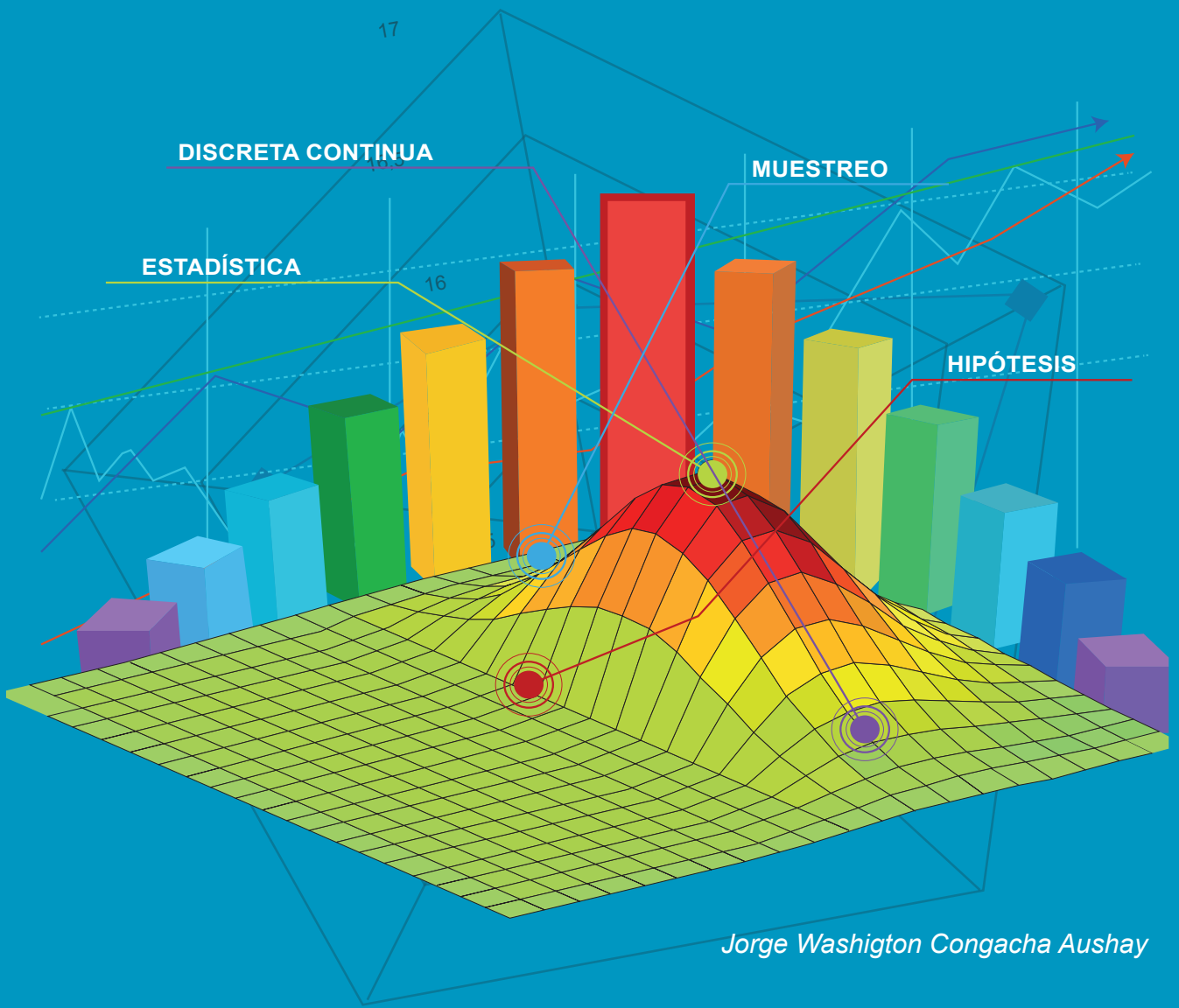
$$n_i = \frac{N_i}{N} n, \qquad S = \sqrt{\frac{\sum fX^2 - \frac{(\sum fX)^2}{n}}{n-1}} = \sqrt{\frac{48925 - \frac{(1540)^2}{52}}{51}} = 8{,}065$$

ESTADÍSTICA APLICADA A LA EDUCACIÓN

con actividades de aprendizaje



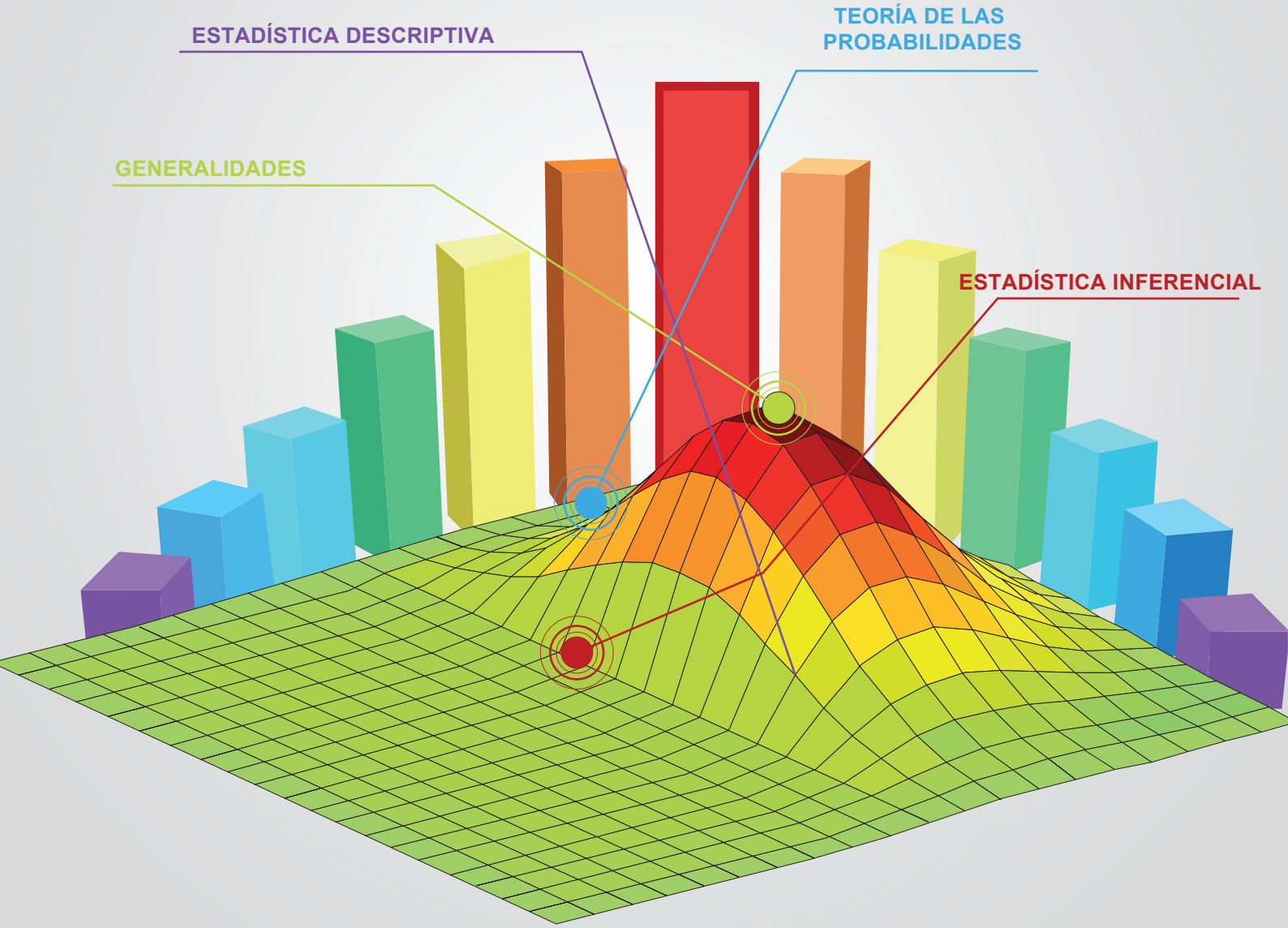
Jorge Washigton Congacha Aushay

RIOBAMBA - ECUADOR
2015

Tomo 1

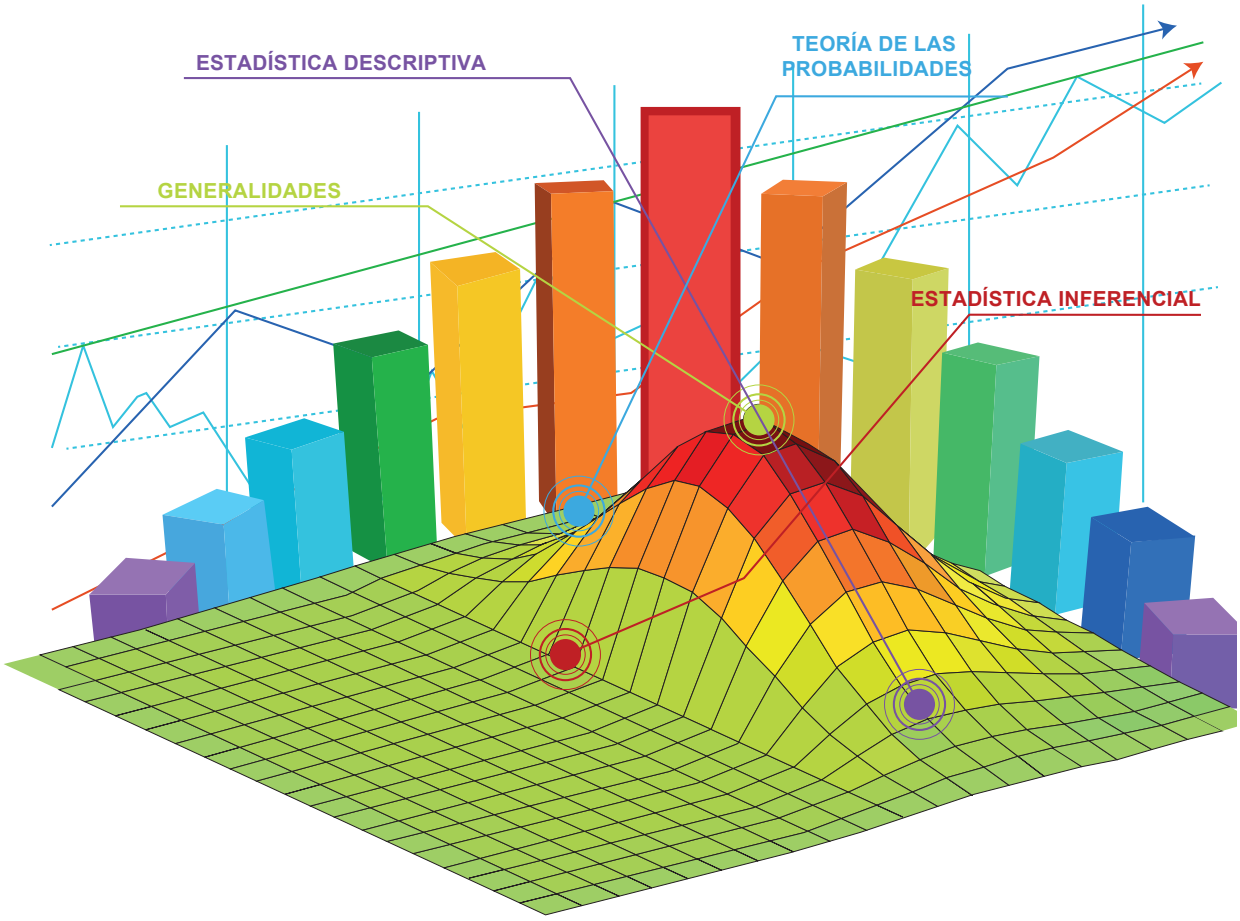
ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO
FACULTAD DE CIENCIAS
ESCUELA DE FÍSICA Y MATEMÁTICA
CARRERA DE INGENIERÍA EN ESTADÍSTICA-INFORMÁTICA

ESTADÍSTICA APLICADA A LA EDUCACIÓN CON ACTIVIDADES DE APRENDIZAJE



Jorge Washington Congacha Aushay

RIOBAMBA - ECUADOR
2016



ESTADÍSTICA APLICADA A LA EDUCACIÓN

con actividades de aprendizaje

Jorge Washington Congacha Aushay
RIOBAMBA-ECUADOR

Autor

Jorge Washington Congacha Aushay

Portada

Centro de Apoyo Diseño Editorial EDG-FIE

Diseño y Diagramación

Centro de Apoyo Diseño Editorial EDG-FIE

ISBN 978-3-8484-6434-0
www.eae-publishing.com

Segunda Edición 2016

Riobamba - Ecuador

eae - Editorial Académica Española
es una marca comercial de AV AKADEMIKERVERLAG GmbH & Co. KG
Heinrich-Böcking-Str. 6-8 D - 66121 Saarbrücken
Telefon: +49 681 3720-310
Telefax: +49 681 3720-3109
www.eae-publishing.com

6 de diciembre de 2012

A quien corresponda,

Expido el presente certificado a fin de informarle que el trabajo académico titulado *"Estadística Aplicada a la Educación con Actividades de Aprendizaje"*, cuya autoría corresponde a Jorge Washington Congacha Aushay fue aceptado para publicación con ISBN 978-3-8484-6434-0. Publicamos el libro en julio de 2012.

Puede consultar nuestro catálogo en www.eae-publishing.com

Ante cualquier consulta, no dude en contactarme.

Saludos cordiales.

Andor Sperling

Departamento de Adquisiciones
a.sperling@lap-publishing.com

Editorial Académica Española es una marca comercial de:
AV Akademikerverlag GmbH & Co. KG
Heinrich-Böcking-Str. 6-8
D – 66123 Saarbrücken

Handelsregister Amtsgericht Saarbrücken HRA 10752 Verkehrsnummer: 12917 - Persönlich haftende Gesellschafterin: VDM Management GmbH Handelsregister Amtsgericht Saarbrücken HRB 18918 - Geschäftsführung: Thorsten Ohm (CEO), Dr. Wolfgang Müller, Esther von Krosigk

Dedicatoria

Con mucho cariño a
mi esposa Miriamín y a mis hijas: Giorgia, Antonella y Sofía, espero que ellas también cultiven la enseñanza de las Aplicaciones de las Matemáticas.

Educación, la mejor herencia.
Mientras más enseño más aprendo.
Estadística, aprendemos de la experiencia, de los datos.

CONTENIDO

PREFACIO

En los programas ministeriales de Matemáticas se establecen argumentos de Estadística y Probabilidades. Se ha visto sin embargo que no se estudian adecuadamente o no se estudian tales temas. Se quiere estimular a una mejor enseñanza de los conceptos básicos de la Estadística y de la Teoría de las Probabilidades.

No se quiere dar un recetario de fórmulas con el único propósito de acatar cumplimiento a un programa establecido, al contrario se pretenderá presentar este texto de ***ESTADÍSTICA APLICADA A LA EDUCACIÓN CON ACTIVIDADES DE APRENDIZAJE*** *de manera atractiva, interesante y aplicativa.*

En los últimos años ha aumentado la atención en enseñar Estadística y Teoría de las Probabilidades a Nivel Primario, Medio, Superior y de Posgrado en cursos diferentes de aquellos que se dictan normalmente en Matemáticas.

El texto ***ESTADÍSTICA APLICADA A LA EDUCACIÓN CON ACTIVIDADES DE APRENDIZAJE*** *pretende ser una alternativa para dichos cursos y para su desarrollo se ha dividido en cuatro capítulos importantes tomando en cuenta los aspectos de generalidades, descripción, herramienta y conclusiones bajo los nombres de Generalidades, Estadística Descriptiva, Teoría de las Probabilidades y Estadística Inferencial o Inferencia Estadística.*

En el primer capítulo se hace una reseña histórica de los temas de Estadística y Probabilidades con el fin de destacar nombres de famosos matemáticos-probabilísticos y estadísticos los que se toman en cuenta en definiciones, propiedades y teoremas. Luego, se da una terminología básica que se utiliza en la investigación estadística, se da de manera sucinta la teoría del muestreo, llamamos la atención con los interrogantes, ¿qué es la Estadística?, ¿para qué aprender Estadística?, tipos de Estadística y finalmente se exponen los temas, de manera general, de la investigación estadística, los tipos de investigación, las etapas de una investigación y las escalas de medida.

Al final de los capítulos del 2 al 4, se puede realizar ejercicios aplicativos a la teoría vista, a lo que llamamos actividades de aprendizaje utilizando paquetes estadísticos como el Minitab, SPSS entre otros, se puede también realizar en la hoja de cálculo EXCEL, en el software estadístico libre R.

En el tercer capítulo exponemos la parte teórica-práctica de las Probabilidades requerida en el cuarto capítulo de Inferencia Estadística denominada también Estadística Inferencial principalmente en los temas de estimación de parámetros, de prueba de hipótesis y de la prueba chi-cuadrada; concluiremos con una introducción al análisis de la varianza o simplemente ANOVA las mismas que son siglas del inglés ANalysis Of VAriance.

INTRODUCCIÓN

Se escucha a menudo decir que la Matemática es una ciencia "abstracta" e incluso "inútil", alejada de los problemas concretos, pero no nos damos cuenta que la Matemática, por ende la Estadística apoya y es base fundamental de los avances tecnológicos. La Matemática, está al servicio de otras ciencias a las cuales provee conceptos, instrumentos de cálculo, estructuras y lenguaje. El calificativo de abstracta que se da a la Matemática no es extraño y el agrado por ella no es del todo espontáneo. Escribe el matemático H.O. Pollak: "Existe un pequeño número de personas que se orientan hacia la Matemática, puesto que esperan escaparse al mundo real. Ellos conciben la Matemática como una bella arquitectura, bien orientada, destacada en la vida y es sorprendente como la Matemática permite ganar para vivir".

Sin embargo, pueden parecer sorprendente las relaciones entre Matemática y realidad, en verdad, son profundas y de diversa naturaleza. Estudiamos a la Estadística como un modelo de la realidad. Podemos de todo esto decir que la Matemática no es algo aislada, tampoco es un hecho técnico, de fórmulas concebidas por mentes privilegiadas, más bien es un factor de ideas. Así su conocimiento no es un lujo, se vuelve una necesidad para las personas que quieren comprender lo que existe a su entorno, para entender las innovaciones tecnológicas, esto es, la tecnología de punta que suscita cada día. Para entender las reacciones que se dan en el mundo de las finanzas, la banca, el comercio, la economía para contrastar resultados de las investigaciones de mercado, psicológicas, educacionales decisiones gubernamentales y otras.

En la actualidad, la Estadística es una herramienta clave para el desarrollo de la investigación educativa, de las Ciencias Sociales y otros campos como: Química, Sicología, Genética, etc., de allí que es importante destacar la enseñanza de la Estadística en todo nivel y se ha considerado menester incluir en éste texto temáticas referentes a la investigación estadística del COMIL - R. (COLEGIO MILITAR "COMBATIENTES DE TAPI") de la ciudad de Riobamba-Ecuador y de la Escuela Superior Politécnica de Chimborazo, ESPOCH.

Aunque se puede aplicar software estadístico libre como el R y de esto hablaremos en otra oportunidad. Los software estadísticos y la hoja electrónica EXCEL; permiten desarrollar destrezas de cálculo y gráfico en el estudiante, ayudando de esta manera a visualizar los argumentos fundamentales de representación, cálculo de índices estadísticos y la construcción gráfica (histogramas, polígonos de frecuencias, diagrama de barras, diagrama de caja (box-plot), diagrama de tallo y hoja (stem and leaf), etc.) de un conjunto de datos como se ve en el segundo capítulo denominado Estadística Descriptiva. También estos software estadísticos generan resultados de ANOVA (análisis de la varianza), regresión, análisis multivariado, econometría, control de calidad, series de tiempo y funciones matemáticas.

1 CAPÍTULO GENERALIDADES

OBJETIVOS

- Presentar argumentos estadísticos de manera atractiva, interesante y aplicativa, generando ambientes que faciliten la investigación
- Motivar al estudio de la Estadística, más que al cálculo, a través de preguntas como ¿Qué es la Estadística? ¿Por qué aprender Estadística?
- Destacar nombres de famosos matemáticos-probabilísticos y estadísticos los que se toman en cuenta en definiciones, propiedades y teoremas.

