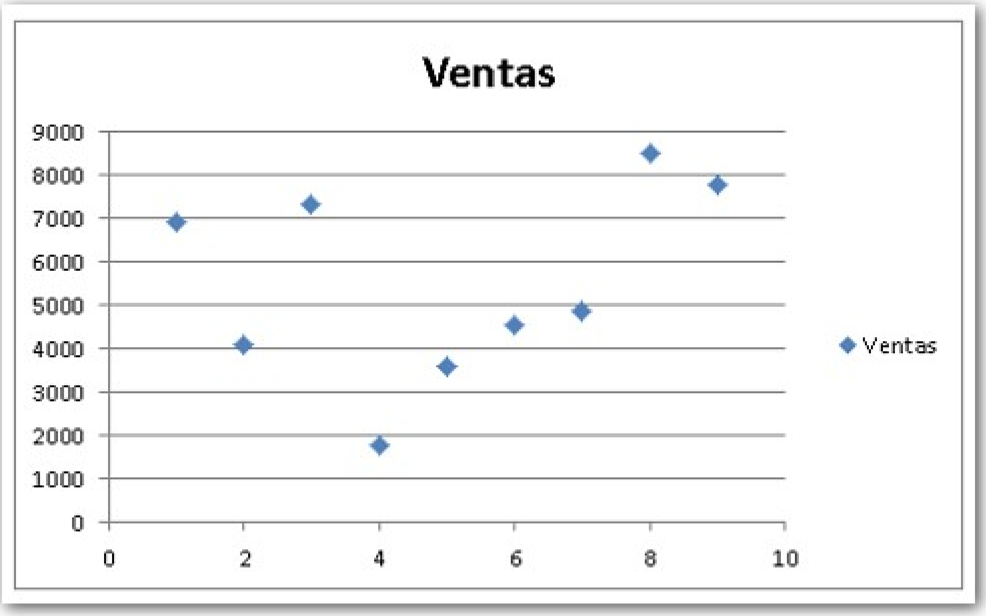
Gráfica de Área

Los gráficos de área enfatizan lo que sería la magnitud de los cambios con el transcurso del tiempo.

Al mostrar la suma de los valores trazados, un gráfico de área también muestra la relación de las partes con un todo.

Gráfica XY (Dispersión)

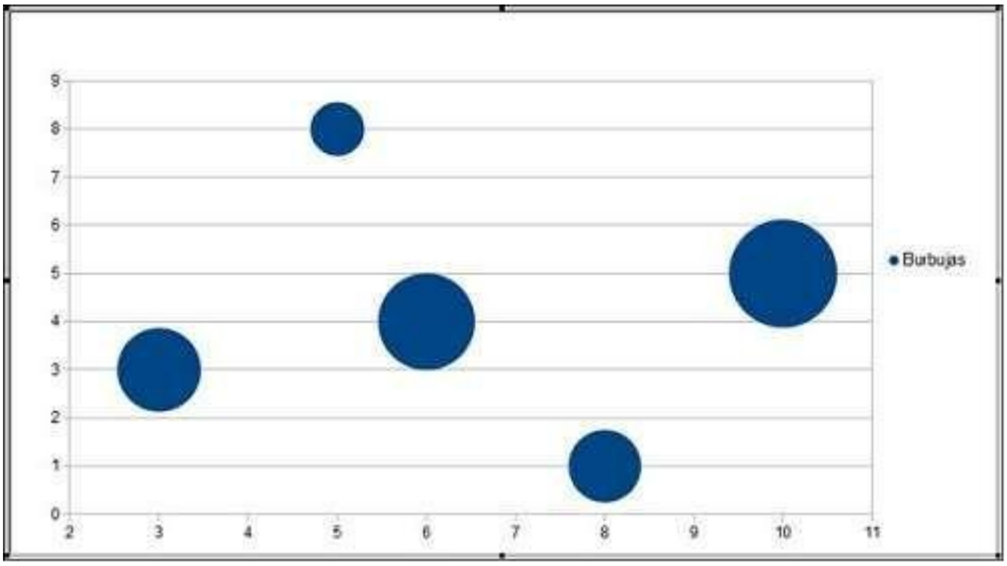
Los gráficos XY (Dispersión) exponen la correspondencia entre los valores numéricos de diferentes grupos de datos o delinean dos series de números como una única serie de coordenadas XY. Es así como esta clase de gráficos muestra los intervalos o agrupaciones de datos; y suele usarse para representar datos de carácter científico.



Si los datos se ordenan, los valores X irán posicionados en una fila o columna, mientras que los valores de Y se corresponderán en las filas o columnas adyacentes.

Gráfica de Burbujas

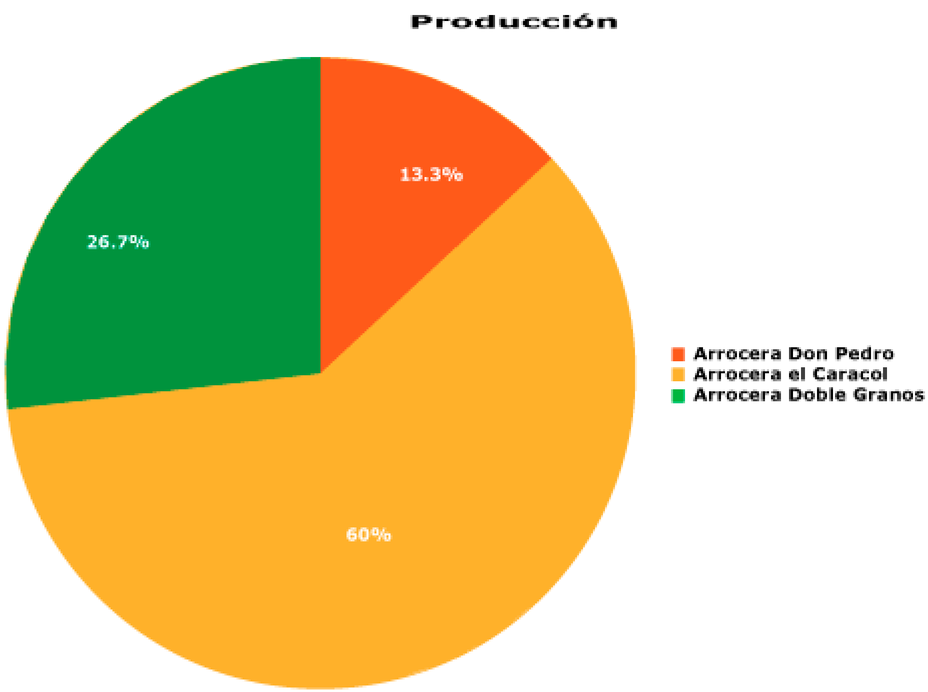
Un gráfico de burbujas es en realidad un tipo de gráfico XY (dispersión). El tamaño del marcador de datos muestra el valor de una tercera variable.



Con el objeto de ordenar los datos, se deben situar los valores X en una fila o columna y, a continuación, debe introducir los valores Y y los tamaños de burbuja correspondientes en las filas o columnas inmediatas.

Gráfica de Burbujas

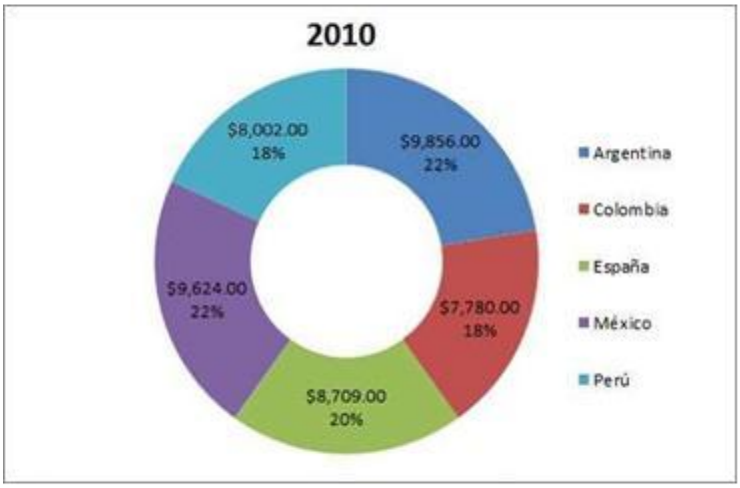
El gráfico circular es aquel que indica el tamaño proporcional de los elementos que componen una serie de datos basándose en la suma de sus elementos. Como resultado, debe mostrar una única serie de datos.



Es un tipo de gráfica ventajosa en los casos donde se busca enfatizar un elemento revelador.

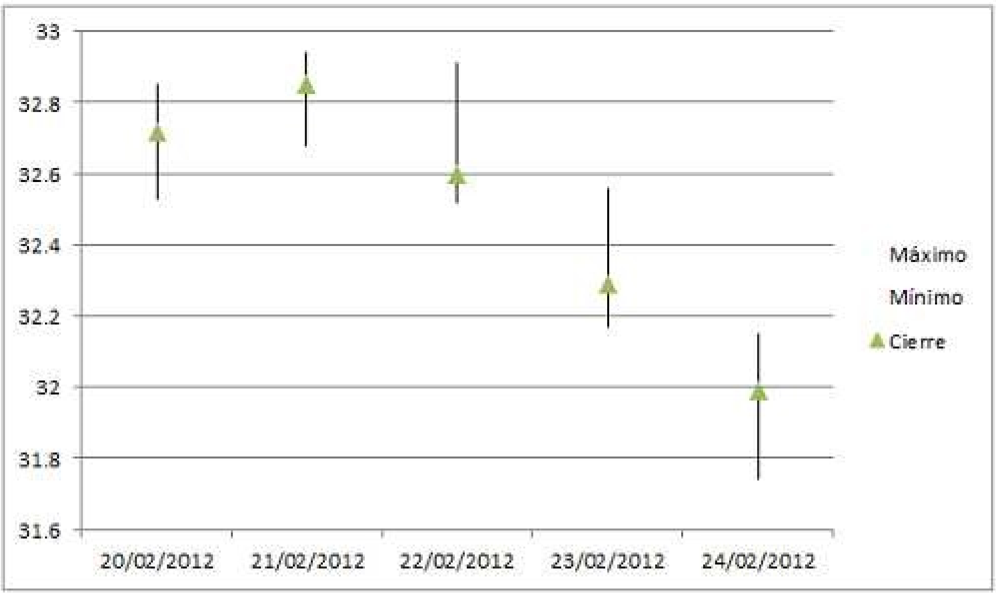
Gráfica de Burbujas

De modo similar a como ocurre con los gráficos circulares, los gráficos de anillos exponen la correspondencia de las partes con un todo; si bien puede contener más de una única serie de datos.



Cada anillo de un gráfico de anillos constituye una serie de datos. Gráfica de Existencias

El gráfico de existencias demuestra el máximo, mínimo, y cierre de existencias, y se usa para ilustrar la cotización de acciones. De manera similar, este tipo de gráficas puede usarse para datos científicos; por ejemplo, para demostrar cambios de temperatura.



Se debe ordenar los datos en el orden adecuado para éstos y otras clases de gráficos de cotizaciones.

Gráfica de Cotizaciones

Un gráfico de cotizaciones es aquel que calcula el volumen que tienen dos ejes de valores; uno que se corresponde a las columnas que miden el volumen y el siguiente para cotizar de los valores.

Se puede incluir el volumen en un gráfico Máximo, Mínimo, Cierre o en un gráfico

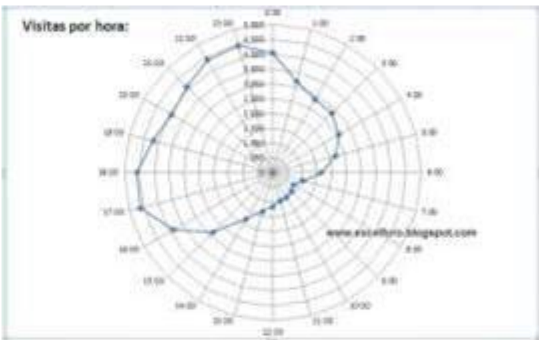
Apertura, Máximos, Mínimos, Cierre. Gráfica de Superficie

Los gráficos de superficie se usan en caso de querer hallar las combinaciones más acertadas entre dos conjuntos de datos.

Tal y como sucede en un mapa topográfico, los colores y los diseños suelen indicar las áreas que se hallan dentro del rango de valores esperado.

Gráfica Radial

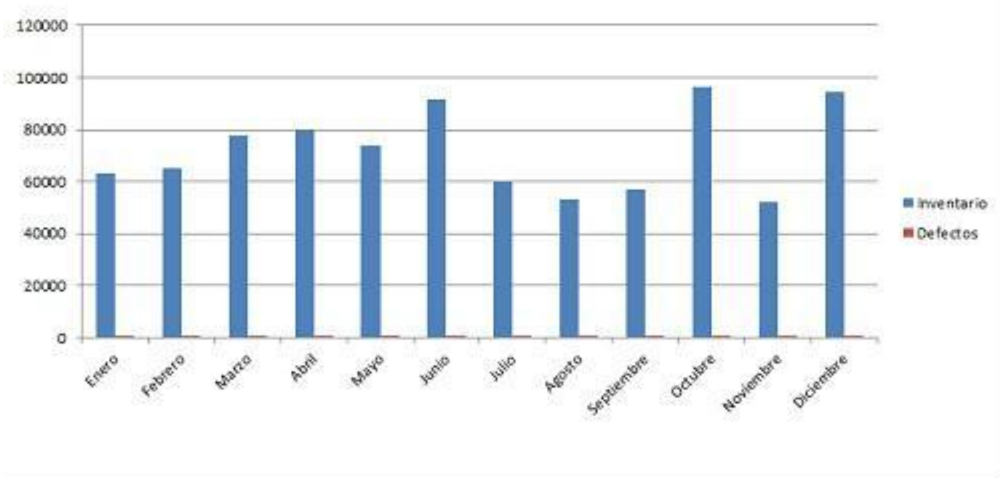
En este tipo de gráficas, cada categoría se inicia en su propio eje de valores, cuyo epicentro se encuentra ubicado en el punto central. De este modo, son las líneas quienes conectan todos los valores de las mismas series.



Tipos de Gráficas

Los tipos de gráficas son muy variados y se pueden describir a continuación: Gráfica de Columna

Los gráficos de columna sirven para exhibir las modificaciones que, en un tramo de tiempo, han sufrido determinados datos, comparándolos entre diversos elementos. Por lo general, la organización horizontal se corresponde con las categorías, y verticalmente se ubican los valores; para así resaltar la variación que se ha producido al pasar el tiempo.

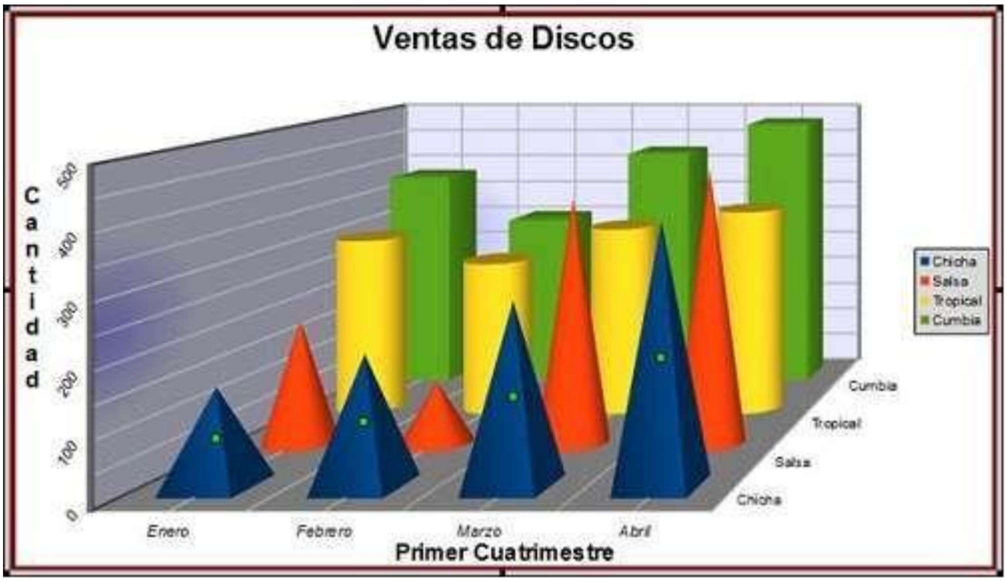


Gráfica de columnas en perspectiva 3D

Una gráfica de columnas en perspectiva 3D se utiliza para establecer comparaciones entre puntos de datos colocados en dos ejes.

Gráfica de Cono, cilindro y pirámide

Las distintas gráficas de datos, dispuestas en forma de cono, cilindro y pirámide, son aquellas capaces de mejorar la presentación de gráficos de columnas y barras 3D, mostrando y comparando datos de la misma manera.



Existe simplemente la diferencia de que estos tipos de gráfico muestran sus datos en diversas formas: cónicas, cilíndricas y piramidales; en vez de en forma de rectángulos horizontales.

Gráfica de Barra

Los gráficos de barra son aquellos que revelan cotejos entre elementos individuales. En este tipo de gráficas, las categorías se muestran organizadas de manera vertical; mientras que los valores se ordenan horizontalmente.

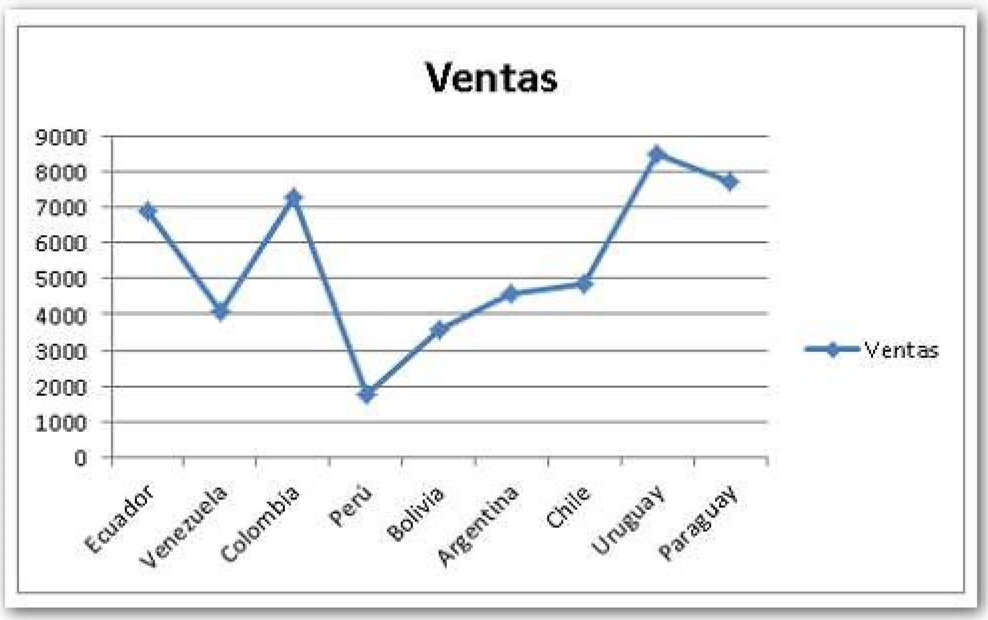
Todo esto, con el propósito de poder concentrarnos en comparar los valores y poner menos firmeza en el tiempo transcurrido.

Gráfica de barras apiladas

Los gráficos de barras apiladas son los que muestran la relación de los elementos individuales con el todo.