CONTENIDOS

1.1 RESEÑA HISTÓRICA

La Estadística se estructuró como disciplina científica, en el siglo XIX, pero ya se conocía y utilizaba en la antigüedad. La configuración actual de esta disciplina es el resultado de una evolución, la misma que se puede catalogar en orden cronológico en los siguientes antecedentes:

Las antiguas civilizaciones, como por ejemplo la de Egipto realizaba levantamientos estadísticos (captación de datos, por cierto, de carácter rudimentario), debido a las inundaciones del río Nilo. Se efectuaban anualmente censos, los mismos que permitían distribuir los bienes y repartir las propiedades para que fueran restituidos luego de las inundaciones. También se sabe que los griegos levantaban censos demográficos (registro de nacimientos, muertes, matrimonios, etc.) y de propiedad.

También podemos indicar que en el antiguo Egipto, la Estadística lo aplicaban los faraones y lograron recopilar, hacia el año 3050 antes de Cristo, datos relativos a la población y la riqueza del país. De acuerdo al historiador griego Heródoto, este registro de riqueza y de población se hizo con el objetivo de preparar la construcción de las pirámides. En el mismo Egipto, Ramsés II hizo un censo de las tierras con el objeto de verificar un nuevo reparto.

En el antiguo Israel, la Biblia da referencias en el libro de los Números, de los datos estadísticos obtenidos en dos recuentos de la población hebrea. El rey David por otra parte, ordenó a Joab, general del ejército, hacer un censo de Israel con la finalidad de conocer el número de habitantes de la población.

También los chinos efectuaron censos hace más de cuarenta siglos. Los griegos efectuaron censos periódicamente con fines tributarios, sociales (división de tierras) y militares (cálculo de recursos y hombres disponibles).

En la época del Imperio Romano se aplicaban censos poblacionales y de bienes a los pueblos sometidos al imperio con objeto de aplicar el régimen de impuestos. Pero fueron los romanos, maestros de la organización política, quienes mejor supieron emplear los recursos de la Estadística. Cada cinco años realizaban un censo de la población y sus funcionarios públicos tenían la obligación de anotar nacimientos, defunciones y matrimonios, sin olvidar los recuentos periódicos del ganado y de las riquezas contenidas en las tierras conquistadas. Para el nacimiento de Cristo sucedía uno de estos empadronamientos de la población bajo la autoridad del imperio.

Durante los mil años siguientes a la caída del imperio Romano se realizaron muy pocas operaciones estadísticas, con la notable excepción de las relaciones de tierras pertenecientes a la Iglesia, compiladas por Pipino el Breve en el 758 y por Carlomagno en el 762 DC. Durante el siglo IX se realizaron en Francia algunos censos parciales de siervos. En Inglaterra, Guillermo el Conquistador recopiló el Domesday Book o libro del Gran Catastro para el año 1086, un documento de la propiedad, extensión y valor de las tierras de Inglaterra. Esa obra fue el primer compendio estadístico de Inglaterra.

Aunque Carlomagno en Francia y Guillermo el Conquistador en Inglaterra, trataron de revivir la técnica romana, los métodos estadísticos permanecieron casi olvidados durante la Edad Media. Durante los siglos XV, XVI, y XVII, Leonardo de Vinci, Nicolás Copérnico, Galileo, Neper, William Harvey, Sir Francis Bacon y René Descartes, hicieron grandes contribuciones al método científico, de tal forma que cuando se crearon los Estados Nacionales y surgió como fuerza el comercio internacional, existía ya un método capaz de aplicarse a los datos económicos.

Para el año 1532 empezaron a registrarse en Inglaterra las defunciones debido al temor que Enrique VII tenía por la peste. Más o menos por la misma época, en Francia la ley exigió a los clérigos registrar los bautismos, fallecimientos y matrimonios. Durante un brote de peste que apareció a fines de la década de 1500, el gobierno inglés comenzó a publicar estadísticas semanales de los decesos.

Esa costumbre continuó muchos años, y en 1632 estos Bills of Mortality (Cuentas de Mortalidad) contenían los nacimientos y fallecimientos por sexo. En 1662, el capitán John Graunt usó documentos que abarcaban treinta años y efectuó predicciones sobre el número de personas que morirían de varias enfermedades y sobre las proporciones de nacimientos de varones y mujeres que cabría esperar.

El primer empleo de los datos estadísticos para fines ajenos a la política tuvo lugar en 1691 y estuvo a cargo de Gaspar Neumann, un profesor alemán que vivía en Breslau. Este investigador se propuso destruir la antigua creencia popular de que en los años terminados en siete moría más gente que en los restantes, y para lograrlo investigó 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.

En la época moderna, la técnica censal adquirió un gran desarrollo, llegando a constituirse un eficaz auxiliar de las tareas gubernamentales, es así que en el siglo XVIII en Alemania se enseñaba en las universidades.

Uno de los profesores de la Universidad de Goetingen, Achenwall (1719-1772), fue al parecer quien introdujo la palabra Estadística atribuyendo a este vocablo el siguiente significado: “Ciencia de las cosas que pertenecen al Estado, llamando Estado a todo lo que en una sociedad civil y al país en que ella habita, con todo lo que se encuentra de activo y de efectivo”, entonces la Estadística se ocupa de los fenómenos que pueden favorecer o defender la propiedad del Estado y agrega: **política** enseña cómo deben ser los Estados, **Estadística** explica como son realmente. Achenwall quien utilizó la palabra estadística que en alemán es Statistik, que extrajo del término italiano statista(estadista). Creía, entonces y con sobrada razón, que la nueva ciencia, Estadística, sería el aliado más eficaz del gobernante consciente, para la planificación de los recursos.

La raíz de la palabra estadística se halla, por otra parte, en el término latino status, que significa estado o situación. Indicando la importancia histórica de la recolección de datos por parte del gobierno de un país, relacionados principalmente a información demográfica (censos por ejemplo).

El segundo antecedente histórico, lo encontramos en el siglo XVII. Nos referimos a los trabajos realizados por John Graunt (1620-1674) un comerciante (vendedor de paños) de Londres de modesta preparación, pero dotado de una gran inteligencia, gracias a ella realizó trabajos que le valieron el honor de ser incorporado como miembro de la Sociedad Real.

Graunt utilizando datos demográficos de las parroquias de Londres, logró efectuar estudios que le permitieron descubrir relaciones y leyes demográficas, por inferencias llegó incluso a estimar con buena aproximación la población de Londres y otras ciudades inglesas.

En 1662, John Graunt usó documentos que abarcaban treinta años y efectuó predicciones sobre el número de personas que morirían de varias enfermedades y sobre las proporciones de nacimientos de varones y mujeres que cabría esperar.

Los trabajos de Graunt constituyen el fundamento de los métodos inferenciales actuales que han dado a la Estadística la posibilidad de estudiar los fenómenos colectivos y constituye el capítulo más interesante de esta disciplina. Se considera a Graunt el verdadero precursor de la Estadística de nuestros tiempos.

Paralelamente al desarrollo de la Estadística, como disciplina científica, pero independiente de ésta se desarrolló a partir del siglo XVII, el Cálculo de Probabilidades o Teoría de las Probabilidades.

Sus iniciadores fueron los matemáticos italianos: Gerolamo Cardano y Galileo Galilei y los franceses de ese siglo, particularmente Fermat (1601-1665) y Pascal (1616-1703), quienes iniciaron los estudios de la Teoría de las Probabilidades, tratando de resolver problemas de juegos de azar propuestos por el caballero de Mére Antoine Gombaud.

La Teoría de las Probabilidades llegó a ser pronto popular por sus alusiones a los juegos de azar y se desarrolló rápidamente a lo largo del siglo XVIII. Quienes más contribuyeron a su progreso y a estructurarla como ciencia fueron Jacob Bernoulli (1654-1705) y Abraham de Moivre (1667-1754).

A fines del siglo XVIII y comienzos del XIX los trabajos del matemático francés Pierre de Laplacé (1749-1827) permitieron dar su definitiva estructuración, prácticamente consistía en un análisis matemático de los juegos al azar en sus obras:

Teoría Analítica de la Probabilidad (1812).

En la cual la Teoría de las Probabilidades encontraba una sistematización clásica desde el punto de vista matemático, primera de las modernas teorías axiomáticas.

Ensayo Filosófico Sobre las Probabilidades (1814).

Completó la obra de Bernoulli y de sus continuadores proveyendo a esta nueva disciplina de recursos matemáticos y que la lleva junto a otros matemáticos como Poisson, Gauss, Chebyshev, Markov, Von Mises y Kolmogorov a un grado de perfeccionamiento que la ha hecho apta para las aplicaciones de diversos campos de la ciencia.

No obstante durante cierto tiempo, la Teoría de las Probabilidades limitó su aplicación a los juegos de azar y hasta el siglo XVIII no comenzó a aplicarse a los grandes problemas científicos.

Thomas Bayes (1702 - 1761), fue uno de los primeros en utilizar la probabilidad inductivamente y establecer una base matemática para la inferencia probabilística. Actualmente, con base en su obra, se ha desarrollado una poderosa teoría que ha conseguido notables aplicaciones en las más diversas áreas del conocimiento como la Medicina, Economía, entre otras.

Godofredo Achenwall, profesor de la Universidad de Goetingen, que ya se mencionó acuñó en 1760 la palabra estadística.

Jacques Quételect es quien aplica la Estadística a las Ciencias Sociales. Él interpretó la Teoría de la Probabilidad para su uso en las Ciencias Sociales y resolver la aplicación del principio de promedios y de la variabilidad a los fenómenos sociales. Entre tanto, en el período del 1800 al 1820 se desarrollaron dos conceptos matemáticos fundamentales para la teoría estadística; la teoría de los errores de observación, aportada por Laplace y Gauss; y la teoría de los mínimos cuadrados desarrollada por Laplace, Gauss y Legendre. A finales del siglo XIX, Sir Francis Galton dio forma al método conocido como regresión. De aquí partió el desarrollo del coeficiente de correlación creado por Karl Pearson y otros cultivadores de la ciencia biométrica como J. Pease Norton, R. H. Hooker y G. Udny Yule, que efectuaron amplios estudios sobre la medida de las relaciones. Más adelante, a partir de 1919 la estadística experimental tuvo su desarrollo cuando Ronald A. Fisher asumió la dirección del departamento de Estadística de la Estación Experimental de Rothampstead en Londres, Inglaterra.

A partir de Laplace las dos disciplinas Teoría de las Probabilidades y la Estadística que hasta entonces habían permanecido separadas se fusionaron dé manera que la Teoría de las Probabilidades se constituyó en el andamiaje matemático de la Estadística.

Conjuntamente a la Teoría de las Probabilidades se desarrolló la Teoría de los Errores especialmente por obra de Gauss, Bessel y el propio Laplace quienes llegaron a establecer los Mínimos Cuadrados Ordinarios, MCO, como procedimiento matemático para resolver el problema fundamental de la Teoría de los Errores.

La Teoría de los Errores es un valioso antecedente de la Estadística, pues constituye la primera rama de la Estadística que pudo construirse con una estructuración teórico-matemático.

Los capítulos más importantes de la Estadística moderna son: Análisis Exploratorio de Datos (AED), la Teoría de la Correlación Regresión, la teoría de las Muestras, la teoría de las Series de Tiempo, Econometría, Control Estadístico de la Calidad, Estadística No Paramétrica, Procesos Estocásticos o aleatorios, Estadística Espacial.

Al avanzar el siglo XX, la obra del estadístico inglés Karl Pearson (1857-1936) tuvo destacados continuadores entre los que sobresale Ronald Fisher (1890-1962) científico británico inventor del método de máxima verosimilitud, del análisis de la varianza y del diseño estadístico de experimentos seguramente la figura más prominente de la Estadística de todos los tiempos.

Para destacar el punto fundamental de su obra diremos que si Pearson fue el iniciador de la teoría de la Estadística Inferencial, Fisher fue quien la desarrolló y estructuró en forma rigurosa, con la colaboración de sus discípulos; en particular la teoría de las pequeñas muestras y la de estimación, adquieren con Fisher la estructuración actual.

1.2 ALGUNA TERMINOLOGÍA NECESARIA

Estadística. Se refiere a la técnica de recolección, representación, procesamiento, análisis, modelación e interpretación de un conjunto de datos en el ámbito de la incertidumbre todo con el fin de tomar decisiones.

Figura 1. Clasificación de la Estadística

La Estadística cumple dos funciones:

La de análisis descriptivo en forma de tablas, gráficas y números de las características observadas por lo general de la muestra, y la de estadística inferencial o inducción, lográndose a través de ésta generalizaciones para un grupo mayor denominado población, partiendo de un grupo menor llamado muestra.

Población. Es un conjunto de medidas o el recuento de todos los elementos o individuos que presentan una característica común. El término población se usa para denotar el conjunto de elementos del cual se extrae la muestra.

Los elementos que integran la población o la muestra pueden corresponder a personas, animales, objetos o cosas.

Además, el elemento puede ser una entidad simple (un estudiante) o una entidad compleja (un curso), y se denomina unidad investigada. Es importante resaltar el hecho de que a pesar de encontrarse una población constituida por un grupo de elementos, a la Estadística no le interesa el elemento en sí, sino su característica.

Características (o caracteres). Corresponden a ciertos rasgos, cualidades o propiedades que poseen los elementos que constituyen la población o la muestra. Algunos caracteres son mensurables y se describen numéricamente denominándose caracteres cuantitativos; otros se expresan mediante palabras, símbolos (o números) por no ser mensurables se denominan caracteres cualitativos o atributos, o categóricos.

Muestra. Se define como un conjunto de medidas o el recuento de una parte de los elementos pertenecientes a una población. Los elementos se seleccionan aleatoriamente, es decir, todos los elementos que componen la población tienen la misma posibilidad de ser seleccionados.

Para que una muestra sea representativa de la población se requiere que las unidades sean seleccionadas al azar, ya sea utilizando el sorteo, las tablas de números aleatorios, la selección sistemática o cualquier otro método que sea el azar.

Estadístico. Es la persona que trabaja en la elaboración y análisis de estadísticas.

Estadísticas. Se refiere a un ordenamiento sistemático de datos presentados en forma de cuadros y gráficas, de nùmeros. En otras palabras, las estadísticas son datos agrupados metódicamente y consignados en publicaciones, elaboradas por las diversas empresas o entidades, buscando sean conocidos por los interesados.

Estadísticas primarias. Son aquellos datos obtenidos ya sea por encuestas directas, mediante la utilización de cuestionarios, o como resultado de la observación directa; esta utilización es una técnica en estudios de carácter científico.

Estadísticas secundarias. En éstas los datos se obtienen de publicaciones, las cuales pueden ser reproducciones totales o parciales. Son fuentes valiosas utilizadas en cualquier tipo de investigación.

Estadísticas temporales. Denominadas series de tiempo o series cronológicas. Son las obtenidas y ordenadas en forma cronológica, siendo el resultado de investigaciones u observaciones periódicas: días, meses, años u otro espacio de tiempo. Cuando las investigaciones son aisladas, es decir, no presentan periodicidad continuada, las estadísticas se llaman atemporales.

Las estadísticas se pueden clasificar como internas y externas.

Estadísticas internas de una institución educativa se originan de los registros internos, tales como promedios, matriculas, pensiones, pesos, estaturas, edades y otros de los estudiantes.

Estadísticas externas son registros originados fuera de la institución educativa; por ejemplo: opinión de la ciudadanía respecto al prestigio, pensiones de la competencia, etc.

Parámetros. Son todas aquellas medidas que describen numéricamente la característica de una población. También se denomina valor verdadero, ya que una característica poblacional tendrá un solo parámetro (media, proporción, varianza, etc.). Sin embargo una población puede tener varias características y por tanto, varios parámetros. Generalmente son cantidades constantes y desconocidas.

Estimador (puntual). La descripción numérica de una característica correspondiente a la muestra, se denomina estimador puntual o estadígrafo como por ejemplo el promedio o media muestral, varianza muestral, proporción muestral, etc.

Por lo general, los estimadores son variables y conocidas. Existe una diferencia entre el estimador y el parámetro, denominado error, es aconsejable utilizar el estimador por intervalos, dentro del cual deberá estar el parámetro con cierto margen de seguridad o nivel de confianza, que hablaremos en el capitulo 4.

1.3 ¿QUÉ ES LA ESTADÍSTICA?

La Estadística es una disciplina científica. Al respecto tengamos presente las siguientes consideraciones: el New Collegiate Dictionary de Webster define la Estadística como una rama de las Matemáticas que trata de la recopilación, el análisis, la interpretación de un conjunto de datos. Por otro lado Kendall y Stuart afirman: la Estadística trata de los datos reunidos al contar o medir las propiedades de algún experimento.

Fraser, al comentar sobre la experimentación y las aplicaciones estadísticas dice, “la Estadística trata con métodos para obtener conclusiones a partir de los resultados de los experimentos o procesos”. En fin como nos damos cuenta el aspecto más importante de ésta disciplina es la obtención de conclusiones basadas en los datos experimentales, a este procedimiento se lo conoce con el nombre de inferencia estadística.

Para comprender la naturaleza de la inferencia estadística, es necesario entender las nociones de población y muestra, dadas anteriormente, pero sin embargo debemos acotar también que una población es una colección finita de mediciones, o una colección grande, virtualmente infinita de datos acerca de algo de interés. Por ejemplo “número de hijos por familia de la ciudad de Riobamba”.

Por otra parte la muestra es un subconjunto representativo seleccionado de una población. Por ejemplo “65 familias seleccionadas de la ciudad de Riobamba”. Con la muestra, el objetivo no consiste en examinarla, sino en estudiar la población a través de ella.

Las palabras algo de interés del concepto de población se pueden entender e incluso caracterizar por dos conjuntos: el de los individuos y el de los caracteres.

El término individuo puede designar, según el caso: el estudiante de un colegio, una familia, un animal, etc., es la entidad de base sobre la cual el observador realiza la toma o las medidas de los datos. Y los caracteres respectivamente pueden ser: calificaciones, grado o curso, edad del estudiante, estatura del estudiante, etc.; condición económica: ingresos y egresos; número de hijos de una familia de la ciudad de Riobamba, etc.; peso del animal, raza, edad del animal, etc.

Entonces ¿qué es realmente la Estadística? respondemos.
La Estadística es una rama de las Matemáticas que trata:

- Recopilación
- Representación
- Análisis
- Interpretación

de un conjunto de datos en un ambiente de incertidumbre para ayudar a tomar decisiones (poder hacer comparaciones y sacar conclusiones).

Entonces podemos decir, “los datos por si solos son inertes sin ninguna utilidad. Adquieren valor únicamente cuando son recopilados, representados, analizados (modelados) e interpretados y se convierten en información útil y confiable para la toma de decisiones.

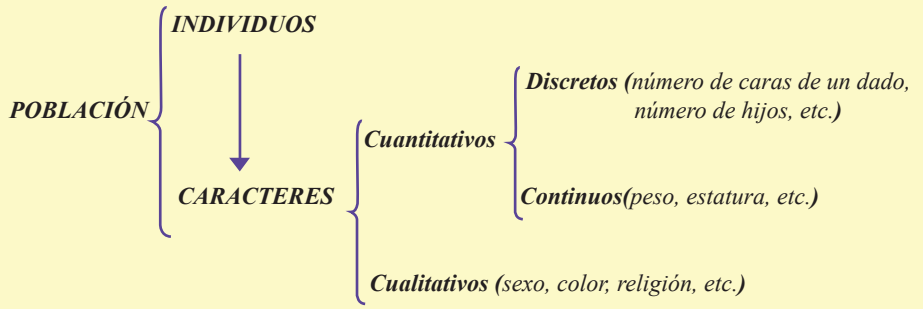
Observación. Se observe que si bien es cierto, para estudiar (la población) el colectivo se requiere de información individualizada (de los individuos), las conclusiones que se obtienen de la investigación estadística no se refiere a cada elemento individualmente, sino al conjunto de los individuos considerados como grupo.
Pues se debe tener en cuenta siempre que la Estadística estudia el comportamiento de los fenómenos de grupo, prescindiendo de aquellos fenómenos individuales que pueden ser considerados como resultados de casos aislados. Ver la actividad de aprendizaje 1.c del párrafo 2.4

Los caracteres son las características de los individuos los mismos que son mensurables cuantitativamente o cualitativamente. Llamamos carácter cuantitativo aquella modalidad numérica, cuyos valores se toma sobre un conjunto finito o infinito numerable, o sobre un subconjunto de números reales.

De acuerdo a esta descripción estos caracteres se subdividen en discretos (naturales, enteros o racionales) y continuos (la recta real numérica, un intervalo o un segmento de la recta), por ejemplo son caracteres discretos: el número de estudiantes de un colegio, el número de hijos de una familia, el número de personas de la cola frente a una ventanilla, el número de estudiantes que asisten normalmente al programa de doctorado en Ciencias de la Educación con mención Enseñanza de la Matemática, etc., y son caracteres continuos: el peso, la estatura de los estudiantes, el salario de un jefe de familia o el tiempo de duración de una persona de la cola frente a una ventanilla, etc..

Se conoce como carácter cualitativo aquel que toma modalidades no numéricas por ejemplo: sexo, profesión, religión, candidato, etc.; a los cuales es posible establecer un nivel jerárquico o un nivel de satisfacción asignándoles un valor; por ejemplo al carácter sexo de un individuo se dan los valores: 1 a hombre y 0 a mujer o viceversa. Se puede presentar lo dicho anteriormente en el siguiente esquema:

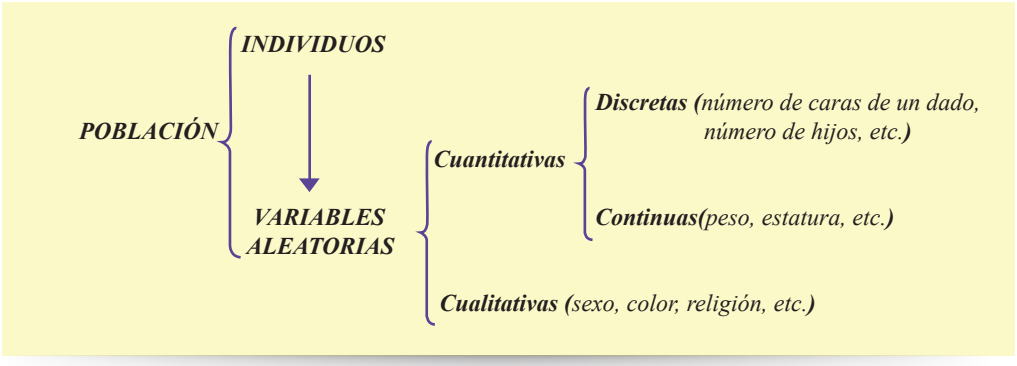
Figura 2. Esquema de Caracteres



Pero, una población (o las características de una población) puede ser analizada (o pueden ser analizadas) a través de una o varias variables aleatorias. Entonces ¿qué es una variable aleatoria? Una variable aleatoria denotamos por v.a.

Definición. Si un caracter es observado sobre una parte de la población, es decir, sobre una muestra y los individuos observados son elegidos al azar, entonces el caracter se denomina variable aleatoria (v.a.) por lo que una v.a. puede ser: cualitativa y cuantitativa discreta o cuantitativa continua.

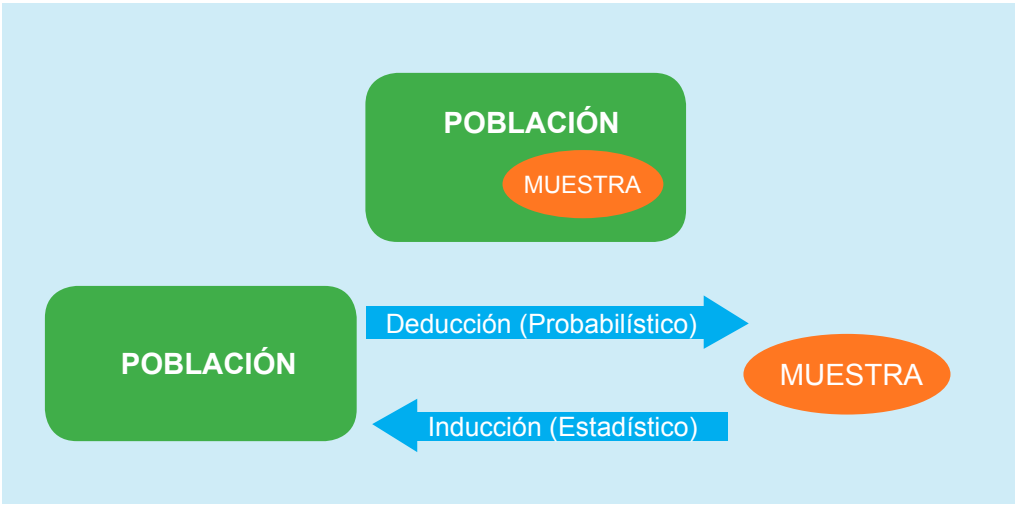
Figura 3. Esquema de Variables



Tengamos también en cuenta que la metodología para hacer inferencias se apoya en la Teoría de las Probabilidades. El razonamiento del especialista en probabilidad parte de una población conocida, para deducir el resultado de un experimento, la muestra. Los censos demográficos son actividades del probabilístico. Al contrario, el estadístico utiliza las probabilidades para calcular la probabilidad de una muestra observada y de ésta hacer inferencias(o sea sacar conclusiones) o de los resultados de la muestra induce con respecto a las características de una población desconocida. Las encuestas son actividades típicas que realiza el estadístico.

Así, la Teoría de las Probabilidades es la herramienta de la Estadística. Esto se puede observar en la siguiente figura:

Figura 4. Población y muestra



En Estadística la inferencia es inductiva, porque se parte de lo específico que es la muestra, hacia lo general que es la población. En un procedimiento de esta naturaleza siempre existe la probabilidad de error. Nunca podrá tenerse el 100% de seguridad sobre una proposición que se base en la inferencia estadística. Siendo así. ¿Qué es lo que hace a la Estadística ciencia? ; es que unida a cualquier proposición existe una medida de confiabilidad de ésta.

En Estadística la confiabilidad se mide en términos de probabilidad, es decir, para cada inferencia estadística se identifica la probabilidad de que la inferencia sea correcta.

1.4 DIVISIÓN DE LA ESTADÍSTICA

La Estadística para su mejor estudio se ha dividido en tres grandes ramas: Estadística Descriptiva, Teoría de las Probabilidades y Estadística Inferencial.

Estadística Descriptiva consiste en la presentación de datos en forma de tablas y gráficas. Esta comprende cualquier actividad relacionada con los datos y está diseñada para resumir o describir los mismos, sin factores pertinentes adicionales; esto es, sin intentar inferir nada que vaya más allá de los datos, como tales.

Es en general utilizada en la etapa inicial de los análisis, cuando se tiene contacto con los datos por primera vez.

Teoría de las Probabilidades puede ser pensada como la teoría matemática utilizada para estudiar la incertidumbre originada de fenómenos de carácter aleatorio, o sea, producto del azar.

Estadística Inferencial se deriva de muestras, de observaciones hechas sólo acerca de una parte de un conjunto numeroso de elementos y esto implica que su análisis requiere de generalizaciones que van más allá de los datos. Como consecuencia, la característica más importante del reciente crecimiento de la Estadística ha sido un cambio en el énfasis de los métodos que describen a métodos que sirven para hacer generalizaciones. La Estadística Inferencial investiga o analiza una población partiendo de una muestra tomada de ella.

La Estadística Descriptiva y la Inferencial comprenden ***la Estadística Aplicada.*** Hay también una disciplina llamada ***Estadística Matemática***, la cual se refiere a las bases teóricas de la materia, e incluye el estudio de las probabilidades, es decir es la Estadística Aplicada más la Teoría de las Probabilidades.

Otra división de la Estadística es:

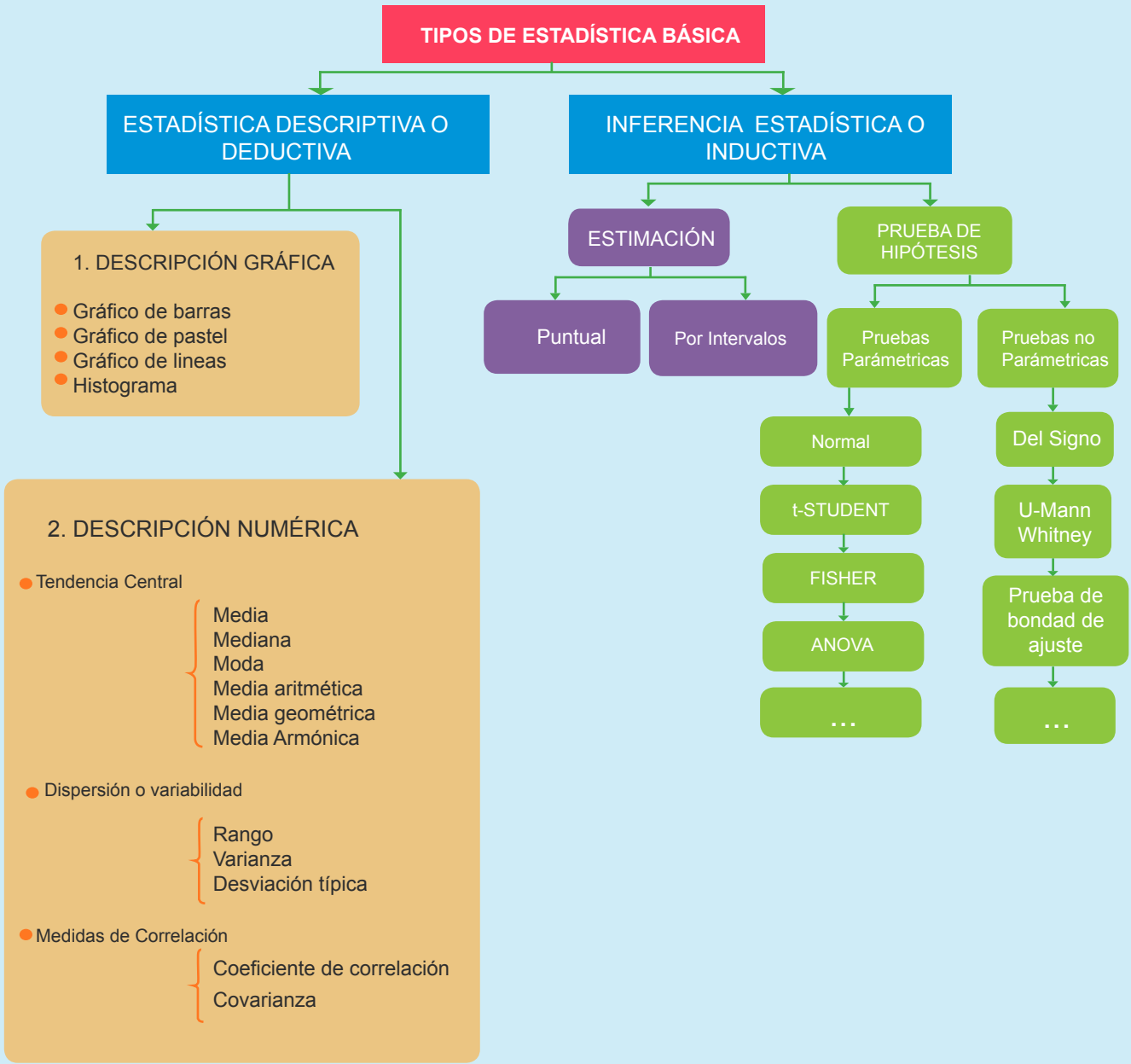
Estadística Paramétrica: en la estadística paramétrica nuestro interés es hacer estimaciones y pruebas de hipótesis acerca de uno o más parámetros de la población. Además, en todas estas estimaciones y pruebas de hipótesis se establece como suposición general que la población o poblaciones de donde provienen las muestras, deben estar distribuidas normalmente, aunque sea en forma aproximada, deben tener la misma variabilidad (homocedasticidad).

Estadística No Paramétrica: estudia las pruebas y modelos estadísticos cuya distribución subyacente no se ajusta a los llamados criterios paramétricos. Su distribución no puede ser definida a priori, pues son los datos observados los que la determinan.

La utilización de los métodos no paramétricos se hace recomendable cuando no se puede asumir que los datos se ajusten a una distribución normal o cuando el nivel de medida empleado no sea, como mínimo, de intervalo. Ver escalas o niveles de medida, que esta al final del capítulo.

Nota. Nótese que en una de las tareas a los estudiantes del cuarto nivel de la carrera de Ingeniería en Estadística Informática de la ESPOCH propusimos la división o tipos de Estadística que ellos han visto hasta el momento y el resultado se presenta a continuación:

ESQUEMA DE CURSO-TALLER
TÉCNICAS Y HERRAMIENTAS DE ESTADÍSTICA PARAMÉTRICA Y NO PARAMÉTRICA



1.5 MUESTREO

En la actualidad, lo más utilizado es el muestreo, por su menor costo, mayor rapidez y menor número de personas que intervienen en la investigación. Existe más de un método de muestreo y se destaca algunos aspectos por cada método:

a) Grado de precisión requerida para los estimadores.
b) Tamaño de muestra.
c) Costo y tiempo.

Muestreo probabilístico. Dentro de éste método existe algunos procedimientos como:

Muestreo aleatorio simple. Este método permite que la selección de todos los individuos o elementos que constituyen la población tenga la misma posibilidad de ser incluidos en la muestra.

Cada elemento que constituye la muestra puede haber sido seleccionado una sola vez, lo que generalmente ocurre, denominándose extracciones sin reposición; en otras ocasiones, cada elemento puede ser elegido más de una vez en la muestra, situación que puede ocurrir cuando la población es pequeña, en este caso se dice que las extracciones son realizadas con reposición.

La elección se puede realizar por sorteo o utilizando las tablas de números aleatorios, siendo esta última la más aconsejable, ya que han sido elaboradas con el fin de facilitar la selección, ahorrando con ello tiempo y dinero (ver tabla de números aleatorios del apéndice).

Para la aplicación de estas tablas se procede de la forma siguiente:

a) Se enumeran las unidades que conforma la población partiendo desde 01 hasta 99, desde 001 hasta 999 y así sucesivamente, dependiendo del tamaño de la población;

b) Se determina un punto de la tabla desde el cual se comenzarán a seleccionar las cifras de dos, tres o más dígitos, dejando establecido si esta selección se hace en forma horizontal o vertical;

c) Los números seleccionados deberán corresponder con los de la población por muestrear, descartando los números superiores al tamaño de la población.

ACTIVIDAD DE APRENDIZAJE DESARROLLADA **01**

Se quiere seleccionar aleatoriamente 5 cadetes del COMIL-R para que representen en la inauguración y exaltación al monumento Simón Bolívar. Si los matriculados en este plantel son 741.

Solución

Siguiendo las instrucciones se enumeran los cadetes como 001, 002,..., 741, luego de la tabla de números aleatorios (Apéndice: Tabla 7), se tomarán desde la primera columna, primera fila y en forma vertical, los tres primeros dígitos de cada número, obteniendo en la muestra los cadetes que representan a los siguientes números:

638, 029, 068, 200 y 513

En este caso se omite el quinto número (866) por cuanto no pertenece a los elementos enumerados en la población.

De esta manera, los cadetes numerados con estas cifras serían los seleccionados para asistir a la inauguración y exaltación al monumento Simón Bolívar.

Muestreo aleatorio estratificado. Llamado también muestreo aleatorio restringido, es aquel donde la población se estratifica, es decir, se forman grupos, en tal forma que el elemento tendrá una característica que sólo le permitirá pertenecer al mismo. Este proceso se realiza cuando la población es heterogénea, presentando una gran variabilidad, siendo por tanto, un diseño más eficiente que el muestreo aleatorio simple, con la ventaja de que se puedan utilizar muestras mucho más pequeñas. Mediante la selección aleatoria en cada estrato se conformará la muestra. Debe para su selección considerar los siguientes casos: Igual tamaño. Cuando los elementos que constituyen la muestra se reparten por igual en los diferentes estratos muestrales.

1. ***Proporcionales.*** Los elementos se distribuyen en los estratos muestrales proporcionalmente al tamaño de los mismos en la población.

2. ***Óptima.*** Cuando el tamaño de la muestra depende del grado de variabilidad en cada estrato poblacional y del costo de investigación.

El objetivo de la estratificación es formar estratos (grupos o clases) de tal forma que haya alguna relación entre estar en un estrato particular y la respuesta que se busca en el estudio estadístico y que en los estratos separados haya tanta homogeneidad (uniformidad) como sea posible.

Usamos para seleccionar una muestra de tamaño n de una población que ha sido estratificada en k estratos (dividida en k grupos) seleccionamos tamaños de muestra para la distribución proporcional de cada estrato mediante la fórmula:

Donde $n_i = \frac{N_i}{N} n, \qquad i = 1,2, \ldots, k$

$$n = n_1 + n_2 + \ldots + n_k$$

N_i son los tamaños de cada estracto i

N es el tamaño de la población

ACTIVIDAD DE APRENDIZAJE DESARROLLADA 02

Se debe tomar una muestra estratificada de tamaño n = 30 en una población de tamaño 2000, que consta de tres estratos de tamaños $N_1 = 1000$; $N_2 = 600$ y $N_3 = 400$. ¿Si la distribución debe ser proporcional, cuán grande debe ser la muestra tomada de cada estrato?

Solución:

Aplicando la fórmula, obtenemos N = 1000+600+400 = 2000

$$n_1 = (1000/2000)30 = 15$$
$$n_2 = (600/2000)30 = 9$$
$$n_3 = (400/2000)30 = 6$$

Este ejemplo ilustra la distribución proporcional, pero se puede agregar otras maneras de distribuir proporciones de una muestra entre los diferentes estratos. Una de esta es conocida como la ***distribución óptima***, o de Neyman, en la que no sólo se maneja el tamaño del estrato, sino que también maneja la variabilidad (o cualquier otra característica) del estrato. La fórmula que se aplica para el tamaño de los estratos es

$$n_i = \frac{nN_i\sigma_i}{N_1\sigma_1 + N_2\sigma_2 + \ldots + N_k\sigma_k}$$

Donde σ_i es la desviación estándar (medida de dispersión) del estrato i.

Muestreo sistemático. Este método de selección es utilizado por algunos contadores para revisar sumas, cuentas, etc., y consiste en determinar en primer lugar un intervalo igual al valor obtenido al dividir el tamaño de la población por el de la muestra. Luego se toma aleatoriamente una observación. Supongamos que entre el 01 y 10 se seleccionó la observación 7 y como el intervalo es 10, la segunda observación será la 17, luego la 27, y así sucesivamente.

Muestreo no probabilístico. En el muestreo no probabilístico se toma la muestra de cualquier tamaño y los elementos son seleccionados de acuerdo a la opinión o juicio que tenga el investigador sobre la población.

En el caso de una población homogénea, la representatividad de tal muestra puede considerarse satisfactoria. Por lo general, los individuos son seleccionados por conveniencia, por capricho o por cuotas, por tal razón no ofrecen confiabilidad alguna.

Por ejemplo el director de una escuela selecciona a los estudiantes de mejor rendimiento académico en la asignatura de Matemática para que representen a la misma en el concurso de Matemática organizado por la Dirección de Estudios de la localidad.

1.6 ¿POR QUÉ APRENDER ESTADÍSTICA?

Al respecto se debe considerar que los medios de comunicación traen cotidianamente resultados de sondeos seleccionados sobre muestras de carácter social (censos), político (elecciones), económico (producción de petróleo o de banano), y de otros aspectos no menos importantes. Como los mismos se deben interpretar cuando se presentan sea de forma esquemática, sea de forma gráfica, es útil conocer o disponer de algún instrumento que permita un análisis crítico frente a tales mensajes. Entonces es oportuno que tengamos presente los siguientes objetivos:

- *Adquirir la capacidad de leer correctamente los diferentes gráficos estadísticos.*
- *Conocer los índices estadísticos más útiles a fin de comparar críticamente los resultados.*
- *Entender la necesidad y la dificultad de las investigaciones sobre el muestreo.*
- *Poseer el mínimo instrumento teórico-probabilístico a fin que se puedan estudiar fenómenos no determinísticos simples o complejos.*
- *Valorizar cuantitativamente la probabilidad de un evento según la definición clásica, esto es como un cociente entre casos favorables y casos totales.*
- *Estimar cualitativamente la probabilidad de un evento aleatorio.*
- *Aplicar las técnicas estadísticas de la inferencia para resolver problemas que atañen a la labor educativa, social u otro ámbito.*

La finalidad de este texto de Estadística Aplicada a la Educación con Actividades de Aprendizaje es exponer la información de cualquier departamento académico por ejemplo la del COMIL-R de la ciudad de Riobamba u otra institución educativa como la ESPOCH con el objeto de que nos permita:

- *Tener una visión general de la institución educativa en su conjunto, para que los directivos puedan formular directrices con pleno conocimiento de causa, etc.*
- *Descubrir las relaciones de causa y efecto en las diversas manifestaciones académicas (rendimientos individuales, por paralelos, cursos e institucional), pedagógicas (incidencia del sexo en el aprendizaje, establecer diferencias entre los aspirantes a la hora de calificar y los profesores), sociales (relación entre el status social-económico y los ejes transversales que se cultivan en los estudiantes), etc.*
- *Detectar casos problemas de conducta y bajo aprovechamiento observando las fluctuaciones individuales, de curso, de la institución con las condiciones externas. (Familia y sociedad) para tener una mayor orientación e información en la actividad educativa.*

1.7 INVESTIGACIÓN ESTADÍSTICA

La *Estadística es una herramienta básica en la investigación* de cualquier campo de la ciencia, y su aplicación dependerá de la *facilidad y disponibilidad de los datos que se analicen* y de la *naturaleza de los fenómenos o experimentos que se deseen estudiar.*

La investigación se puede clasificar de acuerdo a la facilidad de los datos en:

Investigación interna. Al contar con la información, algunas veces recopilada sin ninguna metodología, ésta no será suficiente en la realización de una investigación sino también que requiere organizar la información de tal forma que permita la aplicación de métodos estadísticos a fin de obtener conclusiones válidas.