Gráfica de Línea

Los gráficos de líneas son aquellos que muestran las predisposiciones existentes en los datos a intervalos exactos.



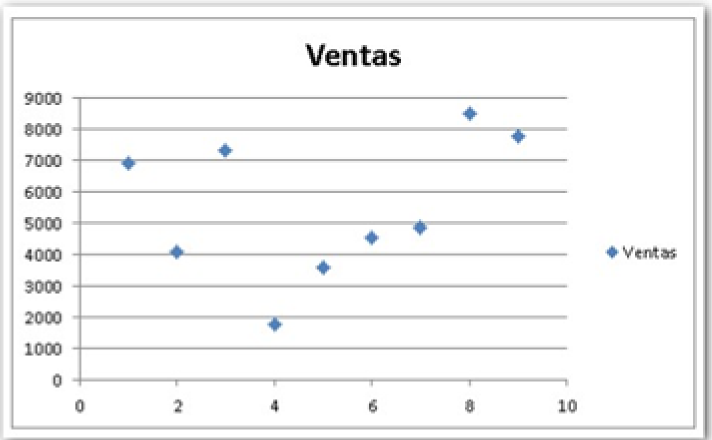
Gráfica de Área

Los gráficos de área enfatizan lo que sería la magnitud de los cambios con el transcurso del tiempo.

Al mostrar la suma de los valores trazados, un gráfico de área también muestra la relación de las partes con un todo.

Gráfica XY (Dispersión)

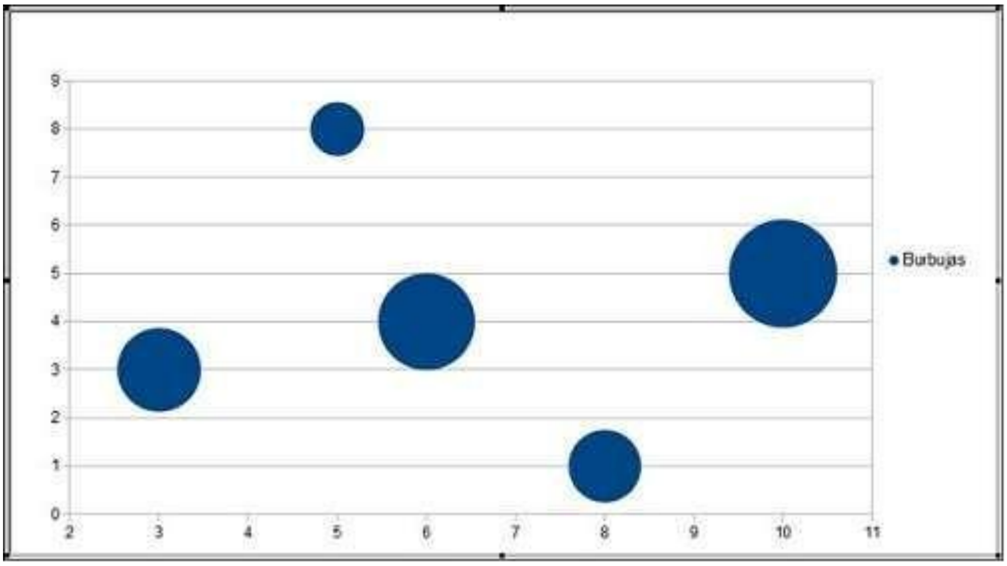
Los gráficos XY (Dispersión) exponen la correspondencia entre los valores numéricos de diferentes grupos de datos o delinean dos series de números como una única serie de coordenadas XY. Es así como esta clase de gráficos muestra los intervalos o agrupaciones de datos; y suele usarse para representar datos de carácter científico.



Si los datos se ordenan, los valores X irán posicionados en una fila o columna, mientras que los valores de Y se corresponderán en las filas o columnas adyacentes.

Gráfica de Burbujas

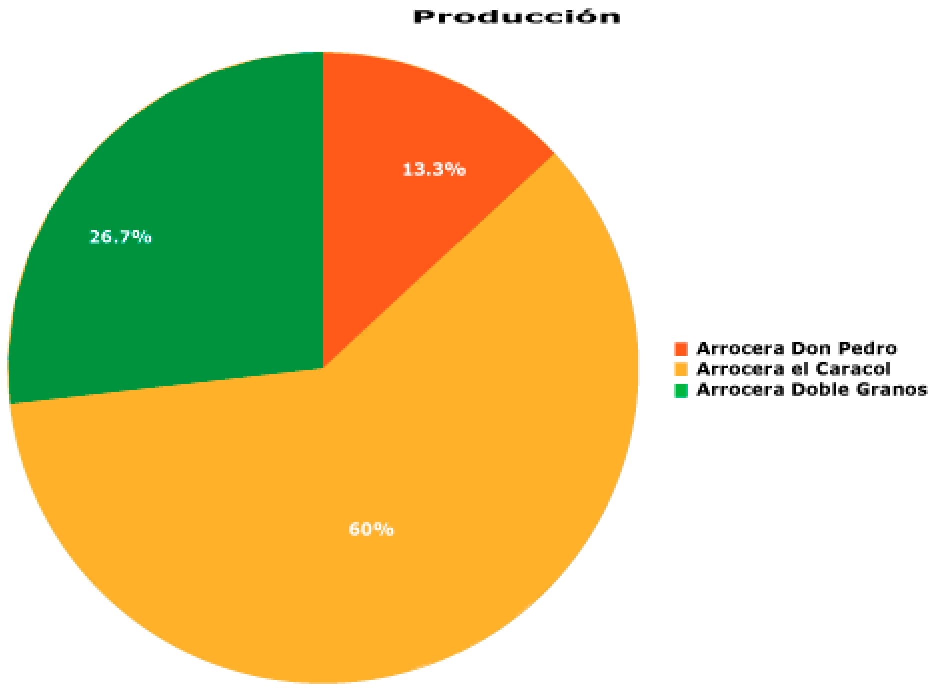
Un gráfico de burbujas es en realidad un tipo de gráfico XY (dispersión). El tamaño del marcador de datos muestra el valor de una tercera variable.



Con el objeto de ordenar los datos, se deben situar los valores X en una fila o columna y, a continuación, debe introducir los valores Y y los tamaños de burbuja correspondientes en las filas o columnas inmediatas.

Gráfica de Burbujas

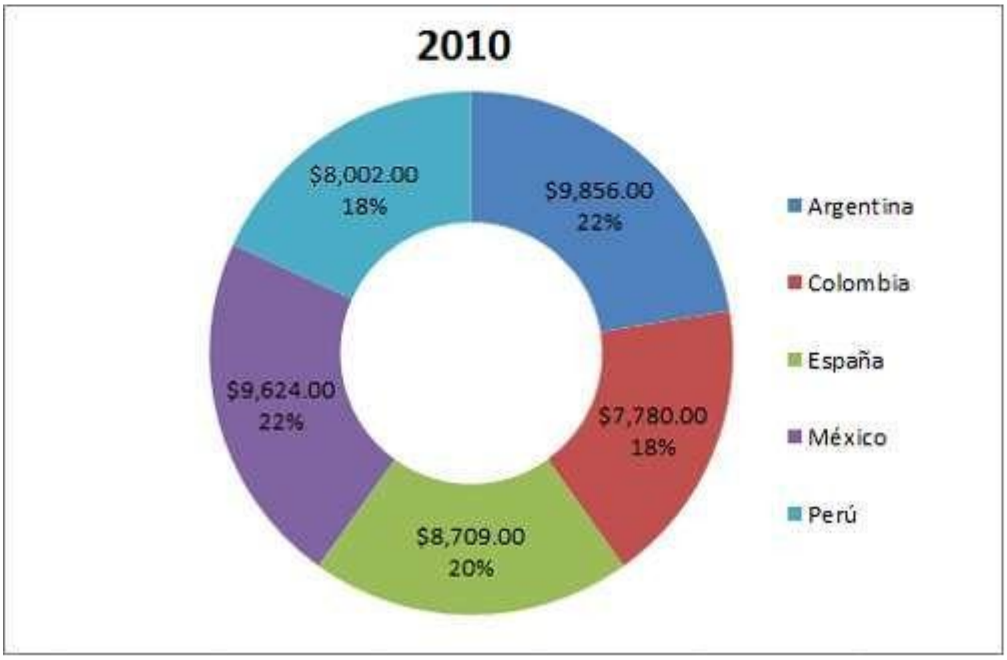
El gráfico circular es aquel que indica el tamaño proporcional de los elementos que componen una serie de datos basándose en la suma de sus elementos. Como resultado, debe mostrar una única serie de datos.



Es un tipo de gráfica ventajosa en los casos donde se busca enfatizar un elemento revelador.

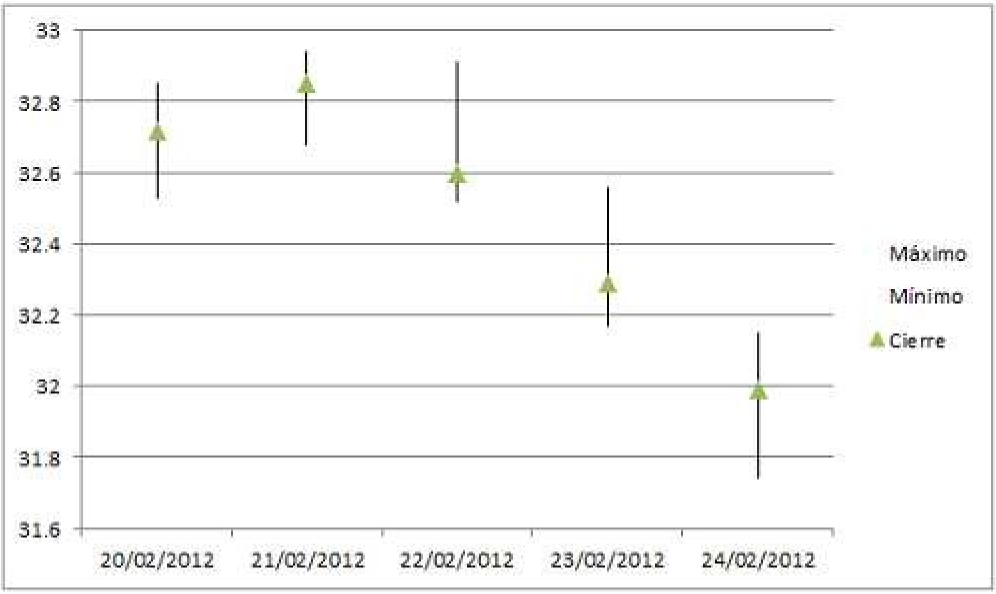
Gráfica de Burbujas

De modo similar a como ocurre con los gráficos circulares, los gráficos de anillos exponen la correspondencia de las partes con un todo; si bien puede contener más de una única serie de datos.



Cada anillo de un gráfico de anillos constituye una serie de datos. Gráfica de Existencias

El gráfico de existencias demuestra el máximo, mínimo, y cierre de existencias, y se usa para ilustrar la cotización de acciones. De manera similar, este tipo de gráficas puede usarse para datos científicos; por ejemplo, para demostrar cambios de temperatura.



Se debe ordenar los datos en el orden adecuado para éstos y otras clases de gráficos de cotizaciones.

Gráfica de Cotizaciones

Un gráfico de cotizaciones es aquel que calcula el volumen que tienen dos ejes de valores; uno que se corresponde a las columnas que miden el volumen y el siguiente para cotizar de los valores.

Se puede incluir el volumen en un gráfico Máximo, Mínimo, Cierre o en un gráfico

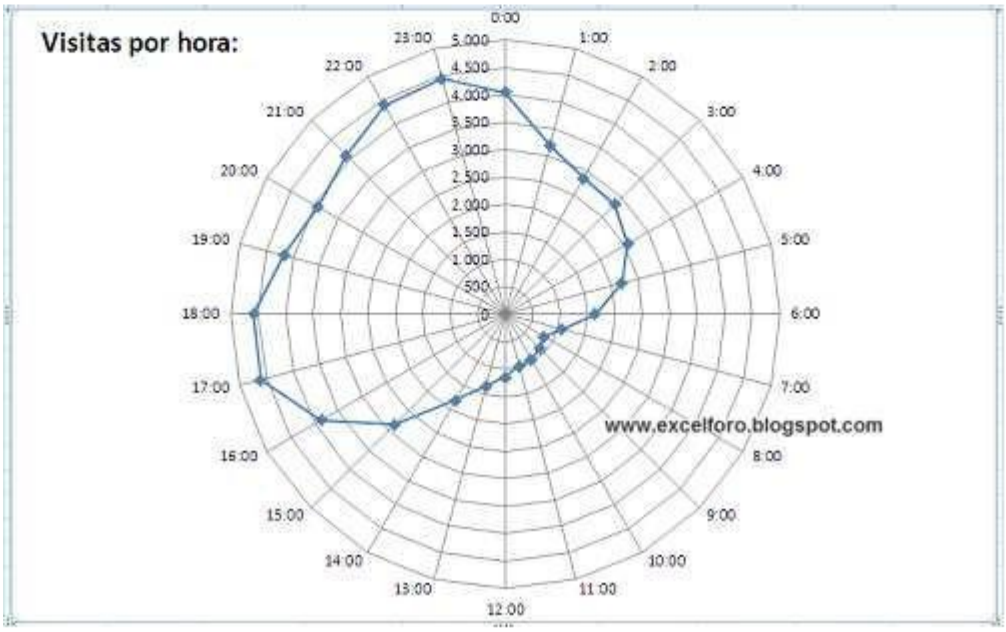
Apertura, Máximos, Mínimos, Cierre. Gráfica de Superficie

Los gráficos de superficie se usan en caso de querer hallar las combinaciones más acertadas entre dos conjuntos de datos.

Tal y como sucede en un mapa topográfico, los colores y los diseños suelen indicar las áreas que se hallan dentro del rango de valores esperado.

Gráfica Radial

En este tipo de gráficas, cada categoría se inicia en su propio eje de valores, cuyo epicentro se encuentra ubicado en el punto central. De este modo, son las líneas quienes conectan todos los valores de las mismas series.



El gráfico radial es un tipo de gráfica que contrasta los valores agregados de muchas series de datos. En estos, la mayor parte del área es abarcada por una serie de datos que simboliza la mayor representación del conjunto total de datos.

1.6.- Diagrama de caja y bigotes.

También conocido como diagrama de caja y bigote, box plot, box-plot o boxplot. Es un método estandarizado para representar gráficamente una serie de datos numéricos a través de sus cuartiles. De esta manera, el diagrama de caja muestra a simple vista la mediana y los cuartiles de los datos1, pudiendo también representar los valores atípicos de estos.

Cómo expresarlo gráficamente

Para la interpretación de este tipo de gráfico, primero obtenemos la media de cada intervalo, y luego la mediana de la tabla de frecuencias en general. Con estos datos utilizamos la fórmula de la media de cada intervalo elevado a la mediana. Los datos obtenidos en esta fórmula son la interpretación.

```
                +-----+-+
  *    o   |----------|   ||--|
                +-----+-+

+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
0   1   2       4   5       7       9   10      12          15
```

- Ordenar los datos y obtener el valor mínimo, el máximo, los cuartiles Q1, Q2 y Q3 y el rango intercuartílico (RIC):

En el ejemplo, para trazar la caja:

- Valor 7: es el Q1 (25% de los datos)
- Valor 8.5: es el Q2 o mediana (el 50% de los datos)
- Valor 9: es el Q3 (75% de los datos)
- Rango intercuartílico (Q3–Q1)

- Los bigotes», las líneas que se extienden desde la caja, se extienden hasta los valores máximo y mínimo de la serie o hasta 1,5 veces el RIC.

Cuando los datos se extienden más allá de esto, significa que hay valores atípicos en la serie y entonces hay que calcular los límites superior e inferior, Li y Ls.

Para ello, se consideran atípicos los valores inferiores a Q1–1.5•RIC o superiores a

Q3+1.5•RIC. En el ejemplo:

- Inferior: 7–1.5•2 = 4

- Superior: 9+1.5•2 = 12

Ahora se buscan los últimos valores que no son atípicos, que serán los extremos de los bigotes.

- En el ejemplo: 4 y 10
- Marcar como atípicos todos los datos que están fuera del intervalo (Li, Ls). En el ejemplo: 0,5 y 2,5
- Además, se pueden considerar valores extremadamente atípicos aquellos que exceden

Q1–3•RIC o Q3+3•RIC.

De modo que, en el ejemplo:

- Inferior: 7–3•2 = 1
- Superior: 9+3•2 = 15

Utilidad

- Proporcionan una visión general de la simetría de la distribución de los datos; si la mediana no está en el centro del rectángulo, la distribución no es simétrica.
- Son útiles para ver la presencia de valores atípicos también llamados outliers.
- Pertenece a las herramientas de las estadísticas descriptivas. Permite ver como es la dispersión de los puntos con la mediana, los percentiles 25 y 75 y los valores máximos y mínimos.

- Ponen en una sola dimensión los datos de un histograma, facilitando así el análisis de la información al detectar que el 50% de la población está en los límites de la caja.

UNIDAD II DATOS AGRUPADOS Y NO AGRUPADOS

Cuando la muestra que se ha tomado de la población o proceso que se desea analizar, es decir, tenemos menos de 20 elementos en la muestra, entonces estos datos son analizados sin necesidad de formar clases con ellos y a esto es a lo que se le llama tratamiento de datos no agrupados.

Cuando la muestra consta de 30 o más datos, lo aconsejable es agrupar los datos en clases y a partir de estas determinar las características de la muestra y por consiguiente las de la población de donde fue tomada. Antes de pasar a definir cuál es la manera de determinar las características de interés (media, mediana, moda, etc.) cuando se han agrupado en clases los datos de la muestra, es necesario que sepamos cómo se agrupan los datos.

FRECUENCIA DE CLASE

Marca de clase (punto medio): punto que divide a la clase en dos partes iguales. Es el promedio entre los límites superior e inferior de la clase.

Intervalo de clase: para una distribución de frecuencias que tiene clases del mismo tamaño, el intervalo de clase se obtiene restando el límite inferior de una clase del límite inferior de la siguiente.

FRECUENCIA RELATIVA

Es la relación o cociente entre la frecuencia absoluta y el número total de observaciones. Es la proporción entre la frecuencia de un intervalo y el número total de datos.

PUNTO MEDIO

Punto medio es el punto que divide a un segmento en dos partes iguales.

El punto medio de un segmento, es único y equidista de los extremos del segmento. Cumpliendo esta última condición, pertenece a la mediatriz del segmento.

La fórmula para determinar el punto medio de un segmento en el plano, con coordenadas: $(x1, y1)$ y $(x2, y2)$ es: $[(x1 + x2) / 2] + [(y1 + y2) / 2]$

LIMITES

Son los valores extremos que tiene el intervalo de clase, inferior y superior, entre los cuales van a estar los valores de los datos agrupados en ese intervalo de clase

TRATAMIENTO PARA DATOS NO AGRUPADOS.

¿A qué se refiere esto? Cuando la muestra que se ha tomado de la población o proceso que se desea analizar, es decir, tenemos menos de 20 elementos en la muestra, entonces estos datos son analizados sin necesidad de formar clases con ellos y a esto es a lo que se le llama tratamiento de datos no agrupados.

b1. Medidas de tendencia central. Se les llama medidas de tendencia central a la media aritmética, la mediana, la media geométrica, la moda, etc. debido a que al observar la distribución de los datos, estas tienden a estar localizadas generalmente en su parte central. A continuación definiremos algunas medidas de tendencia central y la forma de calcular su valor.