Dentro de una institución se originan una serie de fenómenos, como por ejemplo en una unidad educativa los datos recopilados por el departamento académico, los mismos que deben ser organizados en tal forma que faciliten el análisis y su comparación con períodos anteriores.

En el caso de los promedios o número de estudiantes los valores obtenidos permitirán hacer comparaciones entre trimestres o quimestres o cualquier otro periodo y éstos a su vez facilitarán el análisis para el rendimiento académico general de dicha unidad.

Investigación externa. Los valores de una institución no solo se analizan con datos internos bien organizados, sino comparándolas con instituciones, similares, de la competencia.

Si el objeto de la investigación es establecer la posición relativa de la institución educativa en la sociedad y en especial conocer la tendencia de los clientes (estudiantes - padres de familia), el comportamiento actual o futuro, en relación con la calidad de la enseñanza, dependerá de pensiones, propaganda, etc. En estos casos es indispensable la investigación externa, a fin de obtener la información necesaria, que no se da en la investigación interna.

Investigación exhaustiva. Se llama así a aquella investigación donde se observan todos los elementos que constituyen la población objetivo. Si vamos a investigar todos los hogares de los estudiantes del COMIL-R, prácticamente se está desarrollando una labor censal.

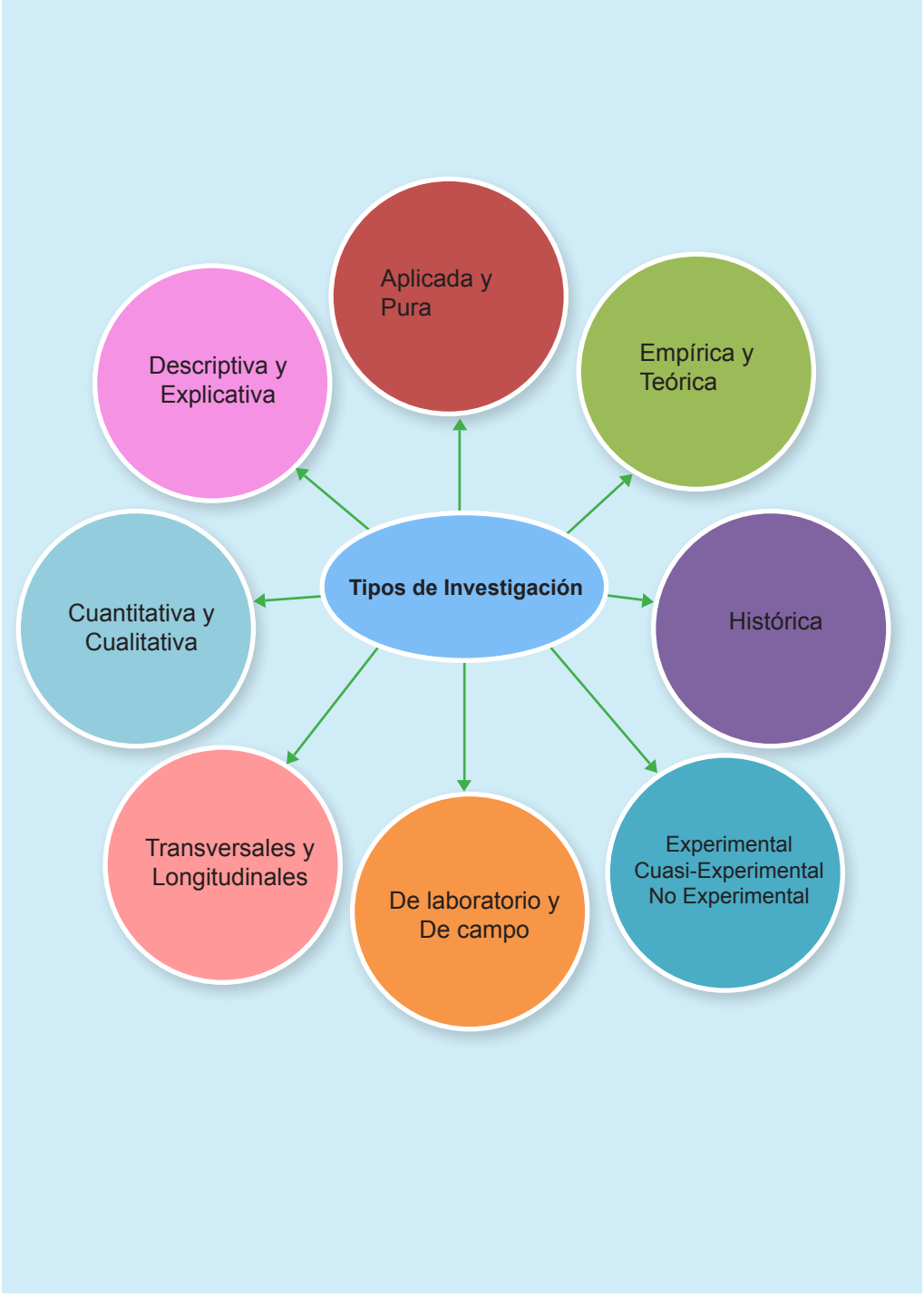
Sin embargo, la población puede referirse a la totalidad de hogares en una zona de la misma ciudad de Riobamba, o puede referirse a los hogares de un cierto barrio. Como se ve, la población constituye todas aquellas unidades que son objeto de estudio. Es decir, la población de algo de interés. Por ejemplo los censos son investigaciones de este tipo.

Por lo general, toda investigación que no sea exhaustiva es parcial y ésta limitación está siempre encaminada a facilitar su ejecución, minimizar tiempo y optimizar recursos humanos o económicos.

Investigación parcial. Se realiza cuando no es posible una investigación exhaustiva y sólo se observa una parte de los elementos o unidades que constituyen la población objetivo, denominándose muestra. Con la muestra, el objetivo no consiste solamente en examinarla, sino también en estudiar la población a través de ella.

La selección de un grupo de hogares de los estudiantes de la sección básica o del bachillerato son ejemplos de muestras tomadas de la población de todos los hogares de una unidad educativa.

La clasificación puede obedecer también a la naturaleza de los fenómenos o experimentos que se deseen estudiar o criterios de problemas científicamente planteados, de acuerdo a esto, exponemos a continuación de manera esquemática los *tipos de investigación.*



1.8 ETAPAS DE UNA INVESTIGACIÓN

Se requiere una investigación de carácter estadístico cuando no se tiene un buen flujo de información que permita que dicha información se organice y condense y por lo general, se encuentra dispersa. Se pueden considerar tres clases de operaciones o etapas de manera general en una investigación: *planeamiento, recolección de datos y análisis de datos.*

1) Planeamiento

Al trazar un plan de investigación, se debe definir y organizar cada una de las actividades necesarias para llevar a cabo el trabajo y poder alcanzar los objetivos propuestos. Dentro de la etapa de planeamiento se podrá considerar ciertos aspectos que se presentan, donde el orden y la necesidad de cada uno de ellos dependerán de la misma naturaleza de la investigación.

Objeto de la investigación

Antes de iniciar cualquier proceso de la investigación, se hace indispensable identificar con claridad y precisión el fin que se propone, formulando el problema en tal forma que nos permita los objetivos generales y los específicos y a ser posible, una jerarquización de los mismos.

En esta etapa se debe contestar los siguientes interrogantes:

a) ¿Qué se va investigar?
b) ¿Cómo se va a realizar la investigación? Se refiere a los medios y condiciones con las cuales se debe realizar el estudio.
c) ¿Cuándo se realiza? El momento en que debe hacerse la observación.
d) ¿Dónde se realiza? El lugar, zona o región donde se hará la investigación.

Con las respuestas a estas interrogantes se sabrá cuál es la población objetivo o de algo de interés que se va investigar, qué tipo de datos se requerirán, el tipo de informante necesario, la dificultad para hacer la observación, número de cuestionarios, tiempo y costo de la investigación, etc.

Unidad de investigación

La unidad es la fuente de información, es decir, a quien va dirigida la investigación, la cual puede ser un estudiante, un curso, una institución, o una persona, una familia, una vivienda, etc., y su selección depende del objeto de la investigación.

La unidad debe ser clara, en tal forma que sea entendida por todos; además, adecuada al tipo de investigación, mensurable, que permita ser medible, y comparable con los resultados obtenidos en investigaciones similares.

Juntamente con la unidad estadística principal se presenta, con mucha frecuencia, la necesidad de establecer otras unidades denominadas secundarias, por ejemplo si se desea investigar en una institución educativa la deserción o el bajo rendimiento académico, un estudiante es la unidad de estudio principal mientras la familia, los amigos de éste, son las unidades secundarias.

Clase de estudio

En primer lugar hay que determinar qué tipo de investigación se va a realizar:

- Investigaciones descriptivas y experimentales (Investigaciones de laboratorio y de campo).
- Investigaciones explicativas y analíticas.
- Investigaciones empíricas y teóricas.
- Investigaciones puras y aplicadas.
- Investigaciones transversales y longitudinales.
- Investigaciones cuantitativas y cualitativas.
- Investigación histórica.

La distinción entre la *investigación descriptiva y analítica,* en algunos casos no es muy clara. Se dice que la primera es la de obtener información con respecto a grupos, en cambio en la analítica permite establecer ciertas comparaciones y sobre todo la verificación de hipótesis.

En la investigación experimental es una situación provocada por el investigador, en condiciones controladas, cuya finalidad es conocer por qué causa se produce un caso particular.

La investigación explicativa busca la causa de un fenómeno a través de su explicación por medio de leyes.

Las investigaciones transversales se realizan en un momento determinado. Por ejemplo, si se hace un estudio sobre los factores que afectan la eficacia laboral de los administradores de la educación de la región central en el sistema educativo ecuatoriano, interesa la situación fundamentalmente en el momento mismo del estudio, no antes ni después. Por su parte, las investigaciones longitudinales se realizan a través del tiempo, de manera que interesan los resultados de un fenómeno o situación dada después de un determinado tiempo.

Un estudio longitudinal consistiría en analizar los factores que afectan la eficacia laboral después de una huelga o de un curso de capacitación o cualquier otro evento y estudiar factores en la nueva situación después de algún tiempo.

Conceptualizando la investigación cuantitativa como la clásica o tradicional, dentro de lo cual se ubica la mayoría de los tipos de investigación presentados anteriormente. En tanto que la investigación cualitativa, se ha concebido como aquel tipo de investigación en el cual participan los individuos y comunidad para solucionar sus propias necesidades y problemas, es una forma moderna de investigar a través de un proceso permanente de interacción y retroalimentación de sus distintas etapas.

La investigación histórica, pretende conocer experiencias pasadas sin tergiversar los hechos y condiciones reales de la época a través de la reunión, examen, selección, verificación y clasificación de los hechos y su adecuada interpretación.

Examen de la documentación y metodología

Es importante determinar si la investigación ha sido con anterioridad tratada, con el fin de prescindir del estudio; averiguar si se cumplió el objetivo propuesto y sí la información está actualizada. En caso contrario, habrá necesidad de realizarla, tratando de solucionar las dificultades que se presentaron en la anterior, en razón, a un mayor conocimiento sobre la población objetivo y además procurando un mejoramiento en la metodología utilizada.

Método de observación

Una vez planteado el objetivo de la investigación, definida la unidad o unidades y determinado que el estudio no fue realizado o que los datos que se tienen requieren actualización, se debe decir el método que se empleará, es decir, si se va a decidir la población en su totalidad o solo una muestra.
El primer caso lo hemos denominado investigación exhaustiva, enumeración completa o censo y el segundo, muestreo. La elección de uno de los métodos censo o muestra, depende entre otros factores, de:

- Tiempo disponible.
- Recursos humanos.
- Recursos financieros.
- Finalidad de la investigación.
- Número de unidades que componen la población.
- Caracteres por investigar.
- Si el elemento que se toma se puede destruir o no en el proceso de medición de la característica.
- El grado de variabilidad.

2) Recolección de datos

Las encuestas se pueden realizar por correo, entrega personal del cuestionario, entrevista, panel, observación directa, teléfono, otros.

Las encuestas por correo tienen algunas ventajas, tales como las de ser poco costosas, ya que el valor de recolección corresponde al valor del envío y retorno del cuestionario. Estas ventajas en el uso de correo, son las mismas que en la entrega personal del cuestionario, agregándose la reducción del extravío del cuestionario. Se presenta a continuación una encuesta a los estudiantes o cadetes de la escuela de la Unidad Educativa del COMIL-R donde se indica el propósito general de esta investigación: *mejorar la educación*.

Ambos procesos: encuesta personal o por correo (cuestionario) presentan así mismo desventajas: extravío del cuestionario, la no-devolución, falta de contestación a determinadas preguntas, demora en la devolución, uso de abreviaturas, preguntas mal respondidas, etc.

La entrevista es un buen proceso de recolección ya que permite recoger el mayor número de cuestionarios, se obtienen respuestas a todas las preguntas, se aclara las dudas del informante, pero su mayor desventaja radica en el costo pues requiere más tiempo y de más recursos económicos. Además las respuestas pueden ser influenciadas por el entrevistador.

La observación puede ser directa como su nombre lo indica, la recolección de los datos se hace observando directamente el hecho.

Es indirecta cuando la tarea de recolección consiste en corroborar los datos que otros han observado.

La encuesta por teléfono se emplea de preferencia para estudios de radio y televisión cuando se requiere determinar la sintonía en el momento de comunicar y las preguntas van encaminadas hacia lo que se ve o escucha.

Para la elaboración del cuestionario se debe considerar los siguientes aspectos técnicos:

- Se incluya preguntas únicamente indispensables.
- Las preguntas deben ser claras, concisas y comprensibles para quién las hace y quién las responda.
- Las preguntas deben ordenarse, comenzando con las fáciles y terminando con las difíciles.
- No se debe emplear abreviaturas.
- La pregunta debe ser de tal calidad que siendo formulada en lenguaje corriente, atienda a la técnica de investigación.

EJEMPLO DE ENCUESTA

Colegio Militar No.6 "Combatientes de Tapi"
Departamento Academico
Sección Estadística
Encuesta a cadetes de la escuela

Estimado niño(a) queremos mejorar tu educación. Colabora contestando con una **X** *en el cuadro que se indica en la presente encuesta con la seriedad y la honestidad que te caracteriza. Recibe nuestro agradecimiento por tu apoyo.*

1. ¿Te gusta la forma de trabajar de tu Maestro(a)?
Mucho ☐ Poco ☐ Nada ☐

2. Indique si tú maestro (a) te evalúa o califica por
Deberes ☐ Consultas ☐ Pruebas Escritas ☐
Pruebas orales ☐ Actuación en clase ☐ Cuaderno ☐

3. ¿Que te enseñaza o te da más tu Maestro(a)?
Conocimientos ☐ Valores ☐
Actividades ☐ Habilidades y Destrezas ☐

4. a) ¿Tú maestro realiza pruebas de recuperación?
Si ☐ No ☐

b) La prueba de recuperación crees tú que es: fácil
Si ☐ No ☐

5. ¿Estás de acuerdo con las calificaciones que te pone tu Maestro(a)?
SI ☐ No ☐
¿Porqué ?_____

6. ¿Tu Maestro (a) te hace participar en clase?
Siempre ☐ Ocasionalmente ☐ Nunca ☐

7. ¿Comprendes las explicaciones que te da tu Maestro(a) en el aula?
Mucho ☐ Poco ☐ Nada ☐

8. ¿Qué sientes cuando no puedes realizar tus tareas escolares?
Miedo ☐ Desesperación ☐ Indiferencia ☐ Pide ayuda ☐

9. Realizas las tareas que te envia tu Maestro(a) en casa:
Sí ☐ No ☐
Porque son:
Extensas ☐ Cortas ☐ Fáciles ☐
Difíciles ☐ Conocidas ☐ Desconocidas ☐

10. ¿Lo que tú dices o haces en clases es respetada por tu maestro(a) y compañeros?
Mucho ☐ Poco ☐ Nada ☐

11. ¿Qué materia(s) mas te gusta(n)? _____

12.- ¿Desayunos antes de venir a la Escuela?
Si ☐ No ☐
¿Porque?_____

Además las preguntas pueden ser de diversas clases, a saber:
Preguntas cerradas. En estas el informante tendrá las posibilidades al responder, como por ejemplos:

a) ¿Tu maestro (a) te hace participar en clases ?

Siempre ☐ Ocasionalmente ☐ Nunca ☐

b) ¿ Desayunas antes de venir a la Escuela?

Si ☐ No ☐

A las preguntas del tipo a) se les denomina politómicas y las del tipo b) se denominan dicotómicas.

Preguntas abiertas. Son aquellas denominadas de opinión o de contestación libre. Por la cantidad de respuestas éstas no podrán ser codificadas y su tabulación tendrá que ser manual. Por ejemplo: ¿Qué ventajas presenta el sistema actual de evaluación en la institución que trabaja?

Preguntas de control. Se hacen con el fin de controlar la veracidad de la información.

3) Análisis de datos (información)

La información obtenida debe ser depurada, clasificada, resumida y analizada. Aplicando para ello técnicas estadísticas. Los aspectos más importantes de esta etapa son:

Codificación

Cumplido el proceso de revisión de cada una de las respuestas obtenidas, se procede a la codificación de las mismas, especialmente cuando va a utilizar la tabulación mecánica. Aquellos formularios en donde la mayor parte de las preguntas son cerradas pueden ser recodificados, es decir, cada respuesta posible tiene el código impreso en el formulario.

Código es un número, letras o símbolos que sustituyen las modalidades no numéricas de una característica.

Por ejemplo si una pregunta tiene dos respuestas se utilizan los dígitos 1 y 2.

Tomando la pregunta de la encuesta 4 a).

¿Tú maestro realiza pruebas de recuperación ?

Si ..... 1 ☐
No ..... 2 ☐

En el caso de la pregunta 4 b) se tiene

La prueba de recuperación crees tú que es:

Muy buena ..... 1 ☐
Buena ..... 2 ☐
Regular ..... 3 ☐
Mala ..... 4 ☐

Ahora, si nos interesa clasificar geográficamente las unidades educativas que existen en nuestro país, se tendrá: 01 Carchi, 02 Imbabura,... ,24 Galápagos.

El proceso de revisión del cuestionario se denomina crítica, cuya finalidad es corregir las deficiencias en la recolección de la información, porque puede haber errores u omisiones, incluso cuando los formularios han sido diligenciados por encuestadores considerados como los aptos o meticulosos y que el crítico puede subsanar directamente o pidiendo al entrevistador que vuelva a la fuente de información.

Tabulación

Puede ser manual, mecánica o digital y su elección dependerá:

a) De la cantidad de formularios que se van a utilizar.
b) Del número de preguntas que tenga el formulario.
c) Del tiempo y de los recursos, ya sea financiero o de equipo, disponibles.

El procesamiento de la información se inicia una vez terminada la crítica, o después de la codificación, cuando se va hacer en forma mecánica o digital.

Análisis e interpretación

Esta etapa se puede considerar como la más importante que tiene el informe, ya que el análisis de los datos tendrá que ver con la formulación del objetivo mismo de la investigación y de las hipótesis establecidas; sin embargo, este proceso de análisis (aplicación de las técnicas de la estadística descriptiva como también de la inferencial) tendrá menos dificultad, si el investigador tiene pleno conocimiento de los problemas que son inherentes al planeamiento de una investigación.

En este proceso se debe considerar la elaboración de distribuciones o tablas de frecuencias, obtenidas a través de una sistematización de la información para poder ser presentada en forma de cuadros.

Con los anteriores resultados se procede luego a hacer un resumen y a la aplicación de las diferentes medidas estadísticas: de tendencia central, de dispersión o de asociación, incluyendo en éstos los porcentajes o proporciones.

Con las cifras resultantes, se pueden hacer comparaciones con otros estudios, para poder llegar a mejores conclusiones. De esta última fase de la metodología se puede decir que encierra dos aspectos.

a) Análisis y evaluación técnica de los resultados.
b) Análisis y evaluación técnica de acuerdo con la naturaleza de la investigación.

Estos dos aspectos permitirán determinar el grado de consistencia y confiabilidad de los resultados obtenidos de la investigación.

El profesor John W. Best en su libro ¿Cómo investigar en educación?, nos da una posible guía del análisis, sugiriendo los siguientes puntos:

1. Título:

a) ¿Es claro y conciso?
b) ¿No promete más de lo que el estudio puede proporcionar?

2. El problema:

a) ¿Se halla establecido con claridad?
b) ¿Está bien delimitado?
c) ¿Se reconoce su significado?
d) ¿Las preguntas son específicas y se encuentran establecidas las hipótesis con claridad?
e) ¿Se establecen supuestos y limitaciones?
f) ¿Se definen los términos importantes?

3. Revisión de la bibliografía relacionada:

a) ¿Es de amplitud adecuada?
b) ¿Se destacan los hallazgos importantes?
c) ¿Está bien organizada?
d) ¿Se procura un resumen efectivo?

4. Procedimientos utilizados.

a) ¿Se describe detalladamente el diseño experimental?
b) ¿Es adecuado este diseño?
c) ¿Se describen las muestras?
d) ¿Se reconocen las variables relevantes?
e) ¿Se procuran controles adecuados?
f) ¿Son idóneos los instrumentos de recogida de datos?
g) ¿Se establecen la validez y la fiabilidad?
h) ¿Es adecuado el tratamiento estadístico?

5. Análisis e interpretación de datos

a) ¿Es adecuado el uso de tablas y figuras?
b) ¿Es concisa y clara la exposición del texto?
c) ¿Es lógico y perceptible el análisis de las relaciones de datos?
d) ¿Se interpreta con precisión el análisis estadístico?

6. Resumen y conclusiones:

a) ¿Se replantea el problema?
b) ¿Se describen con detalle los procedimientos?
c) ¿Se presentan concisamente los hallazgos?
d) ¿Es objetivo el análisis?
e) ¿Los datos presentados y analizados justifican los hallazgos y conclusiones?

Publicación.

Corresponde a la fase final de la investigación y con ella se propone hacer llegar a las personas interesadas el resultado total del estudio, teniendo en cuenta todos los aspectos considerados en el proceso, en tal forma que los datos sean comprensibles, con la correspondiente validez que merezcan las conclusiones.

En términos generales se puede decir que un **informe** deberá contener:

a) A investigar el planteamiento del problema a investigar.
b) Objetivo de la investigación.
c) Hipótesis que se quiere probar.
d) Breve exposición de la metodología adoptada, diseño y tamaño de la muestra. Proceso de selección de las unidades de información y de recolección.
e) Se podrá incluir en el informe una copia del cuestionario utilizado en la recopilación.
f) Descripción de los resultados en forma de cuadros y gráficos, acompañados del análisis y comparaciones obtenidas a través de los datos.
g) *Conclusiones y recomendaciones.*
h) Muchas veces el informe tiene una parte final, denominado apéndice, en donde se incluyen cuadros más generales que permiten aclarar o comprobar rápidamente cualquier información más detallada.

1.9 ESCALAS O NIVELES DE MEDIDA

INFORMACIÓN CUALITATIVA

a) Escala Nominal.- Es la escala más débil en cuanto a la información que proporciona. Como su nombre lo indica, esta escala consiste en "nombrar a las observaciones". Para distinguir los agrupamientos de unidades se emplean símbolos, letras o números. En el caso de que se empleen números, estos solo tienen un carácter simbólico y no numérico.

Ejemplos:

Estado civil de los habitantes de Riobamba (soltero=1, casado=2, divorciado=3, unión libre=4, viudo = 5).
Tipos de uso del suelo (agrícola = 1, forestal = 2, pecuario = 3, etc.) en el municipio de Guamote

b) Escala Ordinal.- En este nivel, las unidades de los grupos guardan cierta relación entre sí, que se pone de manifiesto cuando se está en posibilidad de establecer una relación de tipo mayor o menor que.

Ejemplos:

Nivel de estudios, ya que sus modalidades están ordenadas según la duración de los estudios: Educación primaria, secundaria, diversificado, universitaria; antes, ahora EGB, Bachillerato, Universitaria, posgrado: maestría, PhD

Grado de aceptación de algún producto: buena, regular, mala; Alto en …,Medio en …, Bajo en …

Nivel socioeconómico de una familia (alto, medio, bajo).

INFORMACIÓN CUANTITATIVA

a) Escala de Intervalo.- Este tipo de escala provee información mucho más precisa, a la vez que permite llevar a cabo mediciones mucho más sofisticadas que las escalas nominal u ordinal. La escala de intervalo no sólo informa acerca del orden de unos objetos, sino también acerca de las distancias o diferencias numéricas entre dichos objetos. De hecho, esta escala permite medir y comparar esas distancias o diferencias con precisión. En otras palabras (y de aquí el nombre de escalas de intervalo), las distancias o 'intervalos' de igual tamaño en la escala son de hecho iguales no importando donde en la escala se realice la medición.
Por ejemplo, los resultados numéricos de los exámenes académicos (rango de 0 a 100) pueden ser medidos usando escalas de intervalo.

La escala de intervalo, sin embargo, no posee una definición única del valor cero. En otras palabras, el cero es arbitrario en el sentido de que no representa ausencia absoluta de la característica que se desea medir. En este sentido las escalas de intervalo son equivalentes a termómetros, en los que el valor cero no representa la ausencia absoluta de calor.

En el ejemplo anterior, si un estudiante obtiene un resultado de cero puntos en un examen, ello obviamente no significa que el estudiante no sepa absolutamente nada acerca de la materia evaluada.

El comportamiento humano es casi siempre medido utilizando escalas de intervalo. Otras variables medidas en esta escala son: temperatura, horario meridiano, grados de latitud o de longitud, coeficiente de inteligencia.

La numeración de los años en nuestro calendario utiliza también una escala de intervalos. Las autoridades eclesiásticas y gubernamentales de la época decidieron arbitrariamente fijar como el año 1 el del nacimiento de Cristo y como unidad de medida un lapso de 365 días.

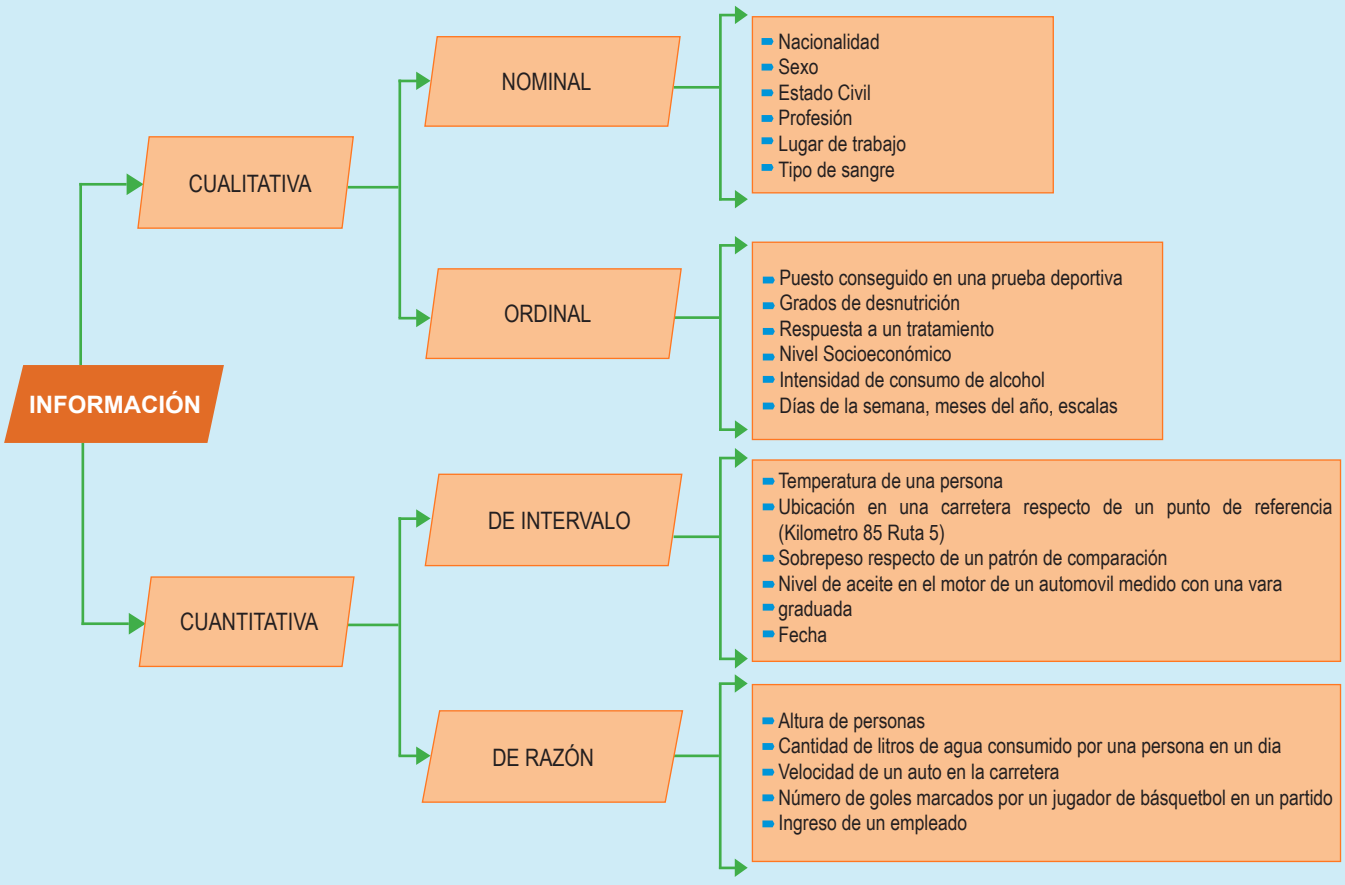
b) Escala de Razón.- Los atributos son cuantitativos organizados en una escala donde tanto el intervalo entre dos valores, como el punto cero, tienen significado real (indica ausencia de valor). Dadas dos medidas en esta escala, podemos decir si son iguales, o si una es diferente, mayor, que tan mayor y cuantas veces la otra. La altura de un individuo es un ejemplo de la medida en esta escala. Si ella fuera medida en centímetros (cm), 0 cm es el origen y 1 cm es la unidad de medida.

Un individuo con 180 cm es dos veces más alto que un individuo con 90 cm, y esta relación continua valiendo si usamos 1 cm como unidad. Otras variables que son medidas en esta escala son: peso, longitud, diámetro, volumen, estatura, densidad, etc.

Nota. *Propusimos a los estudiantes del cuarto nivel de Ingeniería en Estadística Informática de la ESPOCH como clasificaría las variables (datos o información) respecto a la Escala de Medida y obtuvimos el siguiente resultado:*
(Ver figura 6)

CLASIFICACIÓN DE LAS VARIABLES CON RESPECTO A ESCALAS DE MEDIDA

Figura 6. *Clasificación de variables respecto a las escalas de medida*



1.10 ACTIVIDADES DE APRENDIZAJE 1

1. ¿Qué son los censos demográficos?

.....

2. ¿A quién se le atribuye la introducción de la palabra Estadística y cuál fue su significado?

.....

3. ¿Quién fue John Graunt y qué realizó?

.....

4. ¿Quiénes son los iniciadores de la Teoría de las Probabilidades?

.....

5. ¿Qué significado tiene cada uno de los siguientes términos: Población, muestra, parámetros, estimador? E indique un ejemplo de cada uno de los términos.

.....

6. ¿Qué funciones cumple la estadística?

.....

7. ¿Qué trata la estadística?

.....

8. Indique ¿cuándo un caracter es un v.a. y realice un esquema de la clasificación de las v.a.?

.....

9. De tres ejemplos de población determinando: los individuos y las v.a. (cualitativas. cuantitativas discretas y cuantitativas continuas).

.....

10. Indique la definición de muestreo aleatorio simple. Y seleccione una muestra aleatoria de 7 estudiantes de un curso de 25 estudiantes mediante la tabla de números aleatorios de columnas vigésima primera y vigésima segunda y fila tercera.

.....

11. Se debe tomar una muestra de tamaño n=100 de una población consistente en dos estratos para los cuales $N_1 = 10000$, $N_2 = 30000$, $\sigma_1 = 45$, $\sigma_2 = 60$ ¿Qué tan grande tiene que ser una muestra que se debe tomar de cada uno de los dos estratos para lograr una distribución óptima?

12. ¿Por qué está Ud. aprendiendo Estadística?

.....

13. Señalar el literal más adecuado para los siguientes aspectos:

13.1. Antes que nada, la investigación estadística requiere

a) Que exista un objetivo.
b) Que se hayan trazado planes.
c) Que se tenga un problema.
d) Ninguno de los anteriores.

13.2. *En el diseño del cuestionario las preguntas más difíciles deben colocarse:*

a) Al principio, para salir inmediatamente de la parte más difícil.
b) En el centro para que sean precedidas y seguidas por preguntas fáciles.
c) Al final, luego que se haya establecido un clima de confianza, al comenzar por las fáciles hasta llegar a las difíciles.
d) Ninguno de los anteriores

14. *¿Qué tipos de Investigación existen? Y defina dos tipos cualesquiera.*

……

15. *¿Qué le gustaría investigar? Explique a qué tipo de investigación pertenece lo que desea investigar.*

……

16. *Indique 2 problemas que se presenten en el proceso educativo de la institución o facultad que Ud. trabaja.*

……

17. *ESCALAS o NIVELES DE MEDIDA Y VARIABLES*

17.1. *Mencione los niveles de medida.*

……

17.2. *Mencione la diferencia entre el nivel nominal y el nivel ordinal.*

……

17.3. *Diferencia entre el nivel de intervalo y el nivel de razón.*

……

17.4. *Jerarquice las escalas de medida de acuerdo al orden decreciente de perfección.*

……

17.5. *Indique las escalas de medida que corresponda en cada uno de los casos siguientes:*

a) En una unidad educativa se registra la estatura de los alumnos, de un grupo. _____
b) En una unidad educativa aparece la lista de los cinco mejores alumnos en orden decreciente respecto a un aprovechamiento. _____
*c) En una unidad educativa se mide el coeficiente de inteligencia de los alumnos*_____
d) El censo de 2010 señala la ocupación de los habitantes censados. _____
e) En una ciudad se registra la temperatura durante un mes. _____

17.6. *Si un alumno tiene un coeficiente de inteligencia de 75 y otro de 150, ¿Es correcto afirmar que el segundo es doblemente inteligente que el primero? ¿por qué?*

……

17.7. *Si dos establecimientos comerciales registran ventas de 1,000.00 y 2,000.00 dólares respectivamente. ¿Las ventas del segundo establecimiento son el doble que los del primer establecimiento? ¿Por qué?*

……

17.8. *Diga si los siguientes ejemplos corresponden a variables continuas o discretas*

a) El número de clientes atendidos diariamente durante en un mes en una institución bancaria
_____
b) El tiempo de traslado de un grupo de alumnos de su casa a la facultad en un día cualquiera de clases _____
c) El número de productos defectuosos por lote durante la inspección de 25 lotes.
d) Diámetro de los árboles de un huerto _____
e) La velocidad a la que circulan los automóviles que transitan por cierta avenida de la ciudad de Riobamba _____

18. *Mencione tres ejemplos de variables continuas y tres ejemplos de variables discretas.*

……

19. *Realice un análisis crítico de la realidad social y educativa a nivel local, provincial o nacional, para determinar áreas de investigación educativa.*

……

2 CAPÍTULO ESTADÍSTICA DESCRIPTIVA

OBJETIVOS

- Describir gráfica y numéricamente un conjunto de datos de manera que resuman y puedan ayudar más adelante a tomar decisiones.
- Desarrollar destrezas y habilidades en los estudiantes en la elaboración de tablas y gráficos estadísticos.
- Aplicar los métodos descriptivos en el manejo de la información que provenga de la práctica docente como de otros campos

CONTENIDOS

2.1 DESCRIPCIÓN GRÁFICA DE DATOS

La Estadística Descriptiva o deductiva tiene como finalidad colocar en evidencia aspectos característicos (promedios, desviaciones estándar, coeficiente de variación o CV, variabilidad de calificaciones por ejemplo, etc.), que sirven para efectuar comparaciones sin pretender sacar conclusiones de tipo más general. Esta descripción se realiza a través de la elaboración de cuadros, gráficos, cálculo de estadísticas descriptivas como promedios, varianzas, proporciones y mediante el análisis de regresión.

La Estadística Descriptiva es una primera aproximación de las nociones de la Teoría de las Probabilidades tratados con un número finito de observaciones. La Estadística Descriptiva presenta diversos métodos para estudiar y sintetizar un conjunto de datos obtenidos mediante experimentos. En general se indica con S (del inglés Spazio) la población de la investigación estadística. A cada individuo se asocia los valores de las variables que se consideran en la investigación, en general se prefiere representar el conjunto de estos valores con un vector X de dimensión n, si n es el número de las variables consideradas en el modelo, a veces se prefiere representar los datos mediante una matriz, llamada matriz de datos cuyas columnas son los vectores X.

Si las variables están representadas en términos numéricos los vectores X pueden ser vistos como puntos del espacio R^n mientras, si la cardinalidad (número de individuos o elementos de un conjunto) de S es m; las filas de la matriz de datos pueden ser considerados como puntos de R^m, a saber:

Vector de n Variables:

$$X=(X_1,X_2,...,X_n) \in R^n$$

Valores de las variables j-ésima para m individuos

$$\begin{pmatrix} x_{1j} \\ \vdots \\ x_{mj} \end{pmatrix} \in R^m$$

Matriz de datos

Una matriz de datos es un conjunto de n variables sean nominales, ordinales, de intervalo o de razón, las mismas que almacenan información de m individuos, a estos también se los conoce como unidades (de registros).

	Variables						
Individuos	X_1	X_2	. . .	X_j	.	.	. X_n
1	x_{11}	x_{12}		x_{1j}	.	.	. x_{1n}
2	x_{21}	x_{22}		x_{2j}	.	.	. x_{2n}
.	.	.		.			.
.	.	.		.			.
.	.	.		.			.
m	x_{m1}	x_{m2}		x_{mj}	.	.	. x_{mn}

En la matriz de datos x_{2j} representa el valor del segundo individuo de la j-ésima variable.

ACTIVIDAD DE APRENDIZAJE DESARROLLADA (MATRIZ DE DATOS) **01**

El profesor dirigente del tercer curso "A" del COMIL-R, un curso de 25 estudiantes. Al finalizar el primer quimestre le interesa conocer el rendimiento académico del curso en las asignaturas más importantes: Matemáticas, Castellano, Ciencias Sociales, Ciencias Naturales e Inglés. Y conocer el numérico y porcentaje respecto al sexo de los 25 estudiantes.

Solución:

Los individuos son los 25 estudiantes del tercer curso paralelo A y las variables son:

X_1: "Nomina de los 25 estudiantes del tercer curso A"
X_2: "Calificaciones de la materia de Matemáticas de los 25 estudiantes"
X_3: "Calificaciones de la materia de Castellano de los 25 estudiantes"
X_4: "Calificaciones de la materia de Ciencias Sociales de los 25 estudiantes"
X_5: "Calificaciones de la materia de Ciencias Naturales de los 25 estudiantes"
X_6: "Calificaciones de la materia de Inglés de los 25 estudiantes"
X_7: "Género de los 25 estudiantes"

En efecto, las variables X_2, X_3, X_4, X_5, y X_6 son cuantitativas y las variables cualitativas son X_1 y X_7 la última con subíndice 7 es una v.a. cualitativa codificada con 1: masculino y con 2: femenino.

El profesor dirigente se cuestiona: ¿Qué porcentaje de hombres y mujeres tiene este curso? ¿Cuál es la materia o asignatura de mayor rendimiento? ¿Qué podemos decir respecto al rendimiento por género? ¿Qué asignatura tiene mayor variabilidad? ¿Cuál es el ranking de las asignaturas? y ¿cómo interpreta estos resultados?.