2.1.

-

1) Media aritmética (□ x). También se le conoce como promedio ya que es el promedio de las lecturas o mediciones individuales que se tienen en la muestra, se determina con la fórmula siguiente:

n

□ □ xi

x □ i □ l

n

dónde:

□x = media aritmética xi = dato i

n = número de datos en la muestra

Ejemplos:

1. Se han tomado como muestra las medidas de seis cables usados en un arnés para lavadora, las cuales son; 15.2 cm, 15.0, 15.1, 15.2, 15.1 y 15.0, determine su media aritmética.

Solución:

x ☐ 15.2 ☐ 15.0 ☐ 15.1 ☐ 15.2 ☐ 15.1 ☐ 15.0 ☐ 15.1cm

6

2. Se toman varias muestras de cierto tipo de queso y se determina la cantidad de proteína por cada 100 gramos de queso, encontrándose lo siguiente: 26.5 gramos, 24.8, 25.3, 30.5, 21.4, determine la cantidad promedio de proteína encontrada en la muestra por cada 100 gramos de queso que se elabora.

Solución:

X ☐ 26.5 ☐☐24.8 ☐☐25.3 ☐☐30.5 ☐☐21.4 ☐

25.7 grs

5

3. Se hacen varias lecturas de una muestra que contiene cobre, las lecturas se hacen en un espectrofotómetro de absorción atómica y son la siguientes: 12.3%, 12.28, 12.27, 12.3,

12.24, 15.01, determine la concentración promedio de Cu en la muestra.

Solución:

12.3 ☐ 12.28 ☐ 12.27 ☐ 12.3 ☐ 12.24 ☐ 15.01

x

76.4

. %Cu
☐ ☐ ☐ 12 73

6 6

Si observamos las lecturas del espectrofotómetro nos damos cuenta que el valor de

15.01% es un valor diferente al de las lecturas anteriores, por lo que se descarta el valor ya que se considera un valor atípico, es decir un valor que es debido a circunstancias especiales, en este caso puede ser que se deba al hecho de que se está descalibrando el aparato de absorción atómica o simplemente que se ha equivocado el operador del aparato al tomar la lectura, por lo que la media se debe calcular con las primeras cinco lecturas; como se muestra a continuación:

Solución:

12.3 ☐ 12.28 ☐ 12.27 ☐ 12.3 ☐ 12.24

x ☐

5

☐ 61.39

5

☐ 12.278

%Cu

4. Si deseamos determinar la edad promedio de los estudiantes de una escuela de nivel superior al iniciar sus estudios, suponga que se toman las edades de algunos de los alumnos de cierta clase y estas son las que siguen: 20, 18, 18, 19, 18, 19, 35, 20, 18, 18, 19.

Solución:

Luego, la media se determinará con solo 10 de las edades ya que es necesario descartar la edad de 35 años, que es un dato atípico o un caso especial, por lo que;

20 □ 18 □ 18 □ 19 □ 18 □ 19 □ 20 □ 18 □ 18 □ 19

x □

10

□ 187

10

□ 18.7años

Nota: Cuando es necesario determinar aquellas medidas de tendencia central que hagan uso de todos los datos de la muestra se recomienda descartar todos aquellos datos atípicos que se encuentren en la muestra o muestras tomadas.

2) Media geométrica (G). Es la raíz en enésima del producto de los valores de los elementos de la muestra, es usada cuando los valores de los datos de la muestra no son lineales, es decir que su valor depende de varios factores a la vez, se determina de la siguiente forma:

G ☐ n x * x

* ...* x

1   2   n

Dónde:

G = media geométrica xi = dato i

n = número de datos en la

muestra

Ejemplos:

1. Las siguientes temperaturas han sido tomadas de un proceso químico, 13.4oC, 12.8, 11.9, 13.6, determine la temperatura promedio de este proceso.

Solución:

G = 4 13.4x12.8x11.9x13.6 ☐ 4 27758.7968 = 12.9077 oC

2. Las siguientes temperaturas han sido tomadas de un proceso para fabricar queso chihuahua, 21.4oC, 23.1, 20.2, 19.7, 21.0, determine la temperatura promedio de este proceso.

Solución:

G = 5 21.4x23.1x20.2x19.7x21.0 □ 5 4131070.852 = 21.048 oC

3) Media aritmética ponderada (xw). Esta media se usa cuando el peso que tiene cada uno de los datos de la muestra es diferente, se calcula de la siguiente manera:

k wi

□xi

xw □ i □ l

□ wi

i □ l

Dónde:

xw = media aritmética ponderada xi = dato i

wi = peso del dato i

Ejemplo:

A continuación se mencionan las materias que Luis Pérez llevó en el primer semestre de Ingeniería Química, el número de créditos y la calificación obtenida;

MATERIA	NUMERO CREDITOS	CALIFICACIÓN
Metodología de la investigación	8	90.5
Matemáticas I	10	100.0
Programación	8	81.0
Química	10	78.0
Dibujo	4	100.0
Economía	8	84.0

Determine la calificación promedio que obtuvo Luis Pérez en su primer semestre.

Solución:

X w □ (8x90.5) □□(10 x100.0) □□(8x81.0) □□(10 x78.0) □□(4x100) □□(8x84.0) =
8 □ 10 □ 8 □ 10 □ 4 □ 8

□ 724 □ 1000 □□648 □□780 □□400 □□672 □ 4224 □ 88
.0

48 48

Nota: Sí comparamos este promedio con el que se obtiene usando simplemente la media aritmética, que es un 88.91, nos damos cuenta de que este último es mayor, por no tomar en cuenta el peso o número de créditos que aporta cada materia a la carrera que se estudia, el promedio de esta persona es menor al de la media aritmética debido a que obtiene una calificación baja es Química que es una de las materias que aporta más créditos.

4) Media armónica (H). La media armónica se define como el recíproco del promedio de los recíprocos de cada uno de los datos que se tienen en la muestra, y se determina de la siguiente manera:

H □ 1 □

n n

1 / n□1 / xi

i

□1

n

□1 / xi

i □1

Ejemplo: Determine la media armónica de los siguientes datos, 3.1, 2.8, 2.84, 3.05, 3.09

Solución:

H □ 5 □

1 / 3.1 □ 1 / 2.8 □ 1 / 2.84 □ 1 / 3.05 □ 1 / 3.09

□ 5 □ 5 □ 2.9703

0.3226 □ 0.3571 □ 0.3521 □ 0.3279 □ 0.3236

1.6833

5) Mediana (xmed). La mediana es aquel valor que se encuentra en la parte central de los datos que se tienen en la muestra una vez que estos han sido ordenados según su valor o magnitud. Para calcular la mediana se presentan dos casos:

a. Cuando el número de datos en la muestra es impar.- En este caso después de ordenar los datos de la muestra en cuanto a su magnitud, es decir de mayor a menor valor o de menor a mayor valor, se procede a localizar aquel dato que se encuentra justo en el centro de los datos o en la parte central de los mismos, el valor de este dato será el que dé valor a la mediana.

Ejemplo:

Los siguientes datos son las mediciones obtenidas de un circuito utilizado en un arnés de lavadora; se toman como muestra siete circuitos y sus mediciones son: 11.3, 11.2, 11.5, 11.2, 11.2, 11.4, 11.5 cm.

Solución:

Ordenando los datos de menor a mayor valor;

11.2, 11.2, 11.2, 11.3, 11.4, 11.5, 11.5

Se observa que el dato 11.3 es el que queda en la parte central, por lo que este es el que dará valor a la mediana; entonces,

xmed = 11.3 cm.

b. Cuando el número de datos en la muestra es par.- En este caso después de ordenar los datos en cuanto a su magnitud, observamos que en la parte central de los datos no se encuentra dato alguno, en este caso, la mediana tomará el valor del promedio de dos datos; el que se encuentra antes de la parte central y el que se encuentra después de la parte central.

Ejemplo:

Los siguientes datos son las mediciones obtenidas de un circuito utilizado en un arnés de lavadora; se toman como muestra ocho circuitos y sus mediciones son: 11.3, 11.2, 11.5,

11.2, 11.2, 11.4, 11.5, 11.4 cm. Solución:

Ordenando los datos de mayor a menor valor,

11.5, 11.4, 11.4, 11.3, 11.2, 11.2, 11.2, 11,1 cm.

Se observa que en la parte central de los datos no hay dato alguno por lo que la mediana se determina con el promedio de los datos subrayados, entonces,

Xmed ☐ 11.3 ☐ 11.2 ☐ 11.25cm

2

Nota: Es imprescindible para calcular el valor de la mediana el que primero se ordenen los datos en cuanto a su magnitud, ya que de no hacerlo, se incurriría en un grave error.

5) Moda (xmod). La moda se define como aquel valor o valores que más se repiten o que tienen mayor frecuencia entre los datos que se han obtenido en una muestra, la muestra de una población nos genera la distribución de los datos una vez que estos se han graficado y en esta gráfica es posible observar la moda o modas de la misma, es por

esto que una distribución de datos puede ser amodal (carece de moda), unimodal (tiene una sola moda), bimodal (tiene dos modas) o polimodal (tiene más de dos modas).

Ejemplos:

1. Determine la moda de los datos que se muestran a continuación, se refieren a la estatura de un grupo de jóvenes; 1.60m, 1.65, 1.70, 1.71, 1.70, 1.70, 1.70, 1.71, 1.70, 1.93, 1.87, 1.85

Solución:

Estatura	Frecuencia
1.60	1
1.65	1
1.70	5*
1.71	2
1.85	1
1.87	1
1.93	1

La tabla muestra la distribución de frecuencias de los datos o el número de veces que estos se repiten, la mayor frecuencia que es 5 corresponde a una estatura de 1.70m, por lo que esta sería la moda.

Luego, xmod = 1.70m

2. Determine la moda de los siguientes datos que se refieren a la edad de alumnos de primer semestre del tecnológico de Chihuahua, 18 años, 17, 19, 21, 19, 18, 22, 22, 18, 18,

17, 19, 19, 19, 18, 20, 21, 20, 18, 19, 18, 19, 18,19, 22, 35

Solución

Edad	Frecuencia
17	2
18	7*
19	8*
20	2
21	2
22	3
35	1

En este caso se observa que las edades que más frecuencia tienen son las de 18 y 19 años, por lo que se concluye que existen dos modas,

Xmod1= 18 años, Xmod2= 19años

Hay que hacer notar que la frecuencia para ambas modas puede ser de igual magnitud o diferente, como en el caso que se ilustra.

b2. Medidas de Dispersión. Cuando se tiene una muestra de datos obtenida de una población cualquiera, es importante determinar sus medidas de tendencia central así como también es básico el determinar qué tan dispersos están los datos en la muestra, por lo que se hace necesario determinar su rango, la varianza, la desviación estándar, etc., ya que una excesiva variabilidad o dispersión en los datos indica la inestabilidad del proceso en análisis en la mayoría de los casos.

1) Rango o recorrido. El rango es la diferencia entre el valor mayor y el valor menor encontrados en la muestra, también se le denomina recorrido ya que nos dice entre que valores hace su recorrido la variable de interés; y se determina de la siguiente manera:

R = VM – Vm

Dónde:

R = rango o recorrido

VM = valor mayor en la muestra

Vm = valor menor en la muestra

Ejemplo:

1. Se han tomado como muestras las mediciones de la resistencia a la tensión de la soldadura usada para unir dos cables, estas son: 78.5kg, 82.4, 87.3, 78.0, 90.0, 86.5, 77.9,

92.4, 75.9, determine su rango o recorrido.

Solución:

VM = 92.4 kg

Vm = 75.9 kg

R = VM – Vm = 92.4 – 75.9 = 16.5 kg

3. Se toman las mediciones de la cantidad de grasa de la leche en gramos por cada 100 ml de leche que entra a un proceso de pasteurización, a continuación se enumeran; 14.85, 15.32, 12.76,

16.29, 15.84, 17.3, 17.61, 16.33, determine el rango o recorrido de la cantidad de grasa de la leche.

Solución: VM = 17.61

Vm = 12.76

R = 17.61 – 12.76 = 4.85gramos

_

2) Desviación absoluta media (d). Esta medida de dispersión nos representa la diferencia absoluta promedio que existe entre cada dato que se encuentra en la muestra y la media de los datos y se determina de la siguiente manera:

$$d = \frac{\sum_{i=1}^{n} |xi - x|}{n}$$

Dónde:

xi = dato i

_ x = media aritmética de la muestra n = número de datos en la muestra Ejemplo:

1. Determine la desviación absoluta media de los siguientes datos que son las concentraciones de plomo de algunas muestras, las que a continuación se enumeran: 18gr,

12, 21, 19, 16, 20, 22

Solución:

Para determinar la desviación absoluta media o promedio, lo primero que hay que hacer es calcular la media aritmética de los datos de la muestra, la que es 128/7 =18.286, luego se procede a calcular el promedio de las diferencias absolutas entre cada dato y la media calculada.

$$\bar{d} = \frac{|18-18.286| - |12-18.286| - \ldots\ldots - |20-18.286| - |22-18.286|}{7} =$$

$$\bar{d} = \frac{0.286 + 6.286 + 2.714 + 0.714 + 2.286 + 1.714 + 3.714}{7} = \frac{17.714}{7} = 2.5305gr$$

La interpretación de este resultado sería que el grado de alejamiento absoluto promedio de los datos con respecto a su media es de 2.5305 gramos.

¿Por qué sacar el valor absoluto de las diferencias entre cada dato y la media aritmética? Si solo se hicieran diferencias entre cada dato y la media aritmética, estas tendrían signos positivos y negativos ya que algunos datos son menores que la media y otros son mayores que la media, luego al sumar las diferencias, con sus signos correspondientes, éstas se irían anulando unas con otras y no sería posible medir leal grado de alejamiento promedio de los datos en la muestra.

3) Varianza o variancia (s2). Es el promedio de las diferencias elevadas al cuadrado entre cada valor que se tiene en la muestra (xi) y la media aritmética (x) de los datos y se determina de la siguiente manera:

$$s^2 = \frac{\sum_{i=1}^{n}\left(xi - \bar{x}\right)^2}{n-1}$$

Donde n es el número de datos en la muestra. Ejemplo:

$$\bar{x} = \frac{14.2+12.1+15.6+18.1+14.3}{5} = \frac{74.3}{5} = 14.86mg$$

$$s^2 = \frac{(14.2-14.86)^2+(12.1-14.86)^2+....+(14.3-14.86)^2}{5-1} =$$

$$s^2 = \frac{0.4356+7.6176+0.5476+10.4976+0.3136}{4} = \frac{19.412}{4} = 4.853mg^2$$

Los siguientes datos es la cantidad de glucosa en miligramos encontrada en muestras de sangre de algunos pacientes, 14.2, 12.1, 15.6, 18.1, 14.3, determine su varianza. Solución:

Lo primero que hay que calcular es la media aritmética de la muestra como ya se ha hecho anteriormente.

Nota: Dentro de la inferencia estadística se plantea la deferencia entre una variancia maestral s2 y una poblacional, representada por □2.

4) Desviación estándar (s). Es la desviación o diferencia promedio que existe entre cada dato de la muestra y la media aritmética de la muestra. Y se obtiene a partir de la varianza, sacándole raíz cuadrada.

$$s = \sqrt{s^2}$$

donde:

s^2= varianza o variancia

Por tanto la desviación estándar de la muestra anterior sería;

$$s = \sqrt{4.853mg^2} = 2.2029mg$$

La interpretación de este resultado sería, que la cantidad de glucosa encontrada en la muestra es en promedio de 14.86 miligramos y que la cantidad de glucosa en la muestra se aleja o dispersa en promedio 1.9704 mg alrededor de la media.

En este caso solo nos interesa conocer el significado de la desviación estándar, aunque es necesario decir que s es la desviación de la muestra y que □ es la desviación de la población, así como s2 es la varianza de la muestra y □2 es la varianza de la po blación.

1- Media aritmética para datos agrupados

Se calcula sumando todos los productos de marca clase con la frecuencia absoluta respectiva y su resultado dividirlo por el número total de datos:

$$\bar{X} = \frac{\text{Suma (marca clase} \times \text{frecuencia absoluta)}}{\text{Total de datos}}$$

La marca clase de una tabla para datos agrupados en intervalos corresponde al promedio de los extremos de cada intervalo.

El intervalo	$\boxed{26 - 30}$	26: corresponde al extremo inferior del intervalo. 30: corresponde al extremo superior del intervalo

En el intervalo anterior la **marca de clase** es 28 es decir $\bar{x} = \dfrac{26+30}{2} = 28$

2- Moda

Es el valor que representa la mayor frecuencia absoluta. En tablas de frecuencias con datos agrupados, hablaremos de intervalo modal.

La moda se representa por Mo.

2.1- Todos los intervalos tienen la misma amplitud.

$$Mo = L_i + \frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})} \cdot t_i$$

Li Extremo inferior del intervalo modal (intervalo que tiene mayor frecuencia absoluta). fi Frecuencia absoluta del intervalo modal. fi-1 Frecuencia absoluta del intervalo anterior al modal. fi+1 Frecuencia absoluta del intervalo posterior al modal. ti Amplitud de los intervalos.

Si los intervalos tienen amplitudes distintas.

En primer lugar tenemos que hallar las alturas. $hi = fi / ti$

Dónde:

hi: altura correspondiente a cada intervalo.

fi: Frecuencia absoluta del intervalo (también se puede utilizar la frecuencia acumulada o relativa) ti: Amplitud de los intervalos

Luego la clase modal es la que tiene mayor altura.

3- Mediana

Es el valor que ocupa el lugar central de todos los datos cuando éstos están ordenados de menor a mayor. La mediana se representa por Me. La mediana se puede hallar sólo para variables cuantitativas.

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. N / 2

Luego calculamos según la siguiente fórmula:

$$Me = L_i + \frac{\frac{N}{2} - F_{i-1}}{f_i} \cdot t_i$$

Li-I es el límite inferior de la clase donde se encuentra la mediana. N / 2 es la semisuma de las frecuencias absolutas.

Fi-I es la frecuencia acumulada anterior a la clase mediana. fi es la frecuencia absoluta del intervalo mediano.

ti es la amplitud de los intervalos.

Ahora veamos un ejemplo:

- En la siguiente tabla se muestran las edades de un grupo de personas.

1° Calculemos la media aritmética:

Edad	Marca clase (X_i)	Frecuencia absoluta (f_i)	Frecuencia acumulada (F_i)
[0 - 10)	5	3	3
[10 - 20)	15	6	9
[20 - 30)	25	7	16
[30 - 40)	35	12	28
[40 - 50)	45	3	31

N = 31

$$\bar{X} = \frac{5 \times 3 + 15 \times 6 + 25 \times 7 + 35 \times 12 + 45 \times 3}{31} =$$

$$\bar{X} = \frac{15 + 90 + 175 + 420 + 135}{31} = \frac{835}{31} = 26{,}94$$

$$\boxed{\bar{X} = 26{,}94}$$

2° Ahora calculemos la mediana (Me) según las fórmulas explicadas más arriba:

Lo primero que debemos hacer para poder calcular la mediana es identificar la clase mediana. Para esto tenemos que buscar el intervalo en el que se encuentre. N / 2 en este caso N / 2 = 31 / 2 $\Rightarrow$ 15,5

Ahora debemos buscar el intervalo donde la frecuencia acumulada (F_i) contenga el valor obtenido (15,5).