La matriz de datos se puede representar por:

No.	Nómina	Matemáticas	Castellano	Ciencias Sociales	Inglés	Ciencias Naturales	Sexo
1	CARVAJAL ANA	15.36	17.41	16.12	14.00	10.66	2
2	CORTÉS CESAR	20.00	14.35	14.00	12.17	14.19	1
3	AVILA EDUARDO	14.19	17.05	18.00	9.98	17.23	1
4	SILVA EDISON	19.89	14.09	14.67	18.00	17.35	1
5	ALINO FABIAN	18.97	19.45	15.00	19.03	18.69	1
6	SINIDA MARIA	12.56	15.00	16.07	19.78	13.00	2
7	PORTERO NELLY	7.45	14.17	18.00	18.05	11.76	2
8	BERRONES PAUL	20.00	18.06	15.09	16.79	19.11	1
9	DONOSO ANGEL	12.45	20.00	18.79	15.00	20.00	1
10	CONCHA IVAN	16.93	12.62	16.24	17.08	19.78	1
11	ROSERO ENIT	14.00	14.67	17.59	18.98	14.09	2
12	MIRANDA ROSA	20.00	12.99	13.74	10.01	15.54	2
13	MOYA GERMAN	12.90	14.78	17.00	14.59	18.99	1
14	CALI CARLOS	11.97	12.56	14.34	11.97	14.00	1
15	VILLA EDUARDO	19.70	20.00	17.76	18.58	14.45	1
16	AUSAY CARMEN	18.00	20.00	17.16	19.56	16.41	2
17	RIVAS JORGE	14.00	17.78	18.37	12.67	15.69	1
18	ORTEGA MARIA	11.00	20.00	15.38	14.00	18.00	2
19	CANO ROBERTO	19.87	16.00	18.09	18.89	18.89	1
20	PUSAY DIEGO	20.00	19.42	18.00	19.58	18.68	1
21	PEÑA DAVID	18.00	17.47	14.39	14.74	14.74	1
22	MORA PAULINA	19.90	18.56	16.79	18.23	19.45	2
23	VILEMA RITA	14.89	17.00	16.73	14.73	14.81	2
24	PAGUAY MARIA	10.79	18.45	16.39	12.53	13.27	1
25	AUSAY JAVIER	13.99	19.42	18.00	14.44	17.67	1

Por simple conteo de los 25 estudiantes existen 9 mujeres (2) y 16 hombres (1) que son las frecuencias absolutas. Y en frecuencias relativas tanto para mujeres como hombres respectivamente serán:

Género	Codificación (**i**)	Frecuencia Absoluta ($\mathbf{m_i}$)	Frecuencia Relativa ($\mathbf{f_i}$)
Femenino	2	9	36%
Masculino	1	16	64%
Total		25	100%

Diagrama circular: Porcentaje según género

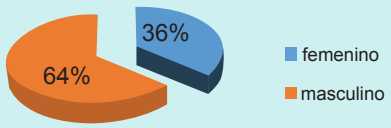
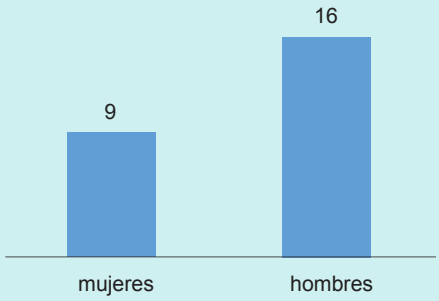


Diagrama de barras: Conteo de hombres y mujeres



Los resultados de la siguiente tabla se obtiene de Excel y se explicará más adelante, consideramos aquí para responder las preguntas planteadas y la necesidad de conocer argumentos probabilísticos (Capitulo III) para describir las estadísticas pedidas en las preguntas.

Estadísticas	Matemáticas	Castellano	Ciencias Sociales	Inglés	Ciencias Naturales
Promedio	15,87	16,85	16,47	15,74	16,26
Desviación estándar	3,72	2,53	1,53	3,09	2,68
CV en %	23,46%	15,02%	9,28%	19,62%	16,46%

Diagrama radial: Comparación de promedios de las 5 asignaturas del curso

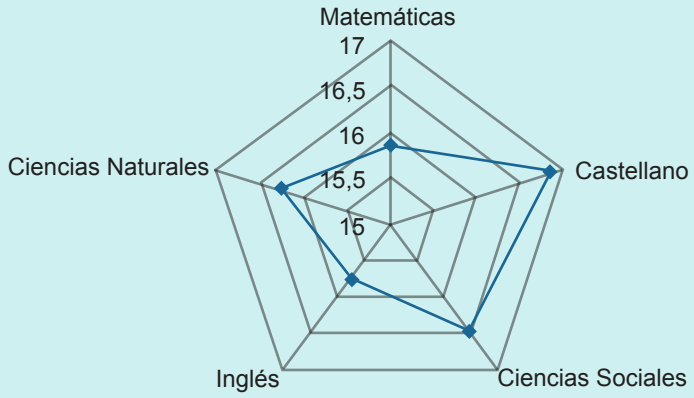
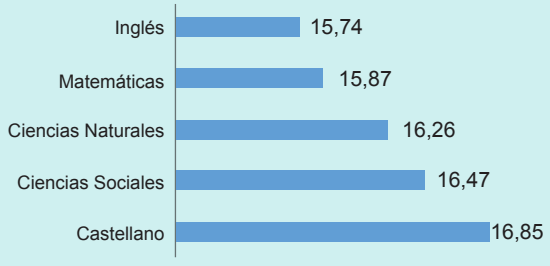


Diagrama de Barras: Ranking del promedio de las asignaturas del curso



Ranking	Promedio
Castellano	16,85
Ciencias Sociales	16,47
Ciencias Naturales	16,26
Matemáticas	15,87
Inglés	15,74

De las 5 asignaturas la de mayor puntaje es Castellano y la que presenta menor variabilidad de ellas es Ciencias Sociales indicando que los estudiantes en esta asignatura es más homogéneo ¿por què?.

2.1.1 VARIABLES CUALITATIVAS

Las variables cualitativas están caracterizadas por observaciones no numéricas, no obstante pueden ser codificadas mediante símbolos numéricos, observando pero que el orden de codificación es del todo arbitrario. Indiquemos con E = {1, 2,..., k} una posible codificación de las observaciones realizadas y sea m_i el número de individuos que tiene resultado i=1,...,k , al mismo que se denomina ***frecuencia absoluta.***
Se indica con $f_i = \frac{m_i}{m}$ ***frecuencia relativa*** del resultado i-ésimo, m indica el número total de individuos. Estas frecuencias satisfacen las siguientes propiedades:

Propiedades de frecuencias:

$$1.\ 0 \le m_i \le m \ \forall_i = 1, ..., k \quad \Rightarrow \quad 0 \le f_i \le 1$$

$$2.\ \sum_1^k m_i = m_1 + m_2 + \cdots + m_k = m \quad \Rightarrow \quad \sum_1^k f_i = f_1 + f_2 + \cdots + f_k = 1$$

Los valores f_i representan la *distribución empírica o distribución* de los datos observados.

ACTIVIDAD DE APRENDIZAJE DESARROLLADA (FRECUENCIAS) **02**

El profesor dirigente referido en la actividad de aprendizaje 1 desarrollada, está interesado en determinar en las asignaturas de Matemáticas y Ciencias Sociales, el número y el porcentaje (f_i x100%) de calificaciones.

Nota. En el colegio Militar de Riobamba se considera la escala de calificaciones siguiente:

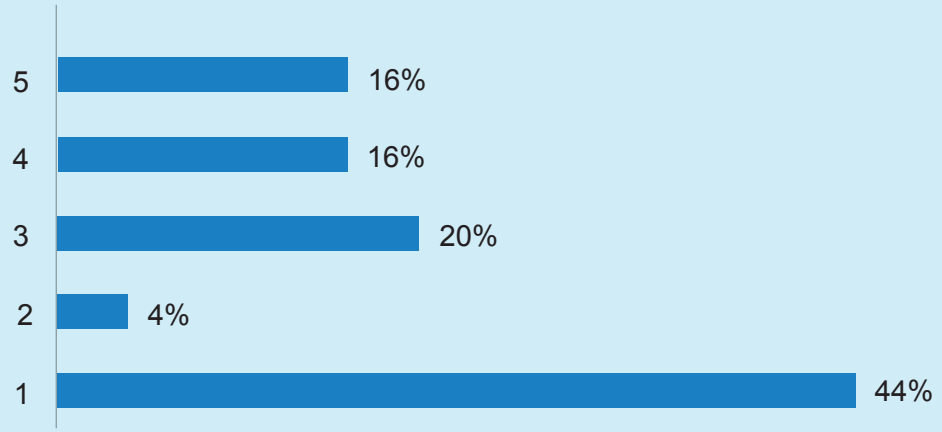
Sobresaliente (S)	18.00 – 20.00
Muy Buena (MB)	16.00 – 17.99
Buena (B)	14.00 – 15.99
Regular (R)	12.00 – 13.99
Insuficiente (I)	menos de 11.99

Solución:

Asignatura: **MATEMÁTICAS**

Codificación	Escala	Frecuencia absoluta	Frecuencia relativa	Porcentaje
i	(1)	(2)	(3)= (2)/25	(4)=(3)*100%
1	S	11	0.44	44%
2	MB	1	0.04	4%
3	B	5	0.20	20%
4	R	4	0.16	16%
5	I	4	0.16	16%
Total		25	1.00	100%

Diagrama de barras: Porcentaje de estudiantes según escala de calificaciones

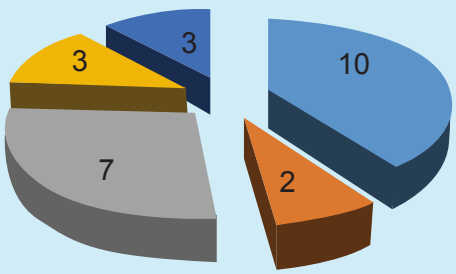


Se observe que las frecuencias absolutas son: m_1 =11, indica el número de sobresalientes; m_2 = 1, indica el número de muy buenas; m_3 = 5, indica el número de buenas; m_4 = 4, indica el número de regulares y m_5 = 4, indica el número de insuficientes. Además se verifica la propiedad 2, es decir, $\sum m_i$= 11+1+5+4+4 = 25, luego m=25. La frecuencia relativa f_1=11/25 =0.44 o en porcentaje 44%, conjuntamente con las otras frecuencias relativas se cumplen las propiedades establecidas al observar la fila de la tabla de frecuencias del total.

Análogamente se procede con la asignatura de **CIENCIAS SOCIALES.**

Codificación i	Escala (1)	Frecuencia absoluta (2)	Frecuencia relativa (3)	Porcentaje (4)=(3)*100%
1	S	10	0.40	40%
2	MB	2	0.08	8%
3	B	7	0.28	28%
4	R	3	0.12	12%
5	I	3	0.12	12%
Total		25	1.00	100%

Conteo de estudiantes según escalas de medida



Opcional. Si se efectúan dos experimentos y sus resultados están recogidos en las variables cualitativas X y Y sean E_1 y E_2 las respectivas codificaciones, es útil representar los datos observados mediante la siguiente tabla, llamada tabla de *contingencia ó de doble entrada.*

X/Y	1	. . .	j	. . .	n	Marginal X
1	f_{11}	. . .	f_{1j}	. . .	f_{1n}	$f_{1.}$
.						
.						
i	f_{i1}	. . .	f_{ij}	. . .	f_{in}	$f_{i.}$
.						
.						
.						
k	f_{k1}	. . .	f_{kj}	. . .	f_{kn}	$f_{k.}$
Marginal Y	$f_{.1}$	. . .	$f_{.j}$	. . .	$f_{.n}$	1

El valor f_{ij} es la frecuencia de los valores i y j de la muestra y es dado por $\frac{m_{ij}}{m}$ con m_{ij} el número de individuos con valor i de la primera observación y valor j de la segunda. Los valores f_{ij} (i=1,...,k y j=1,...,n) constituyen la ***distribución empírica conjunta o distribución conjunta*** de la pareja (X,Y) de los dos experimentos, los valores f_i y f_j son respectivamente las distribuciones de los experimentos X y Y y son llamadas distribuciones empíricas de X y Y respectivamente las mismas que satisfacen:

$$f_{i.} = \sum_{l=1}^{n} f_{il} \quad , \quad \sum_i f_{i.} = 1 \quad , \quad f_{.j} = \sum_{r=1}^{k} f_{rj} \qquad \text{y} \quad \sum_j f_{.j} = 1$$

Son de particular interés las frecuencias relativas de la pareja (i, j), respecto a uno de los resultados, por ejemplo los valores $\frac{f_{ij}}{f_{.j}}$ para i=1,…, k representa la distribución condicionada de la segunda variable respecto al hecho que la primera variable asume el valor i.

Un caso particular está representado por la igualdad válida por cada pareja (i, j), $\frac{f_{ij}}{f_{.j}} = f_{i.}$, que caracteriza el hecho que entre las dos variables X y Y no tienen ninguna relación estadística, es decir, la distribución de una de ellas no está influenciada por los posibles valores de la otra y se dicen ***variables independientes.*** Al respecto, consideremos por ejemplo que el color de los ojos de los estudiantes (X) y la escala de calificaciones (Y) de los mismos no guardan ninguna relación, es decir, son variables independientes.

Definición. Se dice que las variables X y Y son estadísticamente independientes si por cada pareja (i, j) vale la igualdad

$$f_{ij} = f_{i.} * f_{.j}$$

Análogas consideraciones se dan a las variables cuantitativas.

2.1.2 VARIABLES CUANTITATIVAS

Si los resultados de una investigación nos llevan a obtener valores numéricos se tienen entonces las variables llamadas cuantitativas, cuyos valores numéricos tienen su importancia, los datos de una variable X se pueden representar con histogramas, diagrama de caja, diagrama de puntos, entre otros que veremos más adelante. En muchos casos, por ejemplo si las variables asumen valores en un conjunto continuo (en un intervalo o en la recta de los reales), o sea variables continuas, es útil reagrupar los datos en clases sin tener en cuenta la unidad originaria.

Para no perder completamente la información sobre los valores asumidos al interior de las clases, se puede obtener un valor del orden de grandeza de la clase, se tiene así la representación del diagrama de tallo y hoja (*stem and leaf*), la misma es una buena manera de obtener una presentación visual informativa del conjunto de datos $x_1, x_2, ..., x_m$ donde cada número x_i está formado al menos por dos dígitos.

Para construir un diagrama de tallo y hoja, los números x_i se dividen en dos partes: un tallo, formado por uno o más de los dígitos principales y una hoja, la cual contiene el resto de los dígitos. Lo usual es seleccionar entre 5 y 20 tallos, los mismos que son enlistados en la parte izquierda del diagrama. Al lado de cada tallo se ponen todas las hojas que corresponden a los valores observados ordenados tal como se encuentran en el conjunto de datos. Para aclarar lo dicho consideramos la siguiente actividad de aprendizaje:

ACTIVIDAD DE APRENDIZAJE DESARROLLADA 03

Los siguientes datos son los lapsos, en minutos, necesarios para que 50 clientes de un banco comercial lleven a cabo una transacción bancaria:

2.3	0.2	2.9	0.4	2.8	3.1	3.7	7.2	1.6	1.9
2.4	4.4	5.8	2.8	3.3	2.4	4.6	3.8	1.5	2.7
3.3	9.7	2.5	5.6	9.5	0.4	1.3	1.1	5.5	3.4
1.8	4.7	0.7	6.2	1.2	4.2	1.2	0.5	6.8	5.2
7.8	0.8	0.9	0.4	1.3	6.3	7.6	1.4	0.5	1.4

Construya un diagrama de Tallo y hoja (*stem and leaf*).

Solución:

Se trata de una v.a. continua, X:"Tiempo en minutos de transacción de clientes de un banco comercial" puesto que el tiempo de duración se define en general sobre la recta real no negativa, esto es, los valores x≥0.

Su representación mediante un diagrama de tallo y hoja ordenada es:

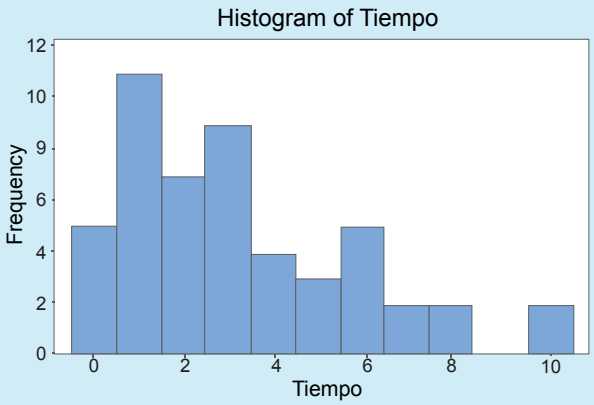
	Hoja	Frecuencia
0	244455789	9
1	12233445689	11
2	34457889	8
3	133478	6
4	2467	4
5	2568	4
6	238	3
7	268	3
8		0
9	57	2

Aplicando el software estadístico Minitab visualizamos el diagrama de tallo y hoja que se presenta un orden de menor a mayor en cada una de las hojas del tallo. Además su histograma está dado por:

Stem-and-Leaf Display: Tiempo

Stem-and-leaf of Tiempo N = 50
Leaf Unit = 0,10

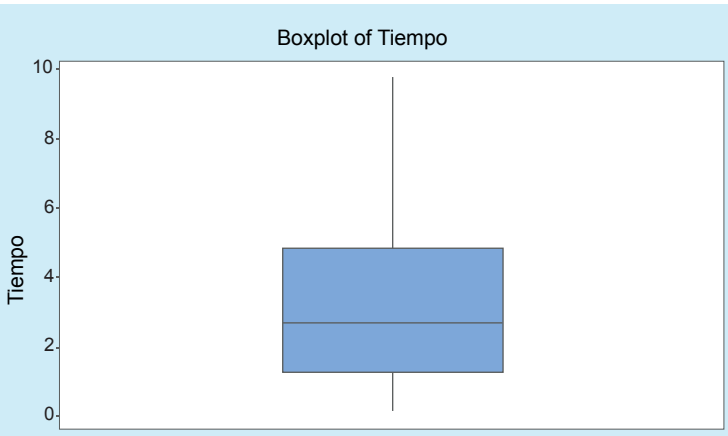
9	0	224445789
20	1	12233445689
(8)	2	34457889
22	3	133478
16	4	2467
12	5	2568
8	6	238
5	7	268
2	8	
2	9	57



Se puede representar también los datos en tablas por sus límites reales y sus marcas de clase como la siguiente:

CLASES	1	2	3	4	5	6	7	8	9	10
Clases por sus limites reales	0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10
Clases por sus marcas	0.5	1.5	2.5	3.5	4.5	5.5	6.5	7.5	8.5	9.5
Rango	1	1	1	1	1	1	1	1	1	1

Observación. El histograma como el diagrama de tallo y hojas son gráficos que describen los mismos resultados en cuanto a tendencia, variabilidad o dispersión (como indica la siguiente gráfica Diagrama de caja).



En la actividad de aprendizaje desarrollada 3, el número de clases que se puede elegir es k = 10, es decir, se toma en cuenta el número de tallos descriptos en el diagrama stem and leaf (Tallo y hoja).

Para determinar el número de clases, k, se puede utilizar también la ***fórmula de Sturgues*** (fórmula empírica) dada por:

$$k = 1 + 3.322ln(n)$$

donde n es el número de observaciones o datos, generalmente la fórmula se aplica cuando n > 50. Se establece las clases de forma que tengan la misma diferencia entre ellas, el mismo rango y el número de clases aumenta en función de n.

Nota. *En la práctica se obtienen buenos resultados si se hace la selección del número de clases considerando la raíz cuadrada de n.*

$$k = \sqrt{n}$$

Cada clase está definida por dos valores, estos valores constituyen los ***límites reales*** de las clases. El límite real superior de una clase es el límite inferior de la siguiente. La diferencia entre los límites reales de una clase constituye el ***intervalo de la clase.*** Se llama ***marca de clase o punto medio*** al valor correspondiente de su intervalo.

Cuando tenemos ya determinado las clases clasificamos y contamos los individuos incluidos en cada clase. El número resultante se denomina ***frecuencia absoluta*** de la clase respectiva.

El número de individuos de una clase se puede expresar también mediante su frecuencia relativa, bien en forma de proporción (cociente entre la frecuencia absoluta de esa clase y el número total de individuos de la muestra) o bien en forma de porcentaje (frecuencia referida a 100 individuos de la muestra).