Veamos:

Edad	Marca clase (X_i)	Frecuencia absoluta (f_i)	Frecuencia acumulada (F_i)
[0 - 10)	5	3	3
[10 - 20)	15	6	9
[20 - 30)	25	7	16
[30 - 40)	35	12	28
[40 - 50)	45	3	31

$\frac{N}{2} = 15,5$

N = 31

Ahora reemplazamos los datos en la fórmula:

$$Me = L_i + \frac{\frac{N}{2} - F_{i-1}}{f_i} \cdot t_i$$

$$Me = 20 + \frac{15,5 - 9}{7} \cdot 10$$

$$Me = 20 + 9,29$$

$$\boxed{Me = 29,285}$$

Recuerda:

Li-I :es el límite inferior de la clase donde se encuentra la mediana, en este caso el límite inferior es 20.

N / 2 :es la semisuma de las frecuencias absolutas, en este caso es 15,5.

Fi-I :es la frecuencia acumulada anterior a la clase mediana, en este caso es 9. fi : es la frecuencia absoluta del intervalo mediano, en este caso es 7 ti :es la amplitud de los intervalos. Se calcula restando el extremo superior menos el inferior del intervalo, en este caso es:

30 - 20 = 10

3° Calculemos la moda Mo :

Lo primero que debemos hacer es identificar el intervalo modal:

Edad	Marca clase (X_i)	Frecuencia absoluta (f_i)	Frecuencia acumulada (F_i)
[0 - 10)	5	3	3
[10 - 20)	15	6	9
[20 - 30)	25	7	16
[30 - 40)	35	12	28
[40 - 50)	45	3	31

N = 31

Intervalo modal: mayor frecuencia absoluta

Ahora podemos reemplazar los datos en la fórmula:

$$Mo = L_i + \frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})} \cdot t_i$$

$$Mo = 30 + \frac{12 - 7}{(12 - 7) + (12 - 3)} \cdot 10$$

$$Mo = 30 + 3{,}57$$

$$\boxed{Mo = 33{,}6}$$

- Si la moda está en el primer intervalo, entonces fi-1= 0. Si la moda está en el último intervalo, entonces fi+1= 0.

- Puede haber más de una moda en el caso en que dos o más valores de la variable presenten la misma frecuencia.

1. 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.

2. CUANTILES

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

Dónde:

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:

LI = limite inferior de la clase que lo contiene

P = valor que representa la posición de la medida

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

Fa-I = 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:

LI = limite inferior de la clase que lo contiene

P = valor que representa la posición de la medida

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

Fa-I = 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:

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

Donde:

LI = limite inferior de la clase que lo contiene

P = valor que representa la posición de la medida

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

Fa-I = 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:

$$\frac{3(n+1)}{4}$$

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

Dónde:

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 = límite 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:

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

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

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

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.

Coeficiente de Pearson

La covariación es el grado de concordancia de las posiciones relativas de los datos de dos variables. En consecuencia el coeficiente de correlación de Pearson opera con puntuaciones tipificadas (que miden posiciones relativas) y se define:

$$r_{xy} = \frac{\sum z_x z_y}{N}$$

El fundamento del coeficiente de Pearson es el siguiente: Cuanto más intensa sea la concordancia (en sentido directo o inverso) de las posiciones relativas de los datos en las dos variables, el producto del numerador toma mayor valor (en sentido absoluto). Si la concordancia es exacta, el numerador es igual a N (o a -N), y el índice toma un valor igual a 1 (o -1).

Ejemplo 1 (Máxima covariación positiva)

X	Y	z_x	z_y
2	4	-1.41	-1.41
3	5	-0.71	-0.71
4	6	-0.00	-0.00
5	7	0.71	0.71
6	8	1.41	1.41

Observa que los datos tipificados (expresados como puntuaciones z) en las dos columnas de la derecha tienen los mismos valores en ambas variables, dado que las posiciones relativas son las mismas en las variables X e Y.

Si obtenemos los productos de los valores tipificados para cada caso, el resultado es:

X	Y	z_x	z_y	$z_x z_y$
2	4	-1.41	-1.41	2
3	5	-0.71	-0.71	0.5
4	6	0.00	0.00	0
5	7	0.71	0.71	0.5
6	8	1.41	1.41	2
				$\Sigma = 5$

El cociente de dividir la suma de productos (5) por N (hay que tener en cuenta que N es el número de casos, NO el número de datos) es igual a 1:

$$\boxed{= \frac{\sum z_x z_y}{N} = \frac{5}{5} = 1}$$

Ejemplo 2 (Covariación positiva de alta intensidad)

X	Y	z_x	z_y	$z_x z_y$
2	5	-1.41	-0.71	1
3	4	-0.71	-1.41	1
4	6	0.00	0.00	0
5	8	0.71	1.41	1
6	7	1.41	0.71	1
				$\Sigma = 4$

Y por tanto,

$$\boxed{= \frac{\sum z_x z_y}{N} = \frac{4}{5} = 0.8}$$

Ejemplo 3 (Ausencia de covariación)

X	Y	Zx	Zy	Zx Zy
2	7	-1.41	0.71	-1
3	8	-0,71	1,41	-1
4	6	0,00	0,00	0
5	4	0,71	-1,41	-1
6	5	1,41	-0,71	-1
				Σ= -4

$$\boxed{=\frac{\sum z_x z_y}{N}=\frac{-4}{5}=-0.8}$$

Ejemplo 4 (Covariación negativa de alta intensidad)

X	Y	Zx	Zy	Zx Zy
2	7	-1.41	0.71	-1
3	8	-0,71	1,41	-1
4	6	0,00	0,00	0
5	4	0,71	-1,41	-1
6	5	1,41	-0,71	-1
				Σ= -4

$$\boxed{=\frac{\sum z_x z_y}{N}=\frac{-4}{5}=-0.8}$$

Ejemplo 5 (Máxima covariación negativa)

X	Y	Zx	Zy	Zx Zy
2	8	-1.41	1.41	-2.0
3	7	-0.71	0.71	-0.5
4	6	0.00	0.00	0.0
5	5	0.71	-0.71	-0.5
6	4	1.41	-1.41	-2.0
				Σ= -5

$$\boxed{=\frac{\sum z_x z_y}{N-1}=\frac{-5}{5}=-1}$$

El valor de la correlación es igual a 1 o -1 si la covariación es de intensidad máxima, y se va acercando hacia el 0 cuanto más pequeña sea la intensidad de la covariación. Además, el índice tiene signo positivo cuando la covariación es directa y negativa cuando es inversa. (Los ejemplos anteriores los puedes practicar con otros datos -pero de la misma escala- si estás leyendo el tema en ordenador, para la que cual hay que clicar dos veces en la siguiente imagen)

	X	Y	z_x	z_y	$z_x z_y$		
	2	7	-1,41	0,71	-1,00		
	3	4	-0,71	-1,41	1,00	r_{XY}=	0
	4	6	0,00	0,00	0,00		
	5	8	0,71	1,41	1,00		
	6	5	1,41	-0,71	-1,00		
				Σ=	0		
M=	4	6					
DT=	1,41	1,41					

Características

a) El coeficiente de correlación de Pearson puede tomar valores entre -1 y 1. b) La correlación de una variable con ella misma siempre es igual a 1.

c) El valor 0 indica ausencia de covariación lineal, pero NO si la covariación es de tipo no lineal

Unidad 3

3.1.- CONCEPTOS DE PROBABILIDAD

Muchos fenómenos de la naturaleza, como la caída libre de un cuerpo en la superficie terrestre, pueden predecirse mediante leyes deterministas. Otros, en cambio, se rigen por el azar, aun cuando se produzcan siempre en unas mismas condiciones. Por ejemplo, ¿qué número saldrá al lanzar un dado? Los sucesos que obedecen al azar se denominan aleatorios o estocásticos, y su comportamiento se estudia a través del cálculo de probabilidades.

Sucesos estocásticos

Por definición, se llama experimento aleatorio, estocástico o estadístico al que puede producir resultados diferentes en unas mismas condiciones. Lanzar una moneda al aire o tirar un dado son ejemplos comunes de experimentos aleatorios.

Cada uno de los resultados de un experimento aleatorio se llama suceso elemental, y el conjunto de todos los sucesos elementales distintos que pueden producirse en el experimento se denomina espacio muestral. Por ejemplo, el espacio muestral de un experimento aleatorio consistente en tirar un dado es E = {1, 2, 3, 4, 5, 6}.

Los diversos subconjuntos del espacio muestral se denominan sucesos estocásticos. Algunos tipos especiales de sucesos estocásticos son:

- Suceso seguro, que se produce siempre. Matemáticamente, corresponde al espacio muestral E.
- Suceso imposible (espacio vacío) de E.

- Suceso contrario o complementario de uno dado (si se produce el suceso A, su complementario no ocurre, y a la inversa). Matemáticamente: = E - A.
- Dos sucesos estocásticos con algún suceso elemental común se dicen compatibles; en caso contrario, se llaman incompatibles.

Operaciones con sucesos

En el espacio muestral E de sucesos estocásticos pueden definirse varias operaciones:

- Unión de sucesos A y B, un suceso estocástico que contiene todos los sucesos = E).
- Intersección de sucesos A y B, que comprende sólo los sucesos elementales
- Diferencia de sucesos A y B, que es un nuevo suceso formado por los sucesos elementales de A que no lo son de B. Se escribe A - B.
- Implicación de sucesos. Se dice que un suceso estocástico A implica a otro B

Propiedades	OPERACIONES	
	Unión	Intersección
Conmutativa	$A \cup B = B \cup A$	$A \cap B = B \cap A$
Asociativa	$(A \cup B) \cup C = A \cup (B \cup C)$	$(A \cap B) \cap C = A \cap (B \cap C)$
Distributiva	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
Complementaria	$A \cup \overline{A} = E$	$A \cap \overline{A} = \varnothing$
Leyes de Morgan	$\overline{(A \cup B)} = \overline{A} \cap \overline{B}$	$\overline{(A \cap B)} = \overline{A} \cup \overline{B}$

Propiedades de las operaciones con sucesos.

Probabilidad de un suceso

Dado un experimento aleatorio, se denomina probabilidad a una función que asigna a cada suceso estocástico un número que refleja el tanto por uno de veces que ocurre el suceso dentro del experimento. Por tanto, el valor de la probabilidad indica la frecuencia relativa de cada suceso estocástico dentro del experimento aleatorio.

La función probabilidad expresa como P (A), y se distingue por las siguientes características:

- La probabilidad del suceso seguro es 1: P (E) = 1.
- La probabilidad de cualquier suces

Propiedades de la función probabilidad

Otras propiedades interesantes de la función probabilidad son las siguientes:

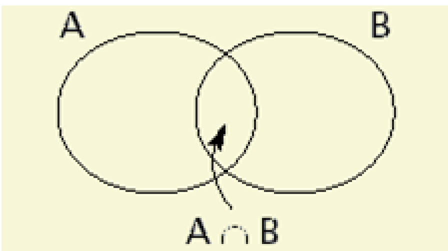
- Dado un conjunto de sucesos incompatibles dos a dos, la probabilidad de su unión es igual a la suma de las probabilidades de cada suceso individual:

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$$

- La probabilidad del suceso contrario viene dada por P ( ) = 1 - P (A).

- Para dos sucesos compatibles, se cumple que la probabilidad de la unión es igual a la suma de las probabilidades de cada suceso menos la probabilidad del suceso intersección entre ambos:

P (A B) = P (A) + P (B) - P (A B)



Representación gráfica de la unión y la intersección de dos sucesos compatibles.

- Regla de Laplace, según la cual la probabilidad de un suceso estocástico formado por h sucesos elementales equiprobables en un espacio muestral de n elementos se determina como el cociente entre el número de casos favorables (h) y el número de casos posibles (n).

$$P(A) = \frac{N^{o} \text{ casos favorables (h)}}{N^{o} \text{ casos posibles (n)}}$$

3.2.- Leyes de la probabilidad

La probabilidad es un método por el cual se obtiene la frecuencia de un suceso determinado mediante la realización de un experimento aleatorio, del que se conocen todos los resultados posibles, bajo condiciones suficientemente estables. La teoría de la probabilidad se usa extensamente en áreas como la estadística, la física, las matemáticas, las ciencias y la filosofía para sacar conclusiones sobre la probabilidad discreta de sucesos potenciales y la mecánica subyacente discreta de sistemas complejos.

La probabilidad constituye un importante parámetro en la determinación de las diversas casualidades obtenidas tras una serie de eventos esperados dentro de un rango estadístico.

Existen diversas formas como método abstracto, como la teoría de Dempster y la teoría de la relatividad numérica, esta última con un alto grado de aceptación si se toma en cuenta que disminuye considerablemente las posibilidades hasta un nivel mínimo ya que somete a todas las antiguas reglas a una simple ley de relatividad.

La probabilidad de un evento se denota con la letra p y se expresa en términos de una fracción y no en porcentajes, por lo que el valor de p cae entre 0 y 1. Por otra parte, la probabilidad de que un evento "no ocurra" equivale a 1 menos el valor de p y se denota con la letra q

$$P(Q) = 1 - P(E)$$

Los tres métodos para calcular las probabilidades son la regla de la adición, la regla de la multiplicación.

Regla de la adición

La regla de la adición o regla de la suma establece que la probabilidad de ocurrencia de cualquier evento en particular es igual a la suma de las probabilidades individuales, si es que los eventos son mutuamente excluyentes, es decir, que dos no pueden ocurrir al mismo tiempo.

P(A o B) = P(A) U P(B) = P(A) + P(B) si A y B son mutuamente excluyente. P(A o B) = P(A) + P(B) − P(A y B) si A y B son no excluyentes.

Siendo: P(A) = probabilidad de ocurrencia del evento A. P(B) = probabilidad de ocurrencia del evento B. P(A y B) = probabilidad de ocurrencia simultánea de los eventos A y B.

Regla de la multiplicación

La regla de la multiplicación establece que la probabilidad de ocurrencia de dos o más eventos estadísticamente independientes es igual al producto de sus probabilidades individuales.

P(A y B) = P(A B) = P(A)P(B) si A y B son independientes. P(A y B) = P(A B) = P(A)P(B|A)

Si A y B son dependientes

La regla de Laplace 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) = N° de casos favorables / N° de resultados posibles

Esto significa que: la probabilidad del evento A es igual al cociente del número de casos favorables (los casos dónde sucede A) sobre el total de casos posibles. -

3.3.-

-Aplicación: Dos aplicaciones principales de la teoría de la probabilidad en el día a día son en el análisis de riesgo y en el comercio de los mercados de materias. Los gobiernos normalmente aplican métodos probabilísticos en regulación ambiental donde se les llama "análisis de vías de dispersión", y a menudo miden el bienestar usando métodos que son estocásticos por naturaleza, y escogen qué proyectos emprender basándose en análisis estadísticos de su probable efecto en la población como un conjunto. No es correcto decir que la estadística está incluida en el propio modelado, ya que típicamente los análisis de riesgo son para una única vez y por lo tanto requieren más modelos de probabilidad fundamentales, por ej. “la probabilidad de otro 11-S”. Una ley de números pequeños tiende a aplicarse a todas aquellas elecciones y percepciones del efecto de estas elecciones, lo que hace de las medidas probabilísticas un tema político.

Un buen ejemplo es el efecto de la probabilidad percibida de cualquier conflicto generalizado sobre los precios del petróleo en Oriente Medio – que producen un efecto dominó en la economía en conjunto. Un cálculo por un mercado de materias primas en que la guerra es más probable en contra de menos probable probablemente envía los precios hacia arriba o hacia abajo e indica a otros comerciantes esa opinión. Por consiguiente, las probabilidades no se calculan independientemente y tampoco son necesariamente muy racionales. La teoría de las fianzas conductuales surgió para describir el efecto de este pensamiento de grupo en el precio, en la política, y en la paz y en los conflictos.

Se puede decir razonablemente que el descubrimiento de métodos rigurosos para calcular y combinar los cálculos de probabilidad ha tenido un profundo efecto en la sociedad moderna. Por consiguiente, puede ser de alguna importancia para la mayoría de los ciudadanos entender cómo se calculan los pronósticos y las probabilidades, y cómo contribuyen a la reputación y a las decisiones, especialmente en una democracia.

Otra aplicación significativa de la teoría de la probabilidad en el día a día es en la fiabilidad. Muchos bienes de consumo, como los automóviles y la electrónica de consumo, utilizan la teoría de la fiabilidad en el diseño del producto para reducir la probabilidad de avería. La probabilidad de avería también está estrechamente relacionada con la

garantía del producto.

Se puede decir que no existe una cosa llamada probabilidad. También se puede decir que la probabilidad es la medida de nuestro grado de incertidumbre, o esto es, el grado de nuestra ignorancia dada una situación. Por consiguiente, puede haber una probabilidad de

1 entre 52 de que la primera carta en un baraja sea la J de diamantes. Sin embargo, si uno mira la primera carta y la reemplaza, entonces la probabilidad es o bien 100% ó 0%, y la elección correcta puede ser hecha con precisión por el que ve la carta. La física moderna proporciona ejemplos importantes de situaciones determinísticas donde sólo la descripción probabilística es factible debido a información incompleta y la complejidad de un sistema así como ejemplos de fenómenos realmente aleatorios-

3.4.- Un diagrama de árbol es una herramienta que se utiliza para determinar todos los posibles resultados de un experimento aleatorio. En el cálculo de la probabilidad se requiere conocer el número de elementos que forman parte del espacio maestral, estos se pueden determinar con la construcción del diagrama de árbol.

El diagrama de árbol es una representación gráfica de los posibles resultados del experimento, el cual consta una serie de pasos, donde cada uno de los pasos tiene un número finito de maneras de ser llevado a cabo. Se utiliza en los problemas de conteo y probabilidad. Para la construcción de un diagrama en árbol se partirá poniendo una rama para cada una de las posibilidades, acompañada de su probabilidad. Cada una de esta ramas se conoce como rama de primera generación En el final de cada rama de primera generación se constituye a su vez, un nudo del cual parten nuevas ramas conocidas como ramas de segunda generación, según las posibilidades del siguiente paso, salvo si el nudo representa un posible final del experimento (nudo final). Hay que tener en cuenta que la construcción de un árbol no depende de tener el mismo número de ramas de segunda generación que salen de cada rama de primera generación y que la suma de probabilidades de las ramas de cada nudo ha de dar 1. Existe un principio sencillo de los diagramas de árbol que hace que éstos sean mucho más útiles para los cálculos rápidos de probabilidad: multiplicamos las probabilidades si se trata de ramas adyacentes (contiguas), el ejemplo de alumna de la primera facultad, o bien las sumamos si se trata de ramas separadas que emergen de un mismo punto, el ejemplo de encontrar un alumno.

¿Para qué sirve?

Un diagrama de árbol es un método gráfico para identificar todas las partes necesarias Para alcanzar algún objetivo final. En mejora de la calidad, los diagramas de árbol se utilizan generalmente para identificar todas las tareas necesarias para implantar una solución.

Se emplea para descomponer una meta u objetivo en una serie de actividades que deban o puedan hacerse. A través de la representación gráfica de actividades se facilita el entendimiento de las acciones que intervendrán. Permite a los miembros del equipo de trabajo expandir su pensamiento al crear soluciones sin perder de vista el objetivo principal o los objetivos secundarios. Ubica al equipo para que se dirija a situaciones reales versus teóricas. Asimismo, se dimensiona el nivel real de complejidad de algún proyecto y se puede prever el encontrarse con soluciones inviables antes de arranque.

¿Cómo se elabora?

- Establezca el objetivo que se analizará a través del diagrama de árbol. Es muy importante que el objetivo quede claro para todos y que Este expresado de manera activa.

- Arme el equipo adecuado. Se sugiere un equipo de 4 a 8 participantes.

Considere que aquellos que seleccione deberán estar involucrados en la problemática a fondo para aportar soluciones y que el diagrama de árbol cuente así con los niveles de análisis necesarios.

- Genere el mayor número posible de "cabeceras del diagrama de árbol" Esto es las ideas o sub-objetivos hacia los que se enfocarán las acciones para lograr el objetivo principal.

- Detenga la descomposición de temas cuando ya se perfilen tareas específicas a realizarse.

- Revise el diagrama de árbol. Asegúrese de que tiene un flujo lógico y que esté lo más completo posible.

- Pregunte al equipo si observa algún punto que sea muy obvio y se haya olvidado incluir.

- Pregúntese junto con el equipo si las tareas resultantes son necesarias para lograr el objetivo.

Ventajas

- Exhorta a los integrantes del equipo a ampliar su modo de pensar al crear soluciones.

- Mantiene a todo el equipo vinculado a las metas y sub metas generales de una tarea.

- Mueve al equipo de planificación de la teoría al mundo real.

Beneficios

- Permite obtener una visión de conjunto del objeto de estudio.
- Permite identificar los medios necesarios para alcanzar una meta o resolver un problema.
- Permite identificar las causas primarias y secundarias de un problema y asignar prioridades al momento de resolver un problema.
- Permite entender la relación causa – efecto de los problemas.
- Permite identificar los objetivos las metas de cada tarea.

Relación con otras herramientas

- Diagrama de afinidad.
- Diagrama de causa – efecto.
- Requisito de la norma ISO 9001:2000

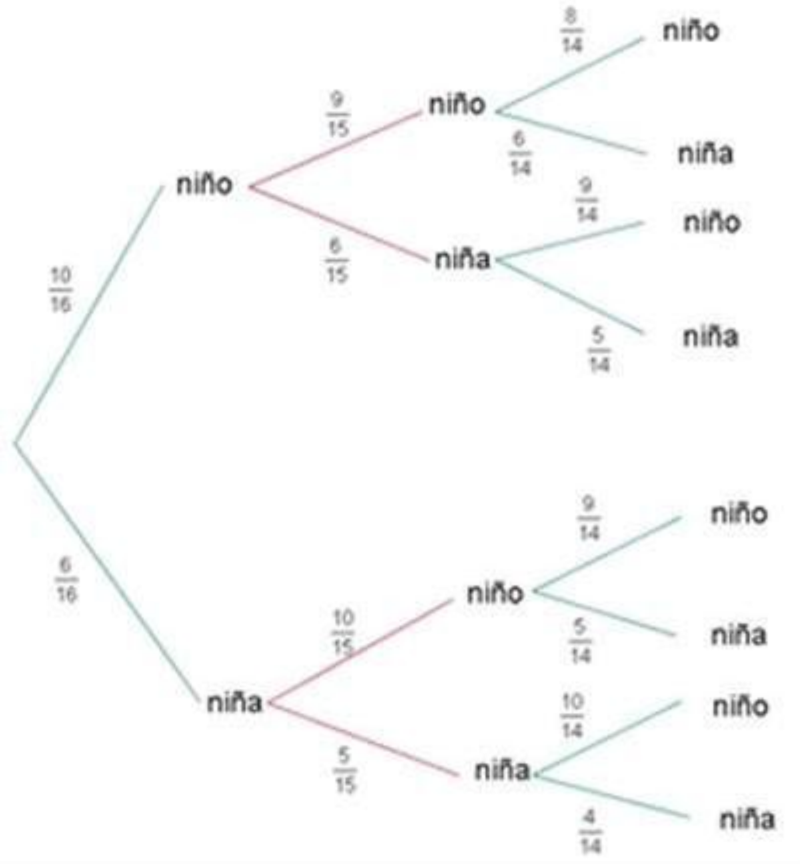
Para la construcción de un diagrama en árbol se partirá poniendo una rama para cada una de las posibilidades, acompañada de su probabilidad.

En el final de cada rama parcial se constituye a su vez, un nudo del cual parten nuevas ramas, según las posibilidades del siguiente paso, salvo si el nudo representa un posible final del experimento (nudo final).

Hay que tener en cuenta: que la suma de probabilidad es de las ramas de cada nudo ha de dar.

Ejemplos

I Una clase consta de seis niñas y 10 niños. Si se escoge un comité de tres al azar, hallar la probabilidad de:



I Seleccionar tres niños.

Son sucesos independientes

$$p(3\,niños) = \frac{10}{16} \cdot \frac{9}{15} \cdot \frac{8}{14} = 0.214$$

2. Seleccionar exactamente dos niños y una niña.

$$p(2\,niños\ y\ 1\ niña) = \frac{10}{16} \cdot \frac{9}{15} \cdot \frac{6}{14} + \frac{10}{16} \cdot \frac{6}{15} \cdot \frac{9}{14} + \frac{6}{16} \cdot \frac{10}{15} \cdot \frac{9}{14} = 0.482$$

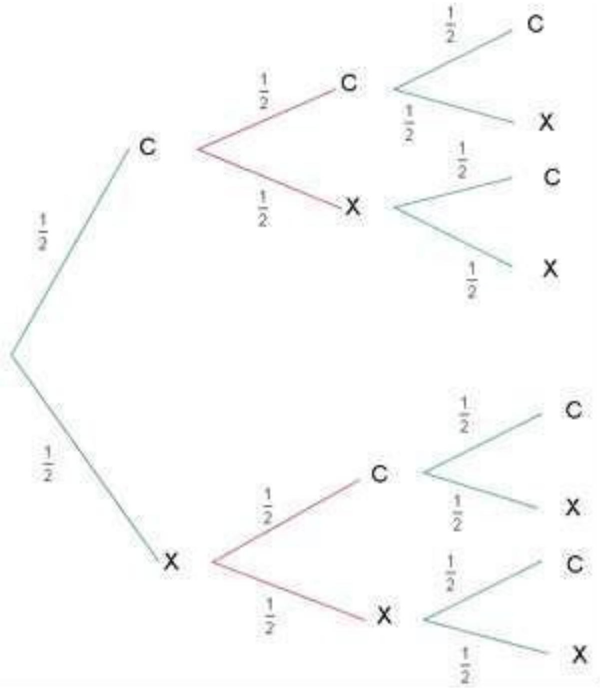
3. Seleccionar exactamente dos niñas y un niño.

$$p\left(2\,niñas\ y\ 1\ niño\right) = \frac{10}{16}\cdot\frac{6}{15}\cdot\frac{5}{14}+\frac{6}{16}\cdot\frac{10}{15}\cdot\frac{5}{14}+\frac{6}{16}\cdot\frac{5}{15}\cdot\frac{10}{14} = 0.268$$

l. Seleccionar tres niñas.

$$p\left(3\,niñas\right) = \frac{6}{16}\cdot\frac{5}{15}\cdot\frac{4}{14} = 0.0357$$

l Calcular la probabilidad de que al arrojar al aire tres monedas, salgan:



l Tres caras.

$$p\left(3c\right) = \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} = \frac{1}{8}$$

Experimentos compuestos

Un experimento compuesto es aquel que consta de dos o más experimentos aleatorios simples.

Es decir, si tiramos un dado, o una moneda, son experimentos aleatorios simples, pero si realizamos el experimento de tirar un dado y posteriormente una moneda, estamos realizando un experimento compuesto.

En los experimentos compuestos es conveniente usar el llamado diagrama en árbol para hacerse una idea global de todos ellos.

3.4.- Teorema de Bayes

El teorema de Bayes es utilizado para calcular la probabilidad de un suceso, teniendo información de antemano sobre ese suceso.

Podemos calcular la probabilidad de un suceso A, sabiendo además que ese A cumple cierta característica que condiciona su probabilidad. El teorema de Bayes entiende la probabilidad de forma inversa al teorema de la probabilidad total. El teorema de la probabilidad total hace inferencia sobre un suceso B, a partir de los resultados de los sucesos A. Por su parte, Bayes calcula la probabilidad de A condicionado a B.

El teorema de Bayes ha sido muy cuestionado. Lo cual se ha debido, principalmente, a su mala aplicación. Ya que, mientras se cumplan los supuestos de sucesos disjuntos y exhaustivos, el teorema es totalmente válido.

Fórmula del teorema de Bayes

Para calcular la probabilidad tal como la definió Bayes en este tipo de sucesos, necesitamos una fórmula. La fórmula se define matemáticamente como:

$$P[A_n/B] = \frac{P[B/A_n] \cdot P[A_n]}{\sum P[B/A_i] \cdot P[A_i]}$$

Donde B es el suceso sobre el que tenemos información previa y A(n) son los distintos sucesos condicionados. En la parte del numerador tenemos la probabilidad condicionada, y en la parte de abajo la probabilidad total. En cualquier caso, aunque la fórmula parezca un poco abstracta, es muy sencilla. Para demostrarlo, utilizaremos un ejemplo en el que en lugar de A(1), A(2) y A(3), utilizaremos directamente A, B y C.

Ejemplo del teorema de Bayes

Una empresa tiene una fábrica en Estados Unidos que dispone de tres máquinas A, B y C, que producen envases para botellas de agua. Se sabe que la máquina A produce un 40% de la cantidad total, la máquina B un 30%, y la máquina C un 20%. También se sabe que cada máquina produce envases defectuosos. De tal manera que la máquina A produce un 2% de envases defectuosos sobre el total de su producción, la máquina B un 3%, y la máquina C un 5%. Dicho esto, se plantean dos cuestiones:

P(A) = 0,40 P(D/A) = 0,02

P(B) = 0,30   P(D/B) = 0,03

P(C) = 0,30   P(D/C) = 0,05

Si un envase ha sido fabricado por la fábrica de esta empresa en Estados Unidos ¿Cuál es la probabilidad de que sea defectuoso?

Se calcula la probabilidad total. Ya que, a partir los diferentes sucesos, calculamos la probabilidad de que sea defectuoso.

P(D) =[P(A) x P(D/A)] + [P(B) x P(B/A)] + [P(C) x P(C/A)] = [0,4 x 0,02] + [

0,3 x

0,03] + [0,3 x 0,05] = 0,032

Expresado en porcentaje, diríamos que la probabilidad de que un envase fabricado por la fábrica de esta empresa en Estados Unidos sea defectuoso es del 3,2%.

2. Siguiendo con la pregunta anterior, si se adquiere un envase y este es defectuoso ¿Cuáles es la probabilidad de que haya sido fabricado por la máquina A?¿Y por la máquina B?¿Y por la máquina C?

Aquí se utiliza el teorema de Bayes. Tenemos información previa, es decir, sabemos que el envase es defectuoso. Claro que, sabiendo que es defectuoso, queremos saber cuál es la probabilidad de que se haya producido por una de las máquinas.

P(A/D) = [P(A) x P(D/A)] / P(D) = [0,40 x 0,02] / 0,032 = 0,25

$P(B/D) = [P(B) \times P(D/B)] / P(D) = [0{,}30 \times 0{,}03] / 0{,}032 = 0{,}28$

$P(C/D) = [P(C) \times P(D/C)] / P(D) = [0{,}30 \times 0{,}05] / 0{,}032 = 0{,}47$

Sabiendo que un envase es defectuoso, la probabilidad de que haya sido producido por la máquina A es del 25%, de que haya sido producido por la máquina B es del 28% y de que haya sido producido por la máquina C es del 47%

3.6.- DEFINICION DE CONJUNTO

La palabra conjunto generalmente la asociamos con la idea de agrupar objetos, por ejemplo un conjunto de discos, de libros, de plantas de cultivo y en otras ocasiones en palabras como hato, rebaño, piara, parcelas, campesinado, familia, etc., es decir la palabra conjunto denota una colección de elementos claramente entre sí, que guardan alguna característica en común. Ya sean números, personas, figuras, ideas y conceptos.

En matemáticas el concepto de conjunto es considerado primitivo y ni se da una definición de este, sino que se trabaja con la notación de colección y agrupamiento de objetos, lo mismo puede decirse que se consideren primitivas las ideas de elemento y pertenencia.

La característica esencial de un conjunto es la de estar bien definido, es decir que dado un objeto particular, determinar si este pertenece o no al conjunto. Por ejemplo si se considera el conjunto de los números dígitos, sabemos que el 3 pertenece al conjunto, pero el 19 no. Por otro lado el conjunto de las bellas obras musicales no es un conjunto bien definido, puesto que diferentes personas puedan incluir distintas obras en el conjunto.

Los objetos que forman un conjunto son llamados miembros o elementos. Por ejemplo el conjunto de las letras de alfabeto; a, b, c, ..., x, y, z. que se puede escribir así:

{ a, b, c, ..., x, y, z}

Como se muestra el conjunto se escribe entre llaves ({}) , o separados por comas (,). El detallar a todos los elementos de un conjunto entre las llaves, se denomina forma tabular, extensión o enumeración de los elementos.

Dos conjuntos son iguales si tienen los mismos elementos, por ejemplo: El conjunto { a, b, c } también puede escribirse:

{ a, c, b }, { b, a, c }, { b, c, a }, { c, a, b }, { c, b, a }

En teoría de conjuntos se acostumbra no repetir a los elementos por ejemplo: El conjunto { b, b, b, d, d } simplemente será { b, d }.

MEMBRESIA

Los conjuntos se denotan por letras mayúsculas: A, B, C,... por ejemplo: A={ a, c, b }

B={ primavera, verano, otoño, invierno }

El símbolo indicará que un elemento pertenece o es miembro de un conjunto. Por el contrario para indicar que un elemento no pertenece al conjunto de referencia, bastará cancelarlo con una raya inclinada / quedando el símbolo como.

Ejemplo:

Sea B={ a, e, i, o, u }, a □□B y c □ B

SUBCONJUNTO

Sean los conjuntos A={ 0, 1, 2, 3, 5, 8 } y B={ 1, 2, 5 }

En este caso decimos que B está contenido en A, o que B es subconjunto de A. En general si A y B son dos conjuntos cualesquiera, decimos que B es un subconjunto de A si todo elemento de B lo es de A también.

Por lo tanto si B es un subconjunto de A se escribe B □ A. Si B no es subconjunto de A se indicará con una diagonal □ .

Note que □ se utiliza solo para elementos de un conjunto y□ solo para conjuntos.

UNIVERSO O CONJUNTO UNIVERSAL

El conjunto que contiene a todos los elementos a los que se hace referencia recibe el nombre de conjunto Universal, este conjunto depende del problema que se estudia, se denota con la letra U y algunas veces con la letra S (espacio muestral).

Por ejemplo si solo queremos referirnos a los 5 primeros números naturales el conjunto queda:

U={ 1, 2, 3, 4, 5 }

Forma alternativa para indicar conjuntos de gran importancia:

- Conjunto de números naturales (enteros mayores que cero) representados por la letra N donde

N={ 1, 2, 3, ... }

- Conjunto de números enteros positivos y negativos representados por la letra Z donde

Z={..., -2, -1, 0, 1, 2, ... }

- Conjunto de números racionales (números que se representan como el cociente de dos números enteros {fracciones }). Estos números se representan por una Q
- Conjunto de números irracionales (números que no puedan representarse como el cociente de dos números enteros) representados por la letra I.
- Conjunto de los números reales que son los números racionales e irracionales es decir todos, representados por R.

Todos estos conjuntos tienen un número infinito de elementos, la forma de simbolizarlos por extensión o por enumeración es de gran utilidad cuando los conjuntos a los que se hace referencia tienen pocos elementos para poder trabajar con ellos se emplean la notación llamada comprensión.

Por ejemplo, la denotar el conjunto de los números naturales menores que 60. Aquí U es el conjunto N y se tiene una propiedad que caracteriza a los elementos del conjunto: ser menores que 60.

Para indicar esta situación empleamos la simbología del álgebra de conjuntos:

En esta expresión se maneja un conjunto de x que pertenece a los números naturales (N) y además que los valores de x son menores que 60.

Ahora si se desea trabajar con conjuntos que manejen intervalos estos pueden ser representados por medio de una expresión algebraica; supongamos que se desea expresar los números enteros (Z) entre -20 y 30 el conjunto quedaría de la manera siguiente:

x/x □ Z ; 20 □ x □ 30 }

También se puede expresar el valor de un conjunto indicando la pertenencia o no pertenencia a uno diferente, por ejemplo L={ 1, 3, 4, 6, 9 }

P={ x/x □ N ; X □ L }

En el conjunto P se indica que los elementos x de un conjunto pertenecen a los números naturales y además x no pertenece al conjunto L.

OPERACIONES CON CONJUNTOS UNION

La unión de dos conjuntos A y B la denotaremos por A □ B y es el conjunto formado por los elementos que pertenecen al menos a uno de ellos ó a los dos. Lo que se denota por:

A □ B = { x/x □ A ó x □ B }

Ejemplo: Sean los conjuntos A={ 1, 3, 5, 7, 9 } y B={ 10, 11, 12 }

A □ B ={ 1, 3, 5, 7, 9, 10, 11, 12 }

INTERSECCION

Sean A={ 1, 2, 3, 4, 5, 6, 8, 9 } y B={ 2, 4, 8, 12 }

Los elementos comunes a los dos conjuntos son: { 2, 4, 8 }. A este conjunto se le llama intersección de A y B; y se denota por A □ B, algebraicamente se escribe así:

A □ B = { x/x □ A y x □ B }

Y se lee el conjunto de elementos x que están en A y están en B.

Ejemplo:

Sean Q={ a, n, p, y, q, s, r, o, b, k } y P={ l, u, a, o, s, r, b, v, y, z }

Q □ P={ a, b, o, r, s, y }

CONJUNTO VACIO

Un conjunto que no tiene elementos es llamado conjunto vacío ó conjunto nulo lo que denotamos por el símbolo □ .

Por ejemplo:

Sean A={ 2, 4, 6 } y B={ 1, 3, 5, 7 } encontrar A □ B.

A □ B= { }

COMPLEMENTO

El complemento de un conjunto respecto al universo U es el conjunto de elementos de U que no pertenecen a A y se denota como A' y que se representa por comprensión como:

A'={ x □ U/x y x □ A }

Ejemplo:

Sea U = { 1, 2, 3, 4, 5, 6, 7, 8, 9 }

A= { 1, 3, 5, 7, 9 } donde A □ U

El complemento de A estará dado por: A'= { 2, 4, 6, 8 }

DIFERENCIA

Sean A y B dos conjuntos. La diferencia de A y B se denota por A-B y es el conjunto de los elementos de A que no están en B y se representa por comprensión como:

A - B={ x/x □ A ; X □ B }

Ejemplo:

Sea A= { a, b, c, d } y

B= { a, b, c, g, h, i }

A - B= { d }

En el ejemplo anterior se observa que solo interesan los elementos del conjunto A que no estén en B. Si la operación fuera B - A el resultado es

B – A = { g, h, i }

E indica los elementos que están en B y no en A.