De la Actividad de Aprendizaje desarrollada 3, se tiene la siguiente tabla.

Marca de clase	Asignación	Recuento
0.5	|||||||||	9
1.5	|||||||||||	11
2.5	|||||||||	8
3.5	||||||	6
4.5	||||	4
5.5	||||	4
6.5	|||	3
7.5	|||	3
8.5		0
9.5	||	2

Asignación de los individuos de una muestra utilizando las marcas de clase

Con los datos de este problema realicemos:

a) Tabla de frecuencias
b) Histograma
c) Polígono de frecuencias
d) Ojiva

Tabla de frecuencias

Existen tres reglas generales para construir la tabla de frecuencias:

1. Determinar el mayor y el menor entre los datos registrados para luego calcular el rango (diferencia entre el mayor y el menor valor de los valores).

   rango = 9.7 - 0.2 = 9.5

2. Dividir el rango en un número conveniente de intervalos de clase del mismo tamaño (igual longitud). Si esto no es posible, entonces utilizar intervalos de clase de diferente tamaño. El número de clases que se emplea para clasificar los datos depende del total de observaciones.

   Como anteriormente se dijo, si el número de observaciones es relativamente pequeño, la experiencia muestra que el número de clases a emplear es generalmente mayor o igual a 5. Si existe una cantidad grande de datos, el número de clases debe encontrarse entre 8 y 12 y generalmente no existirán más de 15 o 20 clases.

   Para esta actividad de aprendizaje se decidió utilizar 10 clases (intente aplicar la fórmula de Sturgues o utilice la raíz cuadrada de n). Entonces la longitud de cada clase será aproximadamente 1, pues
   9.5/10 = 0.95 ≅ 1.

3. Determinar el número de observaciones que caen en cada clase (recuento de datos). La tabla de frecuencias que se obtiene es:

No. i.	Clase	Límite Inferior	Límite Superior	Marca c_i	Frecuencia de clase n_i	Frecuencia Relativa f_i	Porcentaje %
1	(0 - 1]	0	1	0.5	9	0.18	18%
2	(1 - 2]	1	2	1.5	11	0.22	22%
3	(2 - 3]	2	3	2.5	8	0.16	16%
4	(3 - 4]	3	4	3.5	6	0.12	12%
5	(4 - 5]	4	5	4.5	4	0.08	8%
6	(5 - 6]	5	6	5.5	4	0.08	8%
7	(6 - 7]	6	7	6.5	3	0.06	6%
8	(7 - 8]	7	8	7.5	3	0.06	6%
9	(8 - 9]	8	9	8.5	0	0.00	0%
10	(9-10]	9	10	9.5	2	0.04	4%
	Total				n=50	1.00	100%

La columna de las frecuencias relativas se denomina nuevamente distribución de frecuencias relativas o simplemente distribución.

Las tablas estadísticas presentadas son la expresión escrita de la distribución de frecuencias de los individuos de una muestra respecto a una variable.

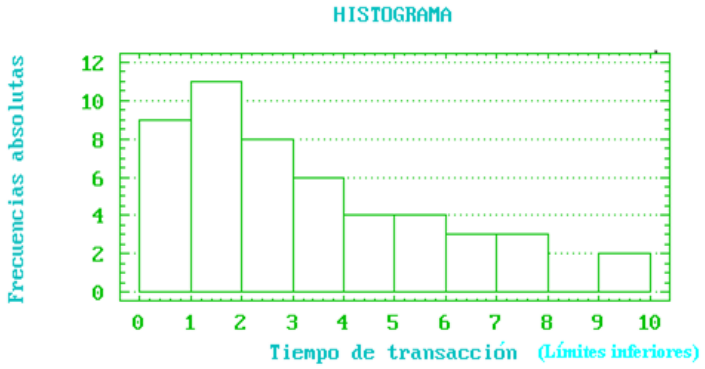
Porcentaje

Las frecuencias relativas f_i expresan la asignación de cada observación a una determinada clase y por la primera propiedad de las frecuencias sus valores siempre están entre 0 y 1, es decir, son números decimales o fracciones que en lo cotidiano no son utilizadas por ejemplo 0.05, 0.5, 0.14, etc.; pero si se escucha a menudo cinco por ciento, cincuenta por ciento o catorce por ciento, etc., los que se indican por 5%, 50% o 14% etc., que son números decimales multiplicados por 100% a los que se les denomina porcentajes.

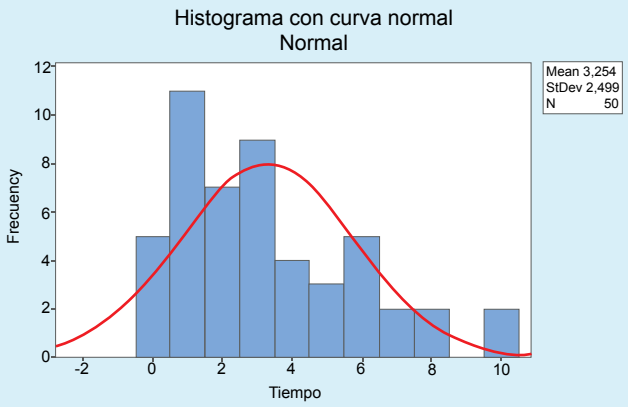
Observación. En el manejo estadístico de datos es importante aprender a utilizar de modo equivalente porcentajes, fracciones y números decimales. Para el hombre de la calle no es sencillo comprender un resultado fraccionado o decimal y la experiencia nos ha enseñado utilizar porcentajes a inusuales fracciones o decimales. Pasando de la fracción al porcentaje, el valor de la fracción se puede redondear o truncar y es claro que si se suman los porcentajes de todos los resultados de las frecuencias relativas no tendremos por lo general el 100% por lo que se debe tener cuidado con el redondeo de estos resultados.

Histograma

Un histograma es un conjunto de rectángulos que se determinan representando las frecuencias relativas en el eje vertical contra los límites reales inferiores para cada una de las clases en el eje horizontal del plano cartesiano. Y las distribuciones pueden ser representadas gráficamente por histogramas



Al pasar ya sea de los datos originales o del diagrama tallo y hoja a la distribución de frecuencias o histograma, se pierde parte de la información debido a que ya no se tienen las observaciones o datos originales. Sin embargo, esta pérdida en la información a menudo es pequeña si se le compara con la concisión y la facilidad de interpretación al utilizar la distribución de frecuencias y el histograma.



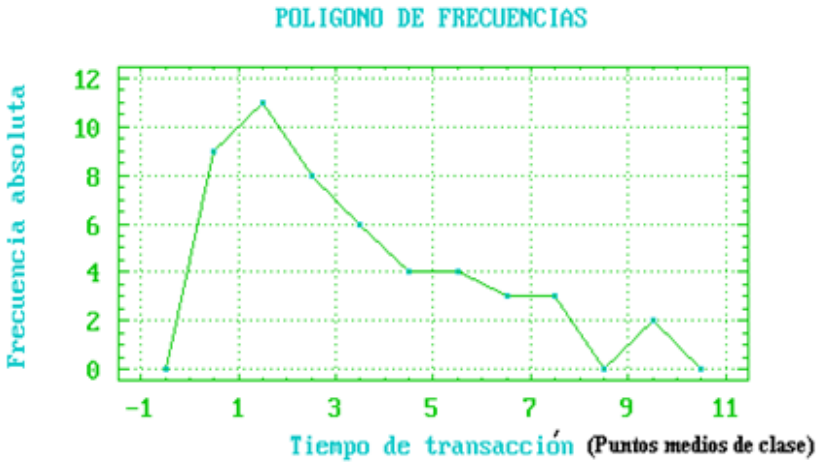
Nota. *Cuando los intervalos de clase son iguales, el área de los rectángulos es proporcional a las frecuencias relativas.*

Sin embargo cuando la longitud de los intervalos es diferente, es necesario ajustar la altura de los rectángulos para que las áreas sean proporcionales a las frecuencias.

Realizamos éstas gráficas en el software ***STATGRAPHICS o Minitab:***

Polígonos de frecuencia

Para trazar un polígono de frecuencias se deben calcular las alturas h_i, en esta actividad de aprendizaje coinciden con las frecuencias absolutas (también relativas), a más de los puntos medios c_i; $i = 1, 2,...,10$ de las clases y luego se une los puntos (c_i , h_i) con segmentos de rectas.



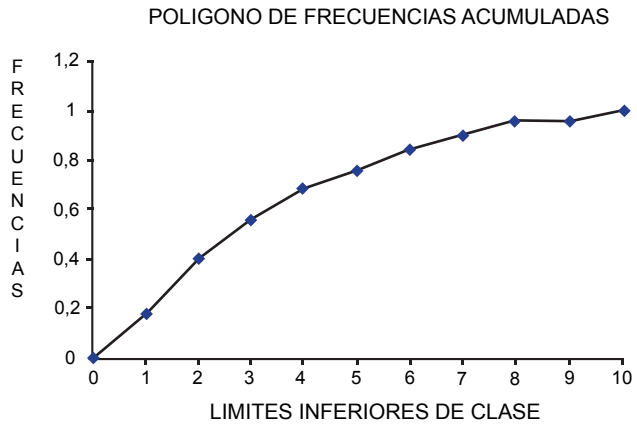
De los dos diagramas presentados se observa que los datos del tiempo de transacción bancaria se agrupan más alrededor de la segunda clase representada por el punto medio 1.5 minutos.

Ojiva

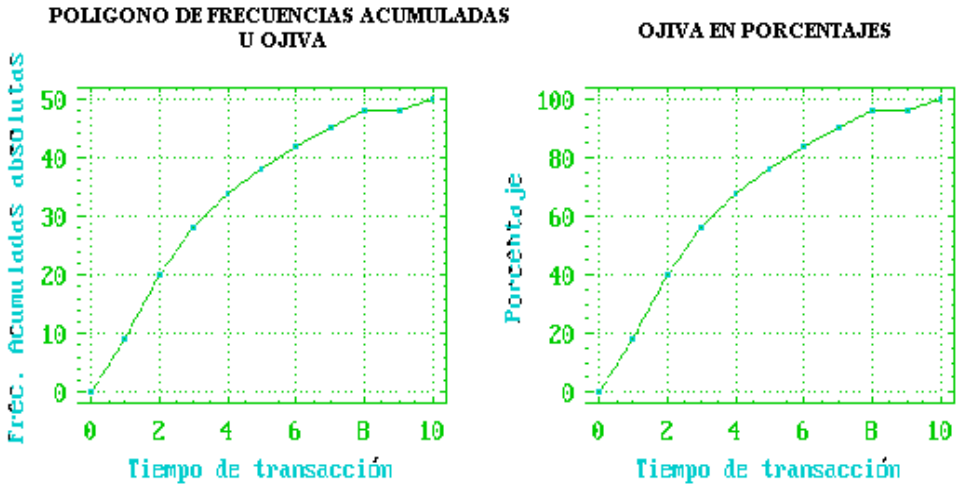
La frecuencia total de todos los valores menores que el límite real superior de un intervalo de clase se conoce como frecuencia absoluta acumulada o relativa acumulada, hasta ese valor de clase inclusive. De esta definición se obtienen los resultados de las dos últimas columnas de la tabla siguiente, es decir, ***las frecuencias acumuladas.***

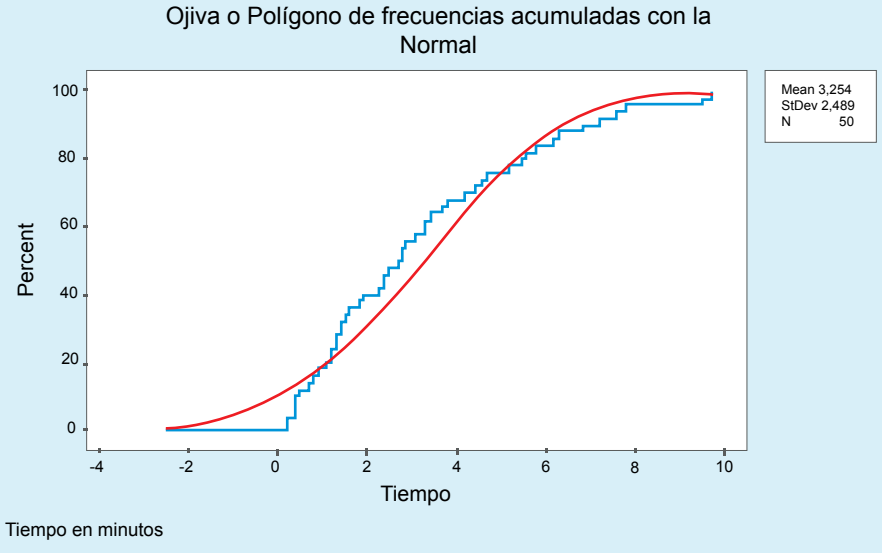
Clase	Límite Inferior	Límite Superior	Marca c_i	Frecuencia de clase n_i	Frecuencia Relativa f_i	Porcentaje f_i *100%	Frecuencia acumulada absoluta F_i	Frecuencia acumulada relativa F_i
(0 - 1]	0	1	0.5	9	0.18	18%	9	0.18
(1 - 2]	1	2	1.5	11	0.22	22%	20	0.40
(2 - 3]	2	3	2.5	8	0.16	16%	28	0.56
(3 - 4]	3	4	3.5	6	0.12	12%	34	0.68
(4 - 5]	4	5	4.5	4	0.08	8%	38	0.76
(5 - 6]	5	6	5.5	4	0.08	8%	42	0.84
(6 - 7]	6	7	6.5	3	0.06	6%	45	0.90
(7 - 8]	7	8	7.5	3	0.06	6%	48	0.96
(8 - 9]	8	9	8.5	0	0.00	0%	48	0.96
(9-10]	9	10	9.5	2	0.04	4%	50	1.00

Las columnas de las frecuencias acumuladas se denominan distribución de frecuencias acumuladas o más brevemente distribución acumulada, la misma que se puede graficar, en el eje vertical se anotará la frecuencia relativa (o absoluta) acumulada de una clase contra el límite inferior de la siguiente sobre el eje horizontal y uniendo con segmentos todos los puntos consecutivos, dan lugar al polígono de frecuencias acumuladas denominada también ojiva.



Diremos que la proporción de clientes en el tiempo menor a 0.0 minutos es 0 o del 0%, la proporción de clientes en el tiempo menor a 1.0 minuto es de 0.18 o del 18%, la proporción de clientes en el tiempo menores a 2 minutos es de 0.4 o del 40%, etc. Se puede realizar también en los software estadísticos Statgraphics y Minitab:





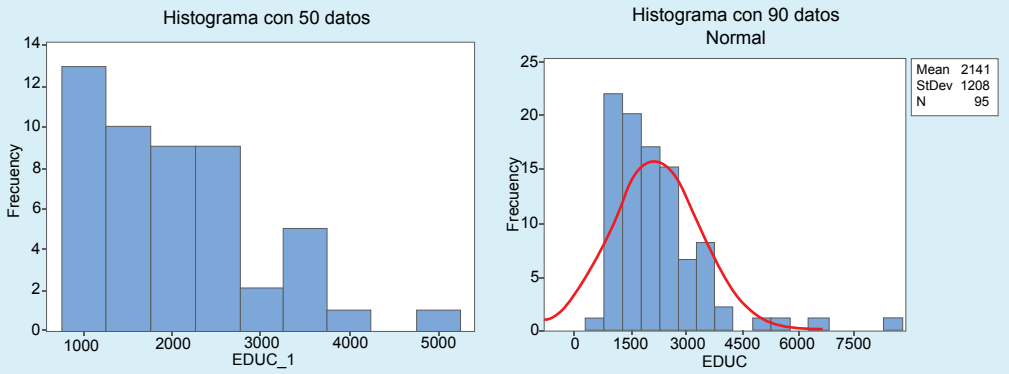
Usamos la distribución acumulada (ojiva) para determinar cuantiles. Con respecto a una distribución acumulada, se define cuantil al valor bajo el cual se encuentra una determinada proporción de los valores de la distribución. El valor del cuantil se lee en dirección opuesta, en el eje horizontal a la proporción correspondiente deseada sobre el eje vertical de una ojiva.

Los cuantiles utilizados comúnmente son los ***percentiles, deciles y cuartiles.*** Los ***percentiles*** son los valores que dividen a una distribución en 100 partes iguales, es decir, cada parte con un porcentaje del 1%; los ***deciles y cuartiles*** son los valores que dividen a la a una distribución en 10 y 4 partes iguales cada una con un porcentaje de 10% y 25% respectivamente. Un cuantil de proporción u orden α denotamos por q_α , Así por ejemplo $q_{0.90} = 7$; $q_{0.80} = 5.5$, en otras palabras se puede decir que $q_{0.9}$ es el valor bajo el cual se encuentra el 90% de los valores de la distribución, es decir, el percentil 90 ó el decil 9.

En términos de probabilidad se definirá el número q_α como:

$$\forall \alpha \in]\,0,1\,[/ P(X < q_\alpha) \leq \alpha,\ P(X > q_\alpha) \leq 1-\alpha$$

Observación. Si aumentamos el número de observaciones de la muestra determinaremos más rectángulos de base pequeña, que producirá un polígono de frecuencias más suave parecido a una campana conocida con el nombre de campana de Gauss o distribución normal.



Realizamos a continuación el manejo estadístico descriptivo gráfico de un problema real, el cual considera datos de una v.a. discreta al mismo tiempo definiremos otras gráficas de importancia y frecuentemente utilizadas.

ACTIVIDAD DE APRENDIZAJE DESARROLLADA 04

Realizar un sondeo de la variable aleatoria "número de hijos por familia en la ciudad de Riobamba".

Cada estudiante realizará un sondeo a 10 familias escogidas al azar, luego estos datos se recopilarán y en grupo deberán construir:

a) Tabla de frecuencias
b) Diagrama de barra, polígono de frecuencias, ojiva
c) Diagrama circular (o pastel).

Observación. Se ha visto que las investigaciones estadísticas sobre un grupo de individuos (muestra) de la población pueden determinar variables medibles cualitativa o cuantitativamente. Las variables cuantitativas pueden tomar valores discretos o continuos definidos sobre un intervalo del conjunto de los números reales, nuestra Actividad de Aprendizaje desarrollada 4 se trata de una variable cuantitativa discreta.

Solución:

Tabla de frecuencias

Los estudiantes al realizar el sondeo denotaron por X la variable discreta "número de hijos por familia en la ciudad de Riobamba", cuyos valores x_i varían entre 0 y 8; de los cuales 6 familias de las 229 encuestadas tienen 0 hijos, 42 tienen 1 hijo, etc., con estos datos se construye la tabla de frecuencias siguiente:

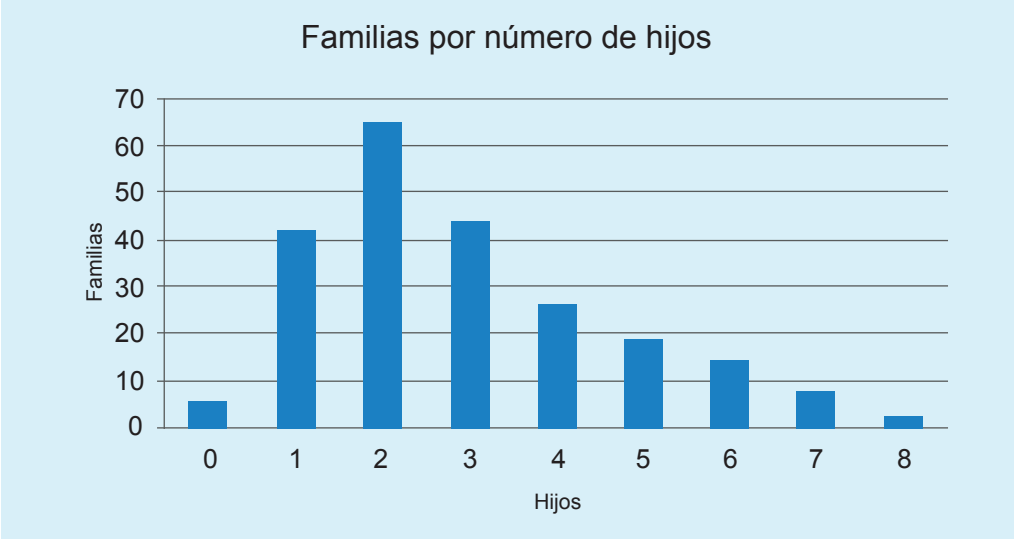
i	Numero de hijos x_i	Frecuencia Absoluta n_i	Frecuencia relativa f_i	Porcentaje f_i*100% n_i	Frecuencia acumulada absoluta F_i	Frecuencia acumulada relativa F_i	Angulo aproximado $\theta = F_i$*360º
1	0	6	0.0262	2,62%	6	0.03	9,43
2	1	42	0.1834	18,34%	48	0.21	75,46
3	2	65	0.2838	28,38%	113	0.49	117,64
4	3	44	0.1921	19,21%	157	0.69	246,81
5	4	27	0.1179	11,79%	184	0.80	289,26
6	5	19	0.0830	8,30%	203	0.89	319,13
7	6	15	0.0655	6,55%	218	0.95	342,71
8	7	8	0.0349	3,49%	226	0.99	355,28
9	8	3	0.0131	1,31%	229	1.00	360,00
Total		229	1.000	100%			

Los valores de la última columna determinan los ángulos (360°´F(x_i)), para cada i, los mismos que nos permitirán trazar el ***diagrama circular*** llamado también ***diagrama pastel o pie en ingles.***

La quinta columna de la tabla anterior expresa los porcentajes de las frecuencias relativas, los cuales se pueden redondear, por ejemplo 8.3% con 8% dicho porcentaje se interpreta como las familias observadas que tienen 5 hijos, el 2,62% es aproximadamente 3%, de las familias entrevistadas no tienen hijos, etc..