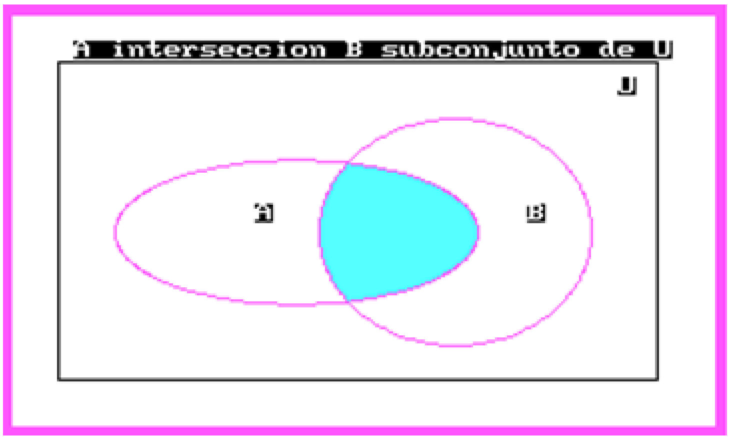
DIAGRAMAS DE VENN

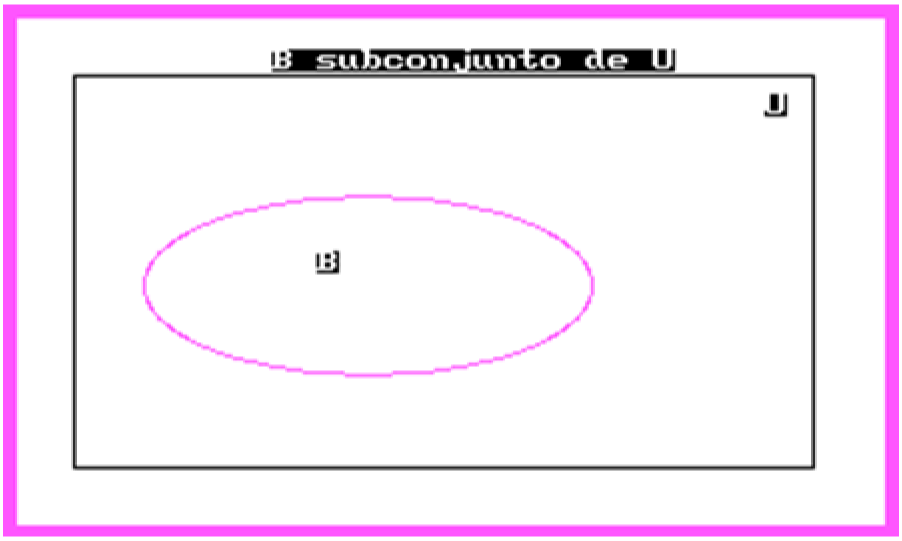
Los diagramas de Venn que deben al filósofo inglés John Venn (1834-1883) sirven para encontrar relaciones entre conjuntos de manera gráfica mediante dibujos o diagramas.

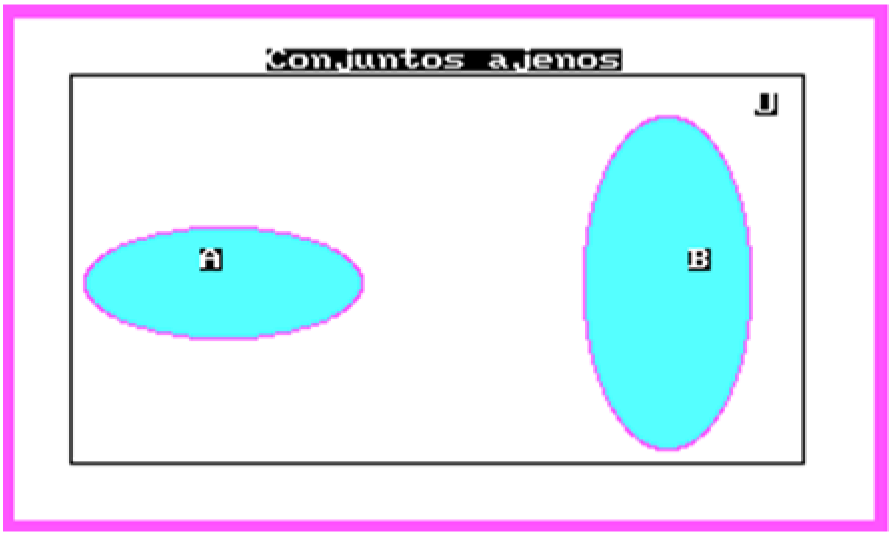
La manera de representar el conjunto Universal es un rectángulo, o bien la hoja de papel con que se trabaje.

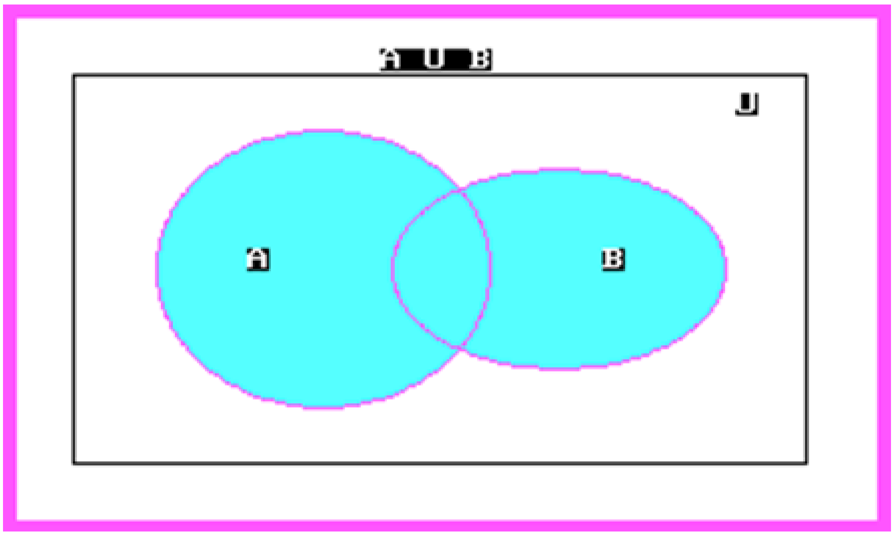
Un ejemplo de la representación del conjunto universal se muestra como:

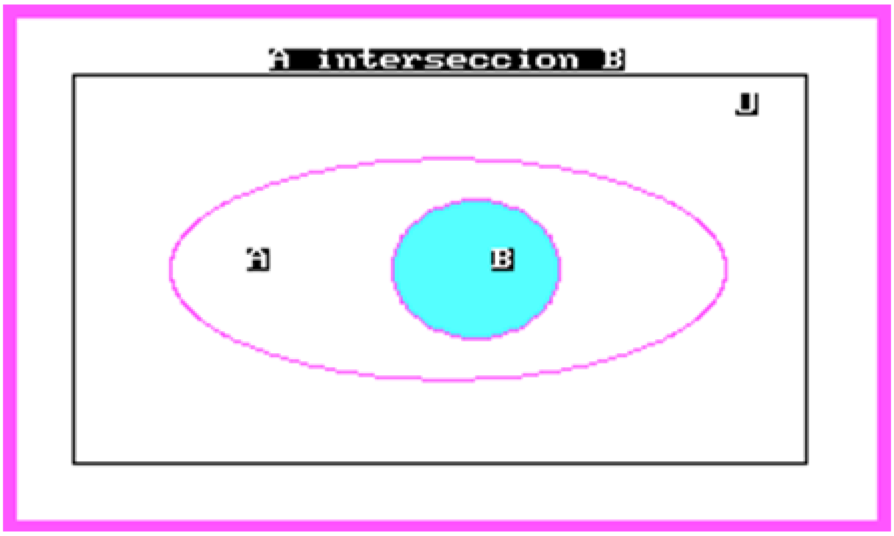
Los conjuntos se representan por medio de dibujos dentro del rectángulo, los aspectos de interés se resaltan sombreando las áreas respectivas. En el caso de este curso las indicaremos por medio de un color azul por ejemplo:

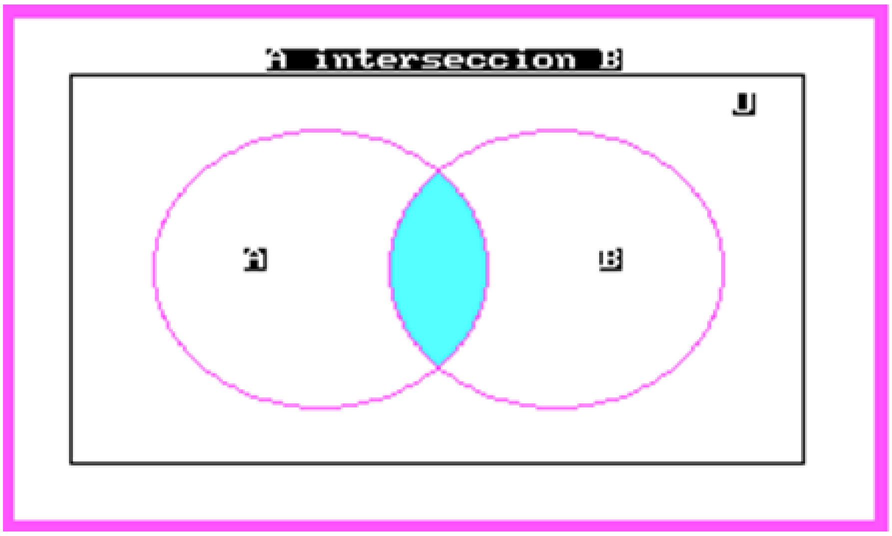


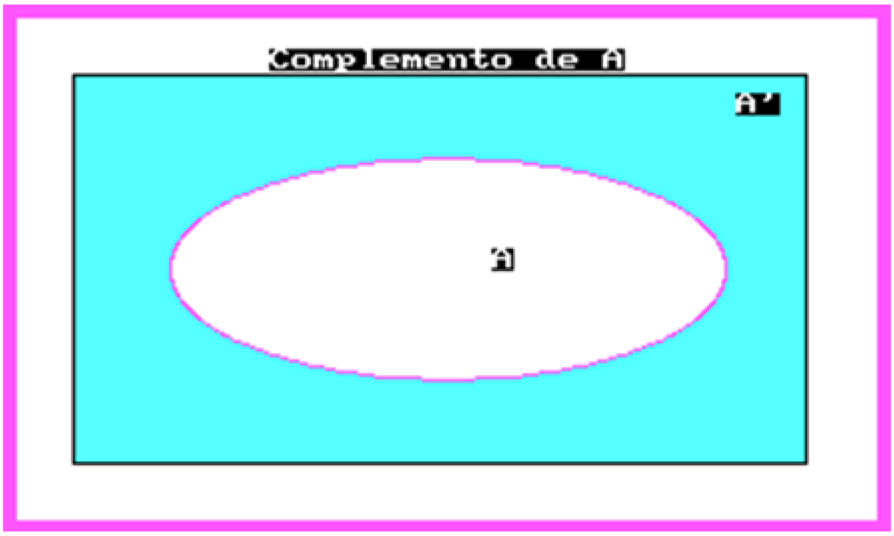

B subconjunto de U
U
B

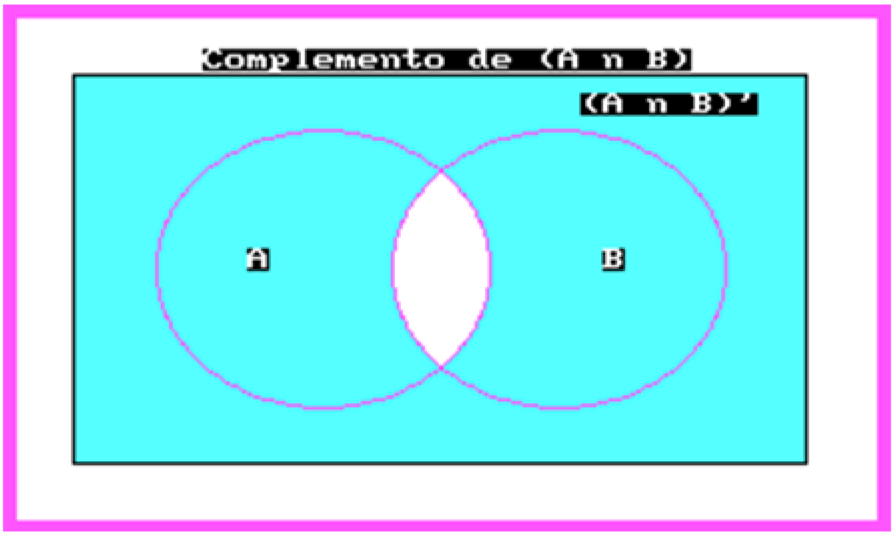

Conjuntos ajenos
U
A
B

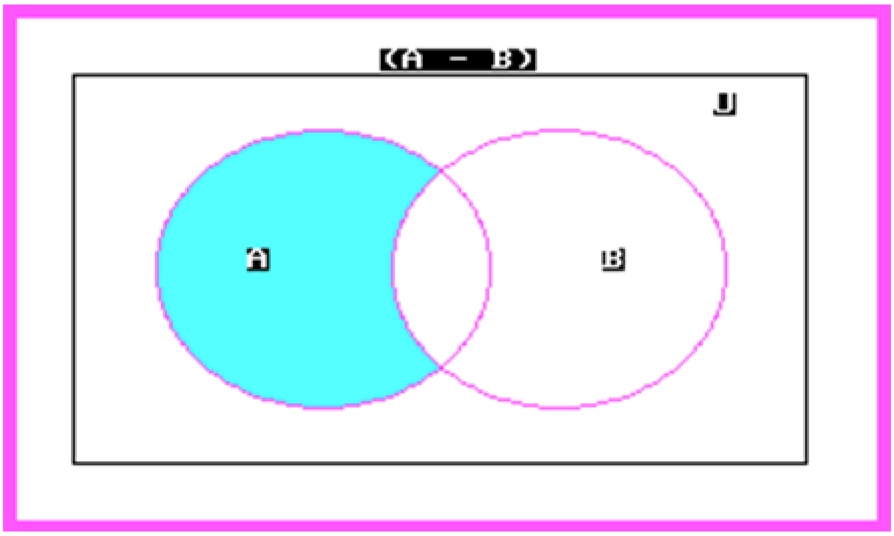

A U B
U
A
B

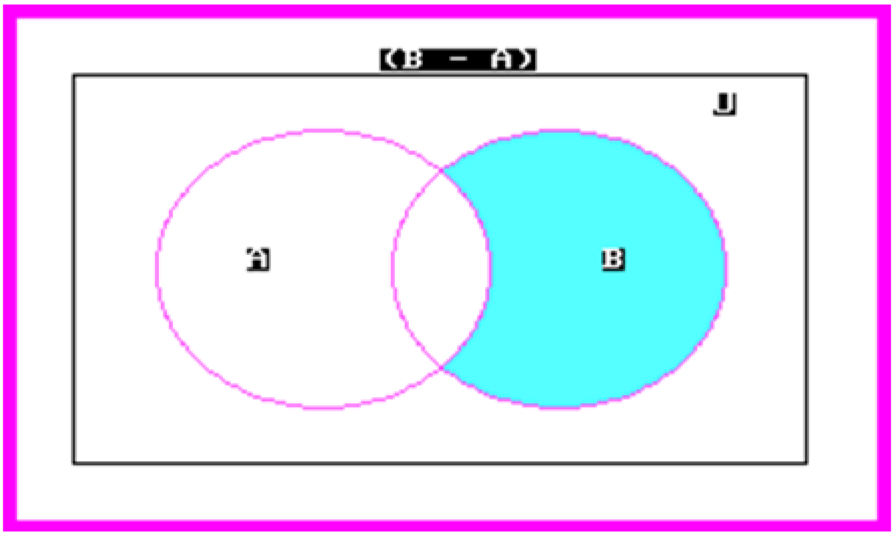

A interseccion B
U
A
B


A interseccion B
U
A
B


Complemento de A
A'
A


Complemento de (A n B)
(A n B)'
A
B


(A - B)
U
A
B



UNIDAD IV

4.1.- DISTRIBUCIONES DE VARIABLE DISCRETA MÁS IMPORTANTES

Las distribuciones de variable discreta más importantes son las siguientes:

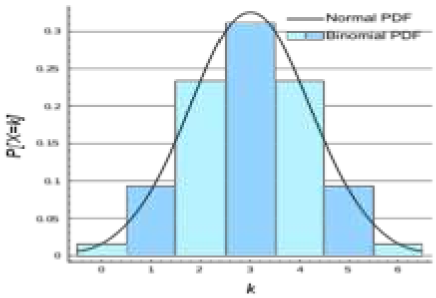
DISTRIBUCIÓN BINOMIAL;

En estadística, la distribución binomial es una distribución de probabilidad discreta que mide el número de éxitos en una secuencia de n ensayos independientes de Bernoulli con una probabilidad fija p de ocurrencia del éxito entre los ensayos.

Un experimento de Bernoulli se caracteriza por ser dicotómico, esto es, sólo son posibles dos resultados. A uno de estos se denomina éxito y tiene una probabilidad de ocurrencia p y al otro, fracaso, con una probabilidad q = 1 – p. En la distribución binomial el anterior experimento se repite n veces, de forma independiente, y se trata de calcular la probabilidad de un determinado número de éxitos. Para n = 1, la binomial se convierte, de hecho, en una distribución de Bernoulli.

Para representar que una variable aleatoria X sigue una distribución binomial de parámetros n y p, se escribe:

$$X \sim B(n,p)$$

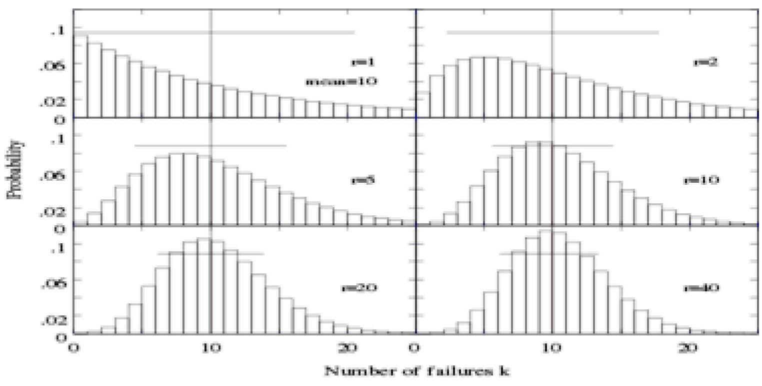


DISTRIBUCIÓN BINOMIAL NEGATIVA

En estadística la distribución binomial negativa es una distribución de probabilidad discreta que incluye a la distribución de Pascal.

El número de experimentos de Bernoulli de parámetro θ independientes realizados hasta la consecución del k-ésimo éxito es una variable aleatoria que tiene una distribución binomial negativa con parámetros k y θ.

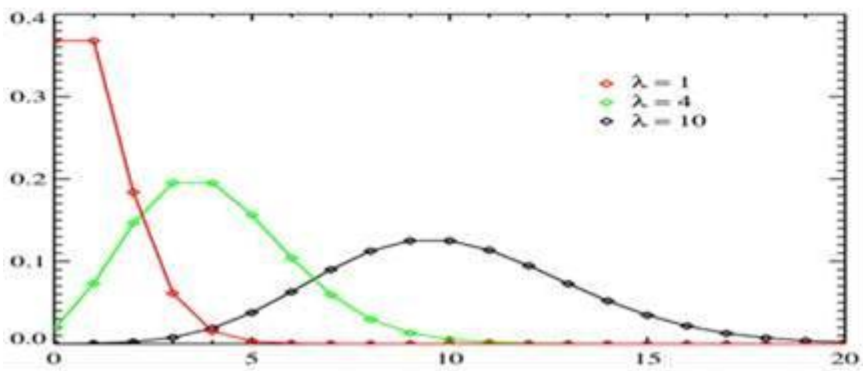
La distribución geométrica es el caso concreto de la binomial negativa cuando k = 1.



DISTRIBUCIÓN DE POISSON

En teoría de probabilidad y estadística, la distribución de Poisson es una distribución de probabilidad discreta. asi tiempo fijo si estos eventos ocurren con una frecuencia media conocida y son independientes del tiempo discurrido desde el último evento.

Fue descubierta por Siméon-Denis Poisson, que la dio a conocer en 1838 en su trabajo Recherches sur la probabilité des jugements en matières criminelles et matière civile (Investigación sobre la probabilidad de los juicios en materias criminales y civiles).



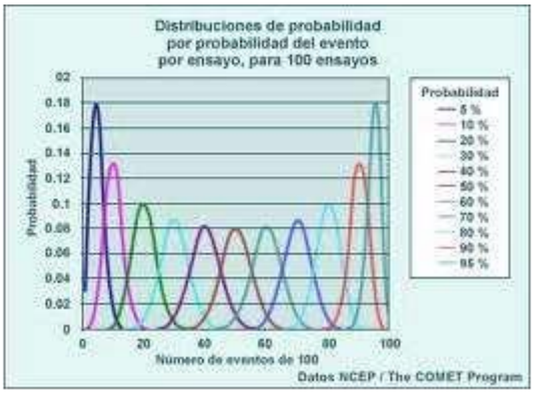
DISTRIBUCIÓN GEOMÉTRICA

En teoría de probabilidad y estadística, la distribución geométrica es cualquiera de las dos distribuciones de probabilidad discretas siguientes:

- La distribución de probabilidad del número X del ensayo de Bernoulli necesaria para obtener un éxito, contenido en el conjunto { 1, 2, 3,...} o

- La distribución de probabilidad del número Y = X − 1 de fallos antes del primer éxito, contenido en el conjunto { 0, 1, 2, 3,... }.

Cuál de éstas es la que uno llama "la" distribución geométrica, es una cuestión de convención y conveniencia.



DISTRIBUCIÓN HIPERGEOMÉTRICA

En teoría de la probabilidad la distribución hipergeométrica es una distribución discreta relacionada con muestreos aleatorios y sin reemplazo. Supóngase que se tiene una población de N elementos de los cuales, d pertenecen a la categoría A y N-d a la B. La distribución hipergeométrica mide la probabilidad de obtener x ( ) elementos de la categoría A en una muestra de n elementos de la población original

Distribución de Bernoulli

En teoría de probabilidad y estadística, la distribución de Bernoulli (o distribución dicotómica), nombrada así por el matemático y científico suizo Jakob Bernoulli, es una distribución de probabilidad discreta, que toma valor 1 para la probabilidad de éxito (p) y valor 0 para la probabilidad de fracaso (q = 1 − p).

Si X es una variable aleatoria que mide "número de éxitos", y se realiza un único experimento con dos posibles resultados (éxito o fracaso), se dice que la variable aleatoria X se distribuye como una Bernoulli de parámetro p.

X˜Be(p)

La fórmula será: f(x) = px(1 − p)1

− x con x = {0,1}

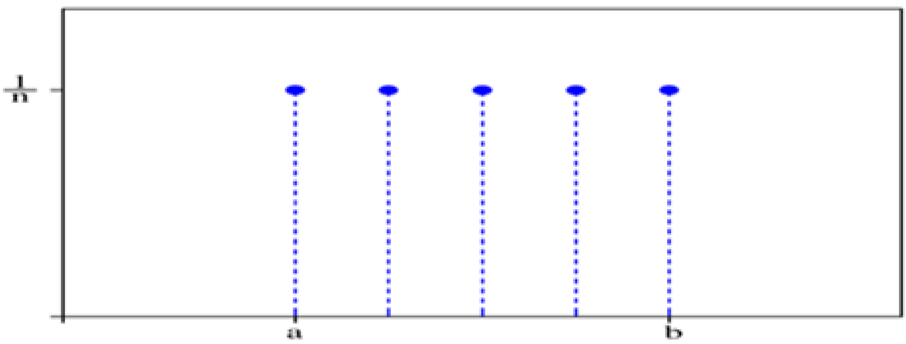
Su función de probabilidad viene definida por:

$$f(x;p) = \begin{cases} p & \text{si } x = 1, \\ q & \text{si } x = 0, \\ 0 & \text{en cualquier otro caso} \end{cases}$$

Un experimento al cual se aplica la distribución de Bernoulli se conoce como Ensayo de Bernoulli o simplemente ensayo, y la serie de esos experimentos como ensayos repetidos.

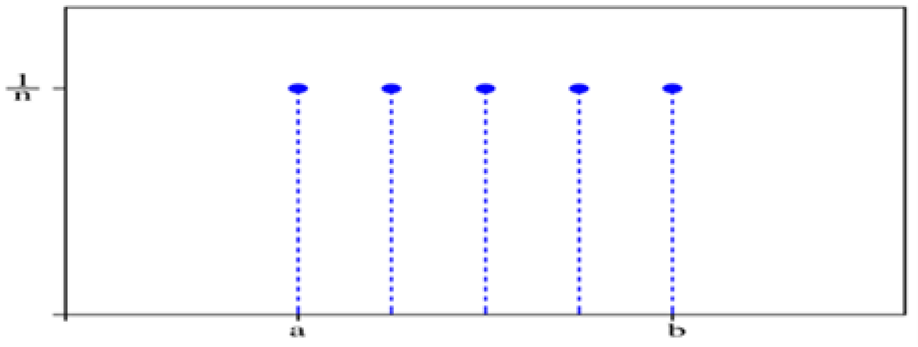
DISTRIBUCIÓN UNIFORME DISCRETA

En teoría de la probabilidad, la distribución uniforme discreta es una distribución de probabilidad que asume un número finito de valores con la misma probabilidad.



DISTRIBUCIÓN UNIFORME DISCRETA

En teoría de la probabilidad, la distribución uniforme discreta es una distribución de probabilidad que asume un número finito de valores con la misma probabilidad.



4.2.- DISTRIBUCIONES DE VARIABLE CONTINUA DISTRIBUCIÓN X²

En estadística, la distribución χ^2 (de Pearson) es una distribución de probabilidad continua con un parámetro k que representa los grados de libertad de la variable aleatoria:

Un experimento al cual se aplica la distribución de Bernoulli se conoce como Ensayo de

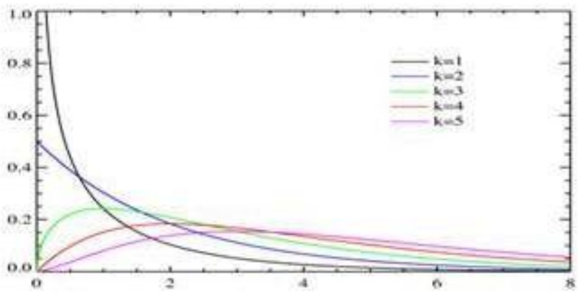
Bernoulli o simplemente ensayo, y la serie de esos experimentos como ensayos repetidos.

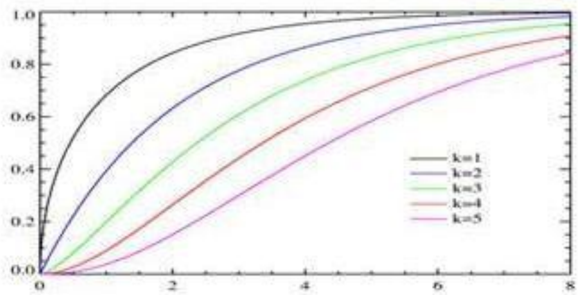
$$X = Z_1^2 + \cdots + Z_k^2$$

Donde Zi son variables de distribución normal, de media cero y varianza uno. El que la variable aleatoria X tenga esta distribución se representa habitualmente así:

Es conveniente tener en cuenta que la letra griega χ se transcribe al latín como chi1 y se pronuncia en castellano comoji.2 3

La distribución χ^2 tiene muchas aplicaciones en inferencia estadística, por ejemplo en la denominada prueba χ^2 utilizada como prueba de independencia y como prueba de bondad de ajuste y en la estimación de varianzas. También está involucrada en el problema de estimar la media de una población normalmente distribuida y en el problema de estimar la pendiente de una recta de regresión lineal, a través de su papel en la distribución t de Student, y participa en todos los problemas de análisis de varianza, por su papel en la distribución F de Snedecor, que es la distribución del cociente de dos variables aleatorias independientes con distribución χ^2.

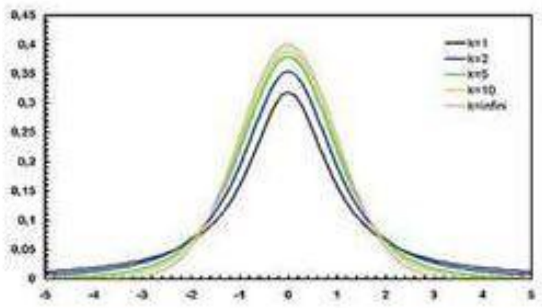


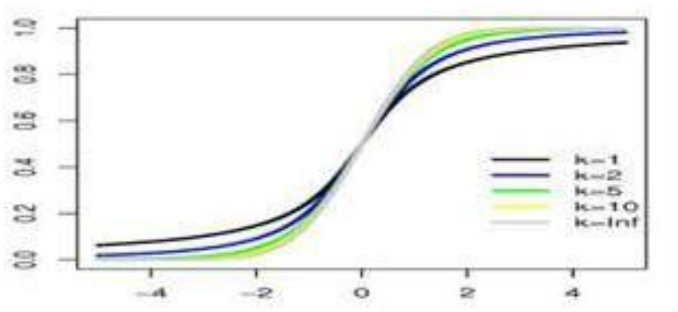


Distribución t de Student

En probabilidad y estadística, la distribución t (de t-Student) es una distribución de probabilidad que surge del problema de estimar la media de una población normalmente distribuida cuando el tamaño de la muestra es pequeño.

Aparece de manera natural al realizar la prueba t de Student para la determinación de las diferencias entre dos medias muéstrales y para la construcción del intervalo de confianza para la diferencia entre las medias de dos poblaciones cuando se desconoce la desviación típica de una población y ésta debe ser estimada a partir de los datos de una muestra.



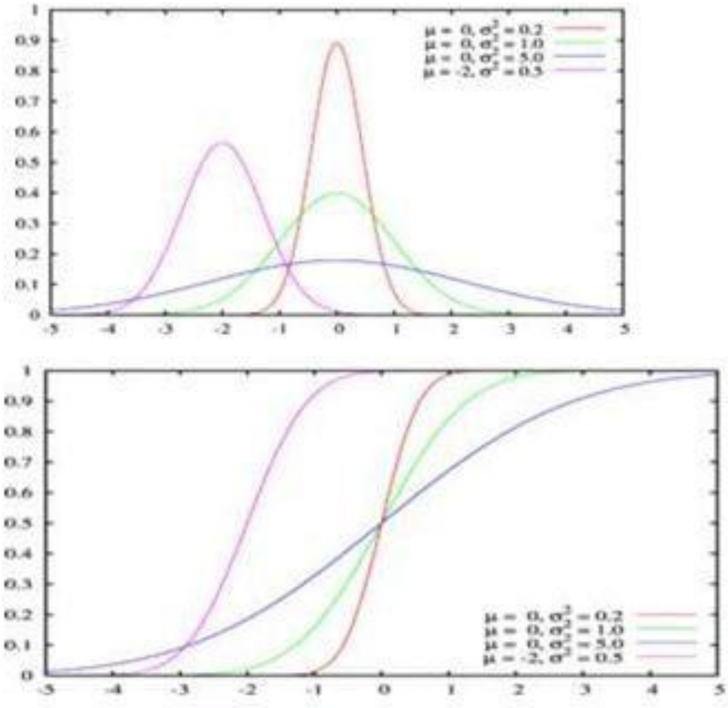


Distribución normal

En estadística y probabilidad se llama distribución normal, distribución de Gauss o distribución gaussiana, a una de las distribuciones de probabilidad de variable continua que con más frecuencia aparece en fenómenos reales.

La gráfica de su función de densidad tiene una forma acampanada y es simétrica respecto de un determinado parámetro. Esta curva se conoce como campana de Gauss.

La importancia de esta distribución radica en que permite modelar numerosos fenómenos naturales, sociales y psicológicos. Mientras que los mecanismos que subyacen a gran parte de este tipo de fenómenos son desconocidos, por la enorme cantidad de variables incontrolables que en ellos intervienen, el uso del modelo normal puede justificarse asumiendo que cada observación se obtiene como la suma de unas pocas causas independientes.



Distribución gamma

En estadística la distribución gamma es una distribución de probabilidad continua con dos parámetros k y λ cuya función de densidad para valores x > 0 es

$$f(x) = \lambda e^{-\lambda x} \frac{(\lambda x)^{k-1}}{\Gamma(k)}$$

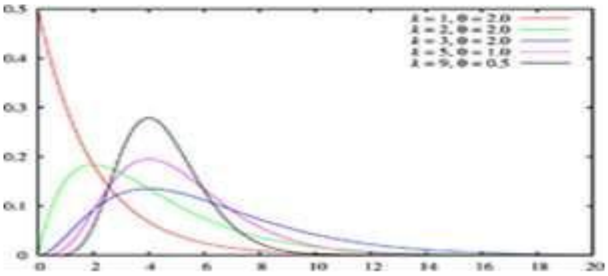
Aquí e es el número e y Γ es la función gamma. Para valores:

La aquella es Γ(k) = (k − 1)! (el factorial de k − 1). En este caso – por ejemplo para describir un proceso de Poisson – se llaman la distribución distribución Erlang con un parámetro θ = 1 / λ.

El valor esperado y la varianza de una variable aleatoria X de distribución gamma son:

E[X] = k / λ = kθ

V[X] = k / λ2 = kθ2



Distribución beta

En estadística la distribución beta es una distribución de probabilidad continua con dos parámetros a y b cuya función de densidad para valores 0 < x < 1 es

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$$

Aquí Γ es la función gamma.

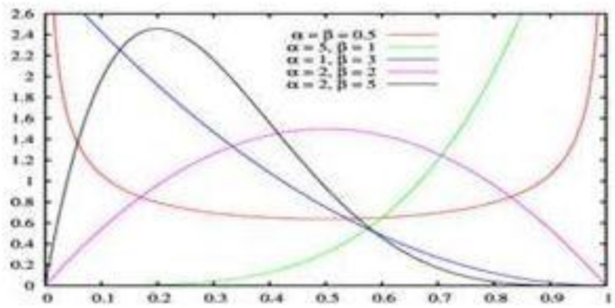
El valor esperado y la varianza de una variable aleatoria X con distribución beta son

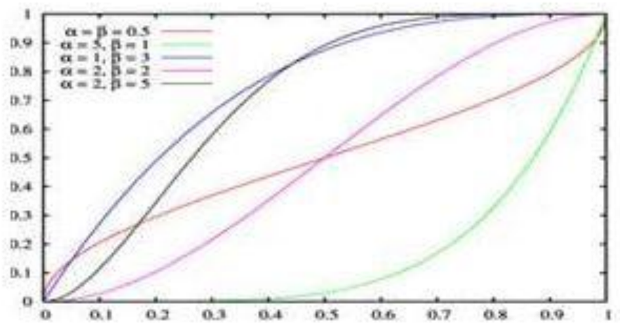
$$E[X] = \frac{a}{a+b}$$

$$V[X] = \frac{ab}{(a+b+1)(a+b)^2}$$

Un caso especial de la distribución beta con a = 1 y b = 1 es la distribución uniforme en el intervalo [0, 1].

Para relacionar con la muestra se iguala E[X] a la media y V[X] a la varianza y de despejan a y b





Distribución F

Usada en teoría de probabilidad y estadística, la distribución F es una distribución de probabilidad continua. También se la conoce como distribución F de Snedecor (por George Snedecor) o como distribución F de Fisher-Snedecor.

Una variable aleatoria de distribución F se construye como el siguiente cociente:

$$F = \frac{U_1/d_1}{U_2/d_2}$$

Donde

- U1 y U2 siguen una distribución chi-cuadrado con d1 y d2 grados de libertad respectivamente, y
- U1 y U2 son estadísticamente independientes.

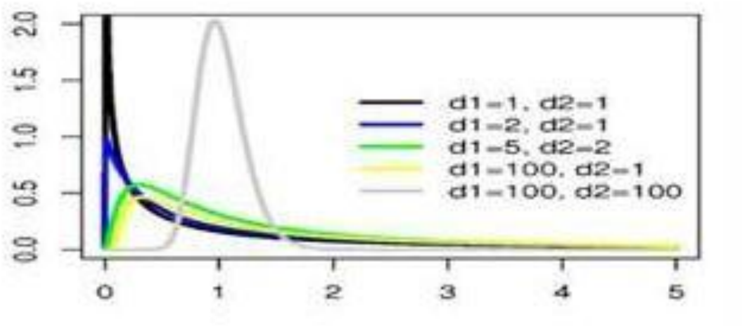
La distribución F aparece frecuentemente como la distribución nula de una prueba estadística, especialmente en el análisis de varianza. Véase el test F.

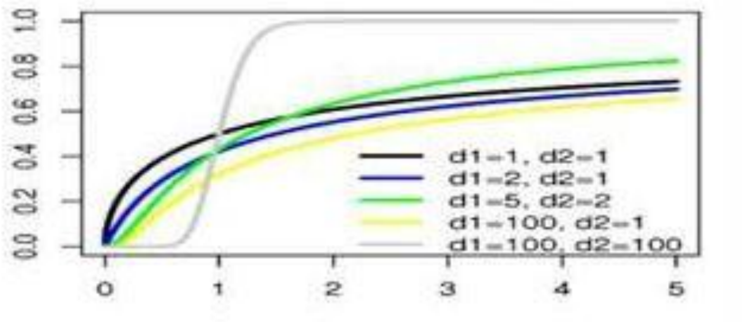
La función de densidad de una F(d1, d2) viene dada por

$$g(x) = \frac{1}{\mathrm{B}(d_1/2, d_2/2)} \left(\frac{d_1\, x}{d_1\, x + d_2}\right)^{d_1/2} \left(1 - \frac{d_1\, x}{d_1\, x + d_2}\right)^{d_2/2} x^{-1}$$

Para todo número real x ≥ 0, donde d1 y d2 son enteros positivos, y B es la función beta. La función de distribución es para todo número real x ≥ 0, donde d1 y d2 son enteros positivos, y B es la función beta. La función de distribución es

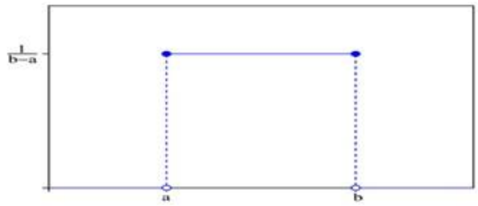
$$G(x) = I_{\frac{d_1 x}{d_1 x + d_2}}(d_1/2, d_2/2)$$

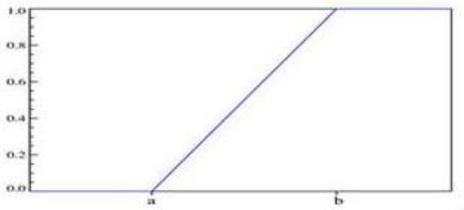




Distribución uniforme continua

En teoría de probabilidad y estadística, la distribución uniforme continua es una familia de distribuciones de probabilidad para variables aleatorias continuas, tales que cada miembro de la familia, todos los intervalos de igual longitud en la distribución en su rango son igualmente probables. El dominio está definido por dos parámetros, a y b, que son sus valores mínimo y máximo. La distribución es a menudo escrita en forma abreviada como U(a,b).





4.4.- MUESTREO

1.- INTRODUCCIÓN

El muestreo estadístico es la herramienta que la Matemática utiliza para el estudio de las características de una población a través de una determinada parte de la misma.

La muestra de estudio debe ser lo más pequeña posible ya que del hecho de que una muestra sea más grande, no se desprende necesariamente que la información sea más fiable.

Además, la muestra elegida debe serlo por un proceso aleatorio para que sea lo más representativa posible.

Términos usuales en un estudio estadístico

- Población: conjunto de todos los individuos que son objeto del estudio.
- Muestra: parte de la población en la que miden las características estudiadas.
- Muestreo: proceso seguido para la extracción de una muestra.
- Encuesta: proceso de obtener información de la muestra. Métodos de muestreo

1.- Muestreo no probabilístico: no se usa el azar, sino el criterio del investigador.

2.- Muestreo probabilístico o aleatorio:

2.1.- Muestreo aleatorio simple: se asigna un número a cada uno de los individuos de la población, y seguidamente se van eligiendo al azar los componentes de la muestra. La elección de un individuo no debe afectar a la del siguiente, por tanto debe reemplazarse el nº, una vez extraído.

2.2.- Muestreo sistemático: se ordenan previamente los individuos de la población, después se elige uno al azar y a continuación, a intervalos constantes, se eligen todos los demás hasta completar la muestra.

2.3.- Muestreo estratificado: se divide la población total en clases homogéneas (estratos). La muestra se escoge aleatoriamente en número proporcional al de los componentes de cada estrato.

Ejemplo: en un I.E.S. hay 120 alumnos en 2º de Bachillerato provenientes de 4 zonas o pueblos.

Zona A: 20 alumnos Zona B: 32 alumnos Zona C: 60 alumnos Zona D: 8 alumnos

Hay que elegir una muestra de 20 alumnos para hacerles una serie de preguntas. Utiliza los tres métodos de muestreo aleatorio para escoger la muestra.

2.- DISTRIBUCIONES DE MUESTREO

Es evidente que los resultados obtenidos del estudio de una muestra no son del todo fiable, pero sí en buena medida. Los parámetros que obtienen de una muestra (estimadores estadísticos) nos permitirán arriesgarnos a predecir una serie de resultados para toda la población. De estas predicciones y del riesgo que conllevan se ocupa la Inferencia Estadística.

4.4.1.- Distribución de medias muéstrales

Si una población tiene N elementos, el nº de muestras distintas de tamaño n que se pueden elegir es. Si pueden repetirse individuos, el número de muestras será igual a

Ejemplo: calcular el nº de muestra de tamaño 21 que pueden elegirse en una población de 120 alumnos:

Repaso de la distribución normal

Ejercicios:

- ☐ Si Z es una N(0, 1), calcular las siguientes probabilidades:

a) p(Z<1) b) p(Z>1´3) c) p(Z<-0´5) d) p(-0´5<Z<1´3)

- ☐ Si X es una N(15, 3), responder a las siguientes cuestiones:

- ☐ tipificarla a una N(0, 1) con el cambio

- ☐ calcular las siguientes probabilidades:

p(X<21) p(X<-7) p(X>31)

Parámetros muéstrales

Elegida una muestra, hallaremos en ella la media y la desviación típica S. Lo que tendremos que estudiar será la representatividad de estos parámetros muestrales con los parámetros reales de la población, es decir: la media poblacional, y la desviación típica de la población.

Si en una población de N individuos tomamos todas las muestras posibles de tamaño n, se puede demostrar que la media de las medias muéstrales coincide con la media poblacional, esto es. Sin embargo, no se cumple lo mismo para la desviación típica de las medias muéstrales, sino que se verifica que, siendo n el tamaño de las muestras. Si las medias muéstrales provienen de una población no normal, pero el tamaño de las mismas es n"30, la distribución de las medias muéstrales también se ajusta a una

Ejemplo: en el último año, el peso de los recién nacidos en una maternidad se ha distribuido según una ley normal de parámetros =3.100 gramos y = 150 gramos.

- ☐ ¿Cuál es la probabilidad de que un recién nacido pese más de 3.130 gramos?
- ☐ ¿Qué distribución seguirán las muestras de tamaño 100 de recién nacidos?

☐ ¿Cuál será la probabilidad de que la media de una muestra de tamaño 100 sea superior a 3.130 gramos?

Ejercicio: en una oposición en la que participan miles de candidatos se hizo un examen tipo test. Las calificaciones se distribuyeron normalmente con media =72 puntos y desviación típica =10.

☐ ¿Cuál es la probabilidad de que un opositor elegido al azar obtenga más de 76 puntos?

☐ ¿Cuál es la probabilidad de que una muestra de 64 opositores obtenga un promedio superior a 76 puntos?

Ejercicios:

☐ Supongamos que la estatura media de las alumnas de bachillerato es 165 cm, con desviación típica 8 cm.

☐ Halla los parámetros de las medias muestrales de tamaños n=36 y n= 64

☐ ¡Cuál es la probabilidad de que una muestra de 36 alumnas tenga una media superior a 167 cm? ¿Y de que una muestra de 64 alumnas supere esa misma medida?

☐ ¡Tiene algo de extraño que una muestra de tamaño 36 tenga una media de 170 cm.?

3.- INTERVALOS DE PROBABILIDAD

A los intervalos simétricos respecto de la media o proporción poblacionales se les denomina intervalos de probabilidad.

Intervalos de probabilidad para la media muestral

Sabemos que la distribución de medias muéstrales es normal de media y desviación típica, donde son los parámetros de la población.

Nos haremos la siguiente pregunta

¿Cuál es la probabilidad de que la media maestral se encuentre entre dos valores simétricos respecto de la media poblacional?, es decir, queremos evaluar las siguientes probabilidades:

Se llama intervalo de probabilidad para la media a uno de la forma tal que se cumple que la probabilidad de que se encuentre en él es igual. Al parámetro se le llama nivel de confianza, y la diferencia es el riesgo asumido.

Si tipificamos la variable, llegaremos a una expresión de la forma:

, donde Z es una variable que se ajusta a una N(0, 1). De este modo podremos evaluar el valor de k consultando la tabla de valores de dicha distribución.

Ejemplo: vamos a hallar el intervalo de probabilidad para el peso medio de una muestra de

100 recién nacidos, con un nivel de confianza de 0,9, sabiendo que =3.100 gramos y =150 gramos.

Solución: como se ha dicho anteriormente, tenemos que evaluar la siguiente expresión si consultamos en la tabla de la N(0, 1), comprobaremos que, por lo tanto, el intervalo de probabilidad será el siguiente:

que simplificado, es el intervalo

(3.075´325 ; 3.124´675)

Ejercicios:

- ☐ Hallar el intervalo de probabilidad con una confianza de 0´95 para la misma distribución.

- ☐ Para las muestras de tamaño 36 extraídas de la distribución de calificaciones en una población de 120 alumnos, con media 5´5 y desviación típica 2´04, halla los intervalos de probabilidad para un nivel de confianza de:

☐ 75´4%

☐ 0´87

Ejercicios:

☐ Si la estatura de las alumnas de segundo de Bachillerato se ajusta a la normal N(165, 8), en cm, halla, para las muestras de tamaño 64:

☐ El porcentaje de ellas que dará una media entre 163 y 167 cm.

☐ El intervalo de probabilidad con un nivel de confianza del 80%.

☐ El nivel medio de colesterol (en mg/dl), en individuos sanos, depende de la edad y el sexo; para los hombres con menos de 21 años su distribución es normal con media =160 y desviación típica =10. Un nivel fuera de resulta extraño: indica que puede haber alguna anomalía. Lo mismo cabe decir de las muestras: un nivel maestral fuera de resulta extraño. ¿Cuál es el intervalo de probabilidad admisible (no extraño) para las muestra de tamaño

☐ 1

☐ 9

☐ 100

☐ ¿Qué porcentaje de individuos o muestras se encuentran en los intervalos hallados para los diferentes tamaños de la muestra?

ESTIMACIÓN ESTADÍSTICA

1.- ESTIMACIÓN A PARTIR DE UNA MUESTRA

Habitualmente, lo normal es que se desconozcan la media y la desviación típica de la población y que, mediante técnicas de muestreo, se busque estimarlas con la fiabilidad necesaria. Así, si para 400 individuos de una región, elegidos al azar, se obtiene una renta per cápita de 1.215.000 ptas, con una desviación típica de 650.000 ptas, podemos hacernos dos preguntas:

- ☐ ¿La renta per cápita de los habitantes de toda la región será de 1.215.000 ptas?
- ☐ ¿Qué seguridad se tiene de tal afirmación?

Cuando se contestan estas preguntas se está haciendo una estimación a partir de la muestra.

2.- INTERVALOS DE CONFIANZA

En este apartado vamos a dar respuesta a las dos preguntas anteriores.

Intervalo de confianza para la media maestral

Al intervalo se le llama intervalo de confianza para la media poblacional, siendo los elementos que aparecen en dicho intervalo, los ya estudiados anteriormente.

La probabilidad de que la media de la población se encuentre en este intervalo es, que es el nivel de confianza. Si la confianza es, suele decirse que el nivel de significación es 1-, o nivel de riesgo.

En el caso en que la desviación típica de la población sea desconocida, no tendríamos más remedio que sustituirla por la desviación maestral s; así el intervalo de confianza para la media poblacional, para, sería con una probabilidad de, siendo y s la media y la desviación típica de la muestra, respectivamente.

A se le llama error típico de la media.

Ejemplo: para una muestra de 400 personas elegidas al azar se obtiene una renta per cápita de 1.215.000 ptas. Si la desviación típica de la renta per cápita para la población es de 700.000 ptas, calcula el intervalo de confianza para la media poblacional con un nivel de significación de:

- ☐ 0,1
- ☐ 0,05

Ejercicios:

- ☐ Para una muestra de 30 alumnos se obtuvo una nota media en el último examen de matemáticas de, con una desviación típica s= 1'92. Determina el intervalo de confianza al 80%. Interpreta el resultado.
- ☐ El peso medio de una muestra de 100 recién nacidos es 3.200 gramos. Sabiendo que la desviación típica de los pesos de la población de recién nacidos es de 150 gramos, halla el intervalo de confianza para la media poblacional para una significación de 0'05

3.- ERROR ADMITIDO Y TAMAÑO DE LA MUESTRA

Error admitido

Cuando decimos que la media poblacional con un nivel de confianza, estamos admitiendo un error máximo de . A este número se le llama error máximo admisible.

Tamaño maestral

El tamaño maestral mínimo de una encuesta depende de la confianza que se desee para los resultados y del error máximo que se esté dispuesto a asumir.

Cómo determinar el tamaño de una maestral

Determinar el tamaño de la muestra que se va a seleccionar es un paso importante en cualquier estudio de investigación de mercados, se debe justificar convenientemente de acuerdo al planteamiento del problema, la población, los objetivos y el propósito de la investigación.

¿De qué depende el tamaño maestral?

El tamaño maestral dependerá de decisiones estadísticas y no estadísticas, pueden incluir por ejemplo la disponibilidad de los recursos, el presupuesto o el equipo que estará en campo.

Antes de calcular el tamaño de la muestra necesitamos determinar varias cosas:

1. Tamaño de la población. Una población es una colección bien definida de objetos o individuos que tienen características similares. Hablamos de dos tipos: población objetivo, que suele tiene diversas características y también es conocida como la población teórica.

La población accesible es la población sobre la que los investigadores aplicaran sus conclusiones.

2. Margen de error (intervalo de confianza). El margen de error es una estadística que expresa la cantidad de error de muestreo aleatorio en los resultados de una encuesta, es decir, es la medida estadística del número de veces de cada 100 que se espera que los resultados se encuentren dentro de un rango específico.

3. Nivel de confianza. Son intervalos aleatorios que se usan para acotar un valor con una determinada probabilidad alta. Por ejemplo, un intervalo de confianza de 95% significa que los resultados de una acción probablemente cubrirán las expectativas el 95% de las veces.

4. La desviación estándar. Es un índice numérico de la dispersión de un conjunto de datos (o población). Mientras mayor es la desviación estándar, mayor es la dispersión de la población.

Cálculo del Tamaño de la Muestra desconociendo el Tamaño de la Población

La fórmula para calcular el tamaño de muestra cuando se desconoce el tamaño de la población es la siguiente:

$$n = \frac{Z_a^{\,2} \times p \times q}{d^2}$$

En donde

Z=nivel de confianza,

P = probabilidad de éxito, o proporción esperada

Q = probabilidad de fracaso

D = precisión (error máximo admisible en términos de proporción) Cálculo del Tamaño de la Muestra conociendo el Tamaño de la Población

La fórmula para calcular el tamaño de muestra cuando se conoce el tamaño de la población es la siguiente:

$$n = \frac{N \times Z_a^{\,2} \times p \times q}{d^2 \times (N-1) + Z_a^{\,2} \times p \times q}$$

En donde, N = tamaño de la población Z = nivel de confianza, P = probabilidad de éxito, o proporción esperada Q = probabilidad de fracaso D = precisión (Error máximo admisible en términos de proporción).

Tipos de muestreo

El muestreo es una herramienta para determinar qué parte de una población debemos analizar cuando no es posible realizar un censo. Depende de los objetivos del estudio el elegir una muestra probabilística o no probabilística.

MUESTREO PROBABILÍSTICO

Se basa en el principio de equiprobabilidad, esto quiere decir que todos los individuos de la muestra seleccionada, tendrán las mismas probabilidades de ser elegidos. Lo anterior nos asegura que la muestra extraída contará con representatividad.

MUESTREO NO PROBABILÍSTICO

No sirven para hacer generalizaciones pero sí para estudios exploratorios. En este tipo de muestras, se eligen a los individuos utilizando diferentes criterios relacionadas con las características de la investigación, no tienen la misma probabilidad de ser seleccionados ya que el investigador suele determinar la población objetivo

Términos básicos en muestreo

¿Hacia quiénes queremos generalizar? = Población Teorética

¿A qué población tenemos acceso? = Población de Estudio

¿Cómo obtenemos el acceso? = Marco de Muestra

¿Quién está en nuestro estudio? = La Muestra

Gráfico o diagrama de control

Un gráfico de control es una herramienta utilizada para distinguir las variaciones debidas a causas asignables o especiales a partir de las variaciones aleatorias inherentes al proceso.

Las variaciones aleatorias se repiten casualmente dentro de los límites predecibles.

Las variaciones debidas a causas asignables o especiales indican que es necesario identificar, investigar y poner bajo control algunos factores que afectan al proceso.

La construcción de gráficos de control está basada en la estadística matemática. Los gráficos de control emplean datos de operación para establecer límites dentro de los cuales se espera hacer observaciones futuras, si el proceso demuestra no haber sido afectado por causas asignables o especiales.

Causas Asignables

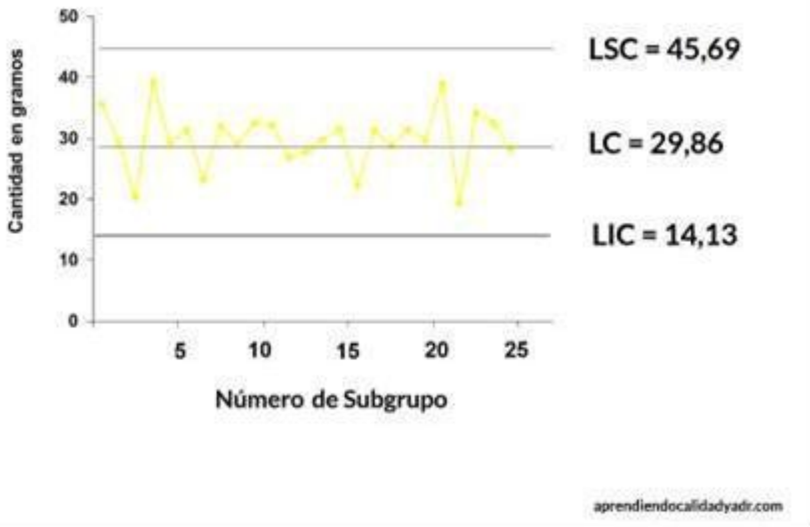
Factores (generalmente numerosos, pero individualmente de relativa importancia) que se pueden detectar e identificar como causantes de un cambio en una característica de la calidad o nivel del proceso.

Nota: En ocasiones, se denominan causas especiales de variación. Causas Aleatorias

Factores generalmente numerosos, pero poco importantes, que contribuyen a la variación y no han sido necesariamente identificados.

Nota: En ocasiones, se denominan causas comunes de variación.

Existe una gran variedad de gráficos de control que se pueden aplicar a todo tipo de características medibles o contables de un proceso, un producto o cualquier salida.



Límite superior de control (LSC): Es el mayor valor aceptado en el proceso.

Límite inferior de control (LIC): Es el valor más pequeño que se acepta en el proceso. Límite central de control (LC): Es la línea central del gráfico. Mientras más cerca estén los puntos a la línea, más estable es el proceso.

¿Para qué sirve un gráfico o diagrama de control?

Diagnóstico: Para evaluar la estabilidad de un proceso.

Control: Para determinar cuándo es necesario ajustar un proceso y cuándo se debe dejar tal y como está.

Confirmación: Para confirmar la mejora de un proceso.

Tipos de gráficos de control

Gráfico de control por variables

Hacen uso de estadísticas obtenidas a partir de datos tales como la longitud o grosor de un elemento.

En los gráficos de control por variables es posible medir la característica de calidad a estudiar. En estos casos conviene describir la característica de calidad mediante una medida de tendencia central (usualmente la media muestral) y una medida de su variabilidad (usualmente el rango o la desviación estándar).

Los gráficos de control por variables son más "sensibles" que los gráficos de control por atributos, razón por la cual son capaces de "avisarnos" de posibles problemas de calidad incluso antes de que éstos sean ya relevantes.

Gráfico de control por atributos

Se basan en frecuencias, por ejemplo el número de unidades defectuosas.

En estos gráficos el control del proceso se realiza si el producto inspeccionado se clasifica como no conforme o conforme (defectuoso o no defectuoso), respecto a las especificaciones para la característica de calidad considerada. Ejemplo: un tornillo es conforme si su longitud está entre 1,9 y 2,1 cm., en caso contrario será no conforme.

Los gráficos de control por atributos tienen la ventaja de sintetizar de forma rápida toda la información referida a diferentes aspectos de calidad de un producto, ya que permiten clasificar éste como aceptable o inaceptable; además, no suelen necesitar de sistemas de medición muy complejos y son más fácilmente entendibles por los no especialistas.

Bibliografía básica y complementaria:

Probabilidad y estadística de George Canavos Estadística de Murray R. Spiegel