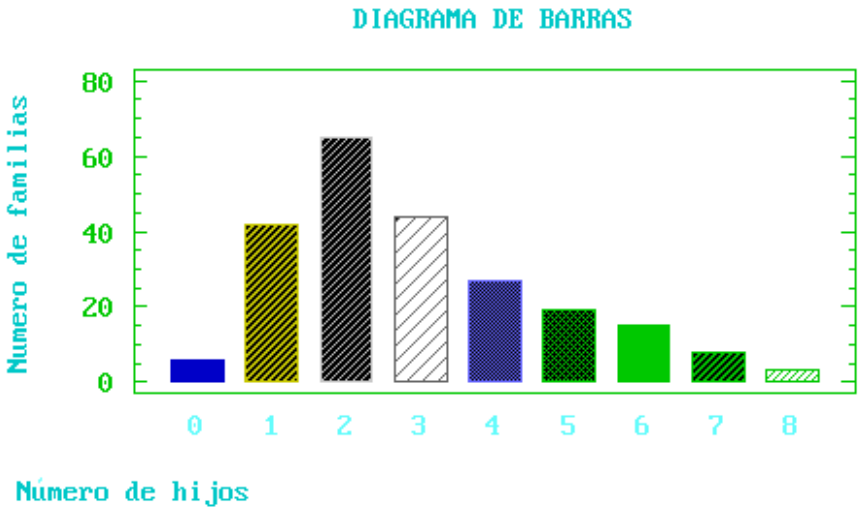
Diagrama de barras, polígono de frecuencias y ojiva

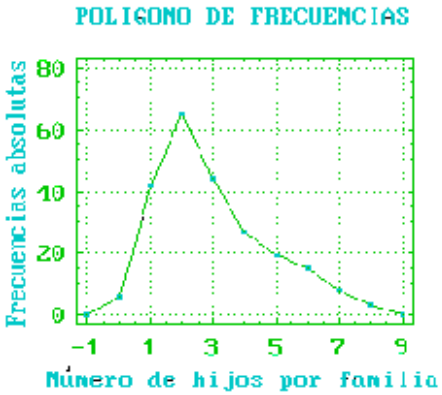
Para el caso discreto como la variable definida X:"número de hijos por familia en la ciudad de Riobamba", los datos pueden ser graficados en un plano cartesiano: los valores $\{x_1,\ldots,x_n\}$ son representados por puntos sobre el eje de las abscisas, las frecuencias son representadas por segmentos paralelos al eje de las ordenadas. El gráfico que se obtiene se llama diagrama de barras.

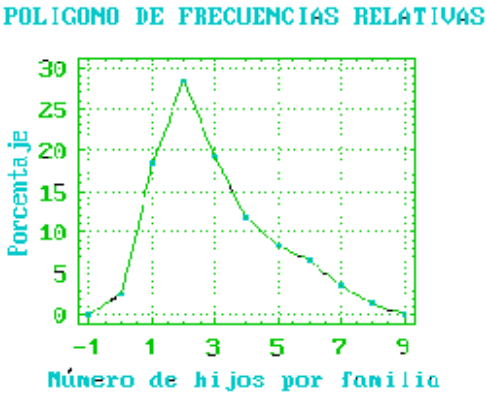


Nota. *Es verificable que si un segmento es proporcional a la frecuencia relativa lo es también con respecto a la frecuencia absoluta, en ambos casos las gráficas serán semejantes.*

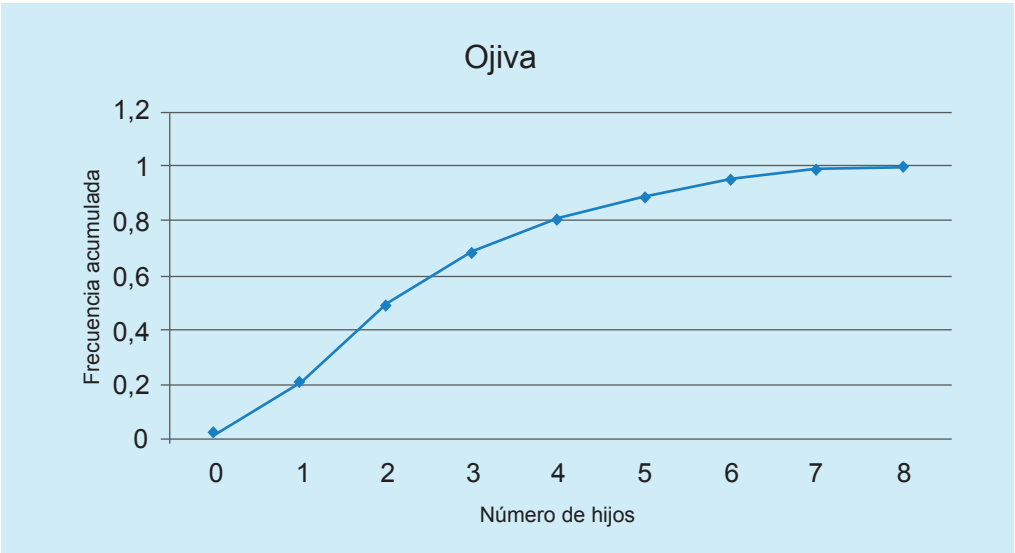
Comúnmente se da preferencia a las frecuencias relativas, porque la escala vertical tiene un intervalo fijo (de 0 a 1 incluidos) para la mayoría de los manejos estadísticos de datos.



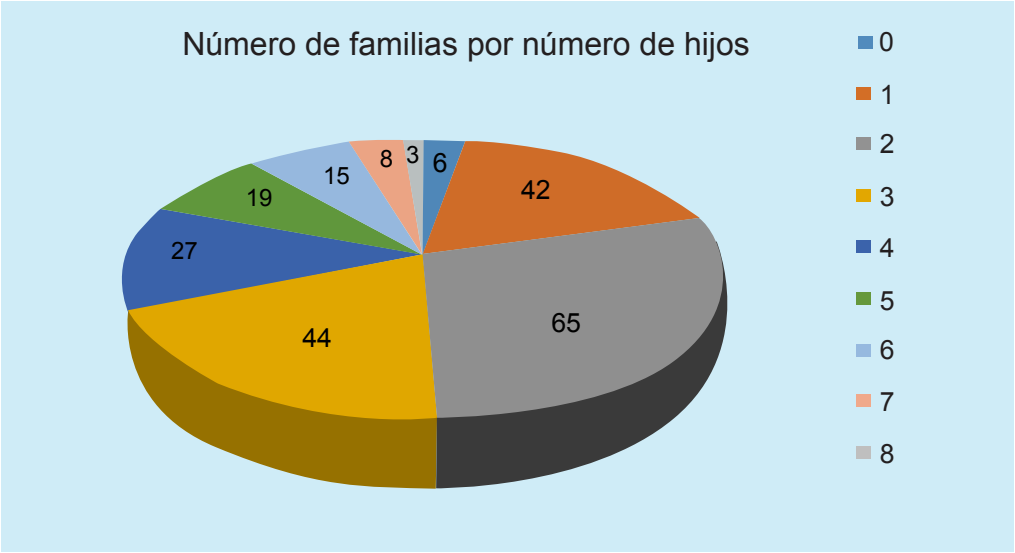




Las columnas de las frecuencias absolutas o relativas acumuladas (antepenúltima y penúltima columnas respectivamente) de la tabla de frecuencias del literal a) denotadas por $F(x_i)$ y es la frecuencia relativa total de todos los valores menores o iguales al valor x_i y se conoce con el nombre de distribución de ***frecuencias acumuladas*** o simplemente ***distribución acumulada***. Un gráfico que muestre las frecuencias acumuladas se denomina ***ojiva.***



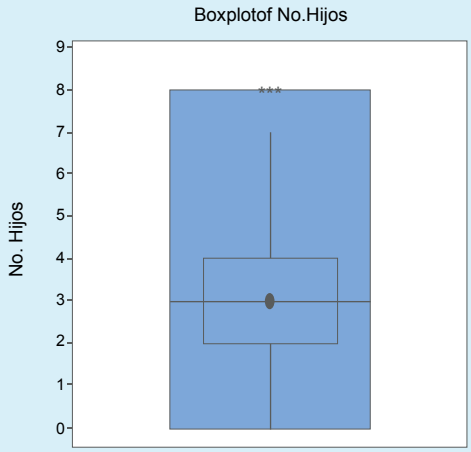
También se realiza el ***diagrama circular*** en la hoja electrónica de Excel y nótese que la porción más grande es la de mayor frecuencia que corresponde a las 65 familias que tienen 2 hijos y el ángulo es 360°´ 0.2838 = 102° redondeados. De la misma manera se calculan los ángulos de las otras porciones del ***diagrama circular o pastel (pie).***



Observación. Tanto los diagramas de barras como los histogramas, así como los diagramas de tallo y hoja o el diagrama de caja (Box and wisher plot) y también el diagrama circular tienen por objeto:

1) Demostrar o presentar el perfil de distribución de los datos. El conocimiento de este perfil es útil en varias situaciones como sugerirán los análisis apropiados de la estadística inferencial: estimación de parámetros, prueba de hipótesis, análisis de la varianza o ANOVA.

2) Dar una idea de la dispersión y la ubicación de algunas medidas de tendencia central (moda, mediana y media), como el box and wisher plot.



Este diagrama de caja representa al número de hijos por familia y se dice también en inglés. Box and Wisher plot

Como vemos el uso gráfico nos ayuda a extraer información acerca de propiedades de un conjunto de datos. Por ejemplo el diagrama tallo y hoja proporciona al observador la asimetría y presentación de datos anómalos denominados también atípicos entre otras propiedades de los datos. La hipótesis de normalidad puede ser convalidada con presentaciones semejantes al diagrama de tallo y hoja, diagrama de caja y bigotes e histogramas o diagrama de barras.

Resumiendo, las presentaciones gráficas pueden dar al analista una impresión de la variabilidad, la agrupación, la tendencia de la distribución.

Antes de describir numéricamente un conjunto de datos tengamos presente que estamos viviendo la "Era de la Información". La cantidad de datos es tan grande que se hace necesario estudiar parte de esa información para tomar una decisión o decisiones. Sin embargo, al seleccionar los datos o al realizar muestreo vemos que no es nada fácil. Y nos cuestionamos, por ejemplo: ¿Qué queremos conocer?, ¿Cuáles son los costos, no solo del proyecto sino también del error cometido en la toma de decisiones?, ¿Qué error podemos cometer?, ¿Cómo interpretamos los resultados del manejo estadístico de datos?. Al realizar el manejo estadístico de datos con base en una muestra debemos tomar muy en cuenta: especificación, proyecto y evaluación.

- ***Especificación:*** Define el máximo error que puede ser cometido.
- ***Proyecto:*** Consiste en obtener la confiabilidad deseada al menor costo posible y utilizando las facilidades físicas y los recursos humanos disponibles.
- ***Evaluación:*** Verifica las diferencias entre los procedimientos utilizados para la comparación de los resultados.

Estos tres aspectos son mutuamente dependientes.

En un ambiente académico, la investigación educativa se desarrollará si tomamos en cuenta los siguientes pasos, bajo ningún modo constituye un itinerario completo para un proyecto de investigación estadística pero si nos ayudará a resolver muchos problemas de ésta importante labor de la docencia.

La recopilación de la información requerida la planteamos en las encuestas, cuestionarios, entrevistas, etc.

El software : ***MINITAB,*** y la Hoja electrónica ***EXCEL*** nos ayudarán al manejo estadístico de datos; esto es, la recopilación, presentación, análisis, modelización e interpretación de la información y todo para ayudar a tomar decisiones y sacar conclusiones en base a la información y resultados obtenidos. Y para conseguir los resultados esperados correctamente aplicamos la ***metodología de la investigación estadística*** siguiendo los pasos:

- Planeación de la investigación: problema, objetivos, hipótesis
- Elaboración de los instrumentos del manejo de datos.
- Prueba piloto
- Selección de la muestra piloto
- Elaboración definitiva de los instrumentos del manejo de datos.
- Selección y entrenamiento de los encuestadores.
- Recolección y presentación de datos
- Análisis, modelización e interpretación estadística de datos
- Toma de decisiones y sacar conclusiones de los resultados obtenidos y emitir
- Informe de la investigación

2.2 DESCRIPCIÓN NUMÉRICA DE DATOS

En el párrafo 1.2 (Alguna Terminología Necesaria) del primer capítulo vimos los términos básicos de población, muestra, variables aleatorias discretas y continuas entre otros.

Para describir y graficar una colección de datos se ha considerado: el diagrama tallo y hoja, el diagrama de caja y bigotes, el diagrama de barras, el histograma, los polígonos de frecuencias, diagramas circulares y ojivas.

Estos diagramas y gráficos proporcionan información útil respecto al conjunto de datos, pero no es muy adecuado para sacar conclusiones basadas sobre el mismo conjunto, sobre todo porque no está bien definido. Es decir, se podría elaborar por ejemplo muchos histogramas similares a partir del mismo conjunto de mediciones los cuales dependen de la escala que se tome y el número de clases que se elija.

Para hacer inferencias, es decir, para sacar conclusiones respecto a una población basadas en la información contenida en una muestra y medir la bondad de inferencia, se requieren de cantidades obtenidas de expresiones rigurosamente definidas.

Es posible obtener mediante las Matemáticas ciertas propiedades de esas cantidades muestrales y establecer conclusiones probabilísticas con respecto a la validez de las inferencias.

En esta parte, intentaré despertar el interés para poder describir de mejor manera los resultados que se han obtenido de los datos observados.

Preguntamos en el curso a los estudiantes la asignatura y el deporte de su preferencia, dichos datos anotamos y contamos.

El número de estudiantes que han preferido una misma asignatura o un mismo deporte. Al hacer esta actividad, introducimos el primer índice estadístico, la ***moda***, esto es, el dato que se presenta con mayor frecuencia.

Las cantidades que se pretenden definir son medidas numéricas descriptivas de un conjunto de datos para determinar características importantes del mismo. Se obtendrán dos tipos de medidas: las medidas de tendencia central y las medidas de dispersión o variación.

2.2.1 MEDIDAS DE TENDENCIA CENTRAL

Las medidas de tendencia central llamadas también de localización que presentamos son: la ***moda,*** la ***mediana*** y la media aritmética o ***media*** simplemente, presentado así por el orden de dificultad que estas tienen.

Es simple darse cuenta que la moda refleja el significado común de la palabra, sin embargo la media es la más conocida. En raras ocasiones se escucha hablar de moda y de mediana no obstante dichos índices describen de manera diferente un conjunto de datos, por lo que seleccionamos estas medidas o índices dependiendo de la naturaleza del fenómeno que se estudie y de los objetivos que se han trazado en la metodología de la investigación estadística.

Moda: Es la observación que se presenta con mayor frecuencia en la muestra; además muestra hacia qué valor tienden los datos a agruparse.

RECORDANDO

Para el caso discreto: Gráficamente en el diagrama de barras la moda está representada por el segmento o barra de mayor longitud. En nuestra Actividad de Aprendizaje desarrollada 4 del número de hijos por familia, la moda es 2 hijos por familia, pues tiene frecuencia 65 y es la mayor de todas.

Para el caso continuo: La clase con la frecuencia más alta recibe el nombre de clase modal con lo que se define a la moda, al punto medio de esa clase. En este caso la clase modal sirve como punto de concentración en el conjunto de datos.

En el histograma la clase modal está representada por el rectángulo de mayor área. En la Actividad de Aprendizaje desarrollada 3, de los tiempos de transacción bancaria, la clase modal está dada por el intervalo (1, 2], entonces la moda es 1.5, es decir, el punto medio de éste intervalo que se calcula por: (1+2)/2.

Otro aspecto que debemos tomar en cuenta es el siguiente. Puede existir en una muestra más de una moda. Por ejemplo, consideremos las siguientes observaciones.

-2, 1, -4, 2, 1, 0, -2, 10, 1, -2, 0, 7, -1 y 4

Las modas son -2 y 1, puesto que ambos valores presentan el mismo número de veces; tres y ningún otro más lo hace con mayor frecuencia. En este caso se dice que la muestra es ***bimodal.*** También debemos estudiar la observación que ocupa el lugar central entre los datos ordenados de forma creciente o decreciente o de manera ascendente o descendente.

Mediana. En un conjunto de datos ordenados de manera creciente, es el valor para el cual, la mitad de éstos es menor que éste valor y la otra mitad mayor. Si el conjunto de n observaciones o datos es impar, esto es 2m+1 con $m \epsilon Z^+$ (Z^+ enteros positivos), la mediana es el (m+1)-ésimo dato del conjunto ordenado. En tanto, si el número de datos es par, esto es 2m con $m \epsilon Z^+$, se considera la mediana como la suma de los valores x_m y x_{m+1} y divididos para 2.

Nota. *Se note que el valor de la mediana de datos no necesariamente es un valor observado.*

La ventaja de la mediana es que los valores extremos no tienen mucha influencia sobre ella. Para ilustrar lo dicho, consideremos que las observaciones de una muestra son:

-10, 2, 3, 4, 5 y 17
La mediana de los datos es 3,5

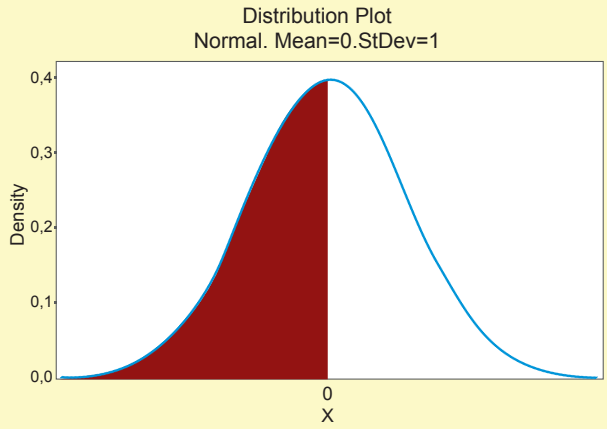
RECORDANDO

Para el caso discreto: Al ordenar los 229 datos de la Actividad de Aprendizaje desarrollada 4 de manera creciente, esto es 0,...,8 y siendo este impar se tiene $2m + 1 = 229$, despejando se tiene que $m = 114$, entonces la mediana es la observación o el dato de posición 115 cuyo valor es 3, es decir, x_{115} = Mediana = 3.

Para el caso continuo: Si los datos de la Actividad de Aprendizaje desarrollada 3, los ordenamos de manera creciente o ascendente tenemos:

0.2	0.4	0.4	0.4	0.5	0.5	0.7	0.9	0.9	1.1
1.2	1.2	1.3	1.3	1.4	1.4	1.5	1.6	1.8	1.9
2.3	2.4	2.4	2.5	2.7	2.8	2.8	2.9	3.1	3.3
3.3	3.4	3.7	3.8	4.2	4.4	4.6	4.7	5.2	5.5
5.6	5.8	6.2	6.3	6.8	7.2	7.6	7.8	9.5	9.7

Donde el número de observaciones es par, se tiene $2m = 50$, despejando se tiene $m = 25$, luego las observaciones de posiciones 25 y 26, es decir, $x_{25} = 2.7$ y $x_{26} = 2.8$, entonces el valor de la mediana es igual a $(x_{25} + x_{26})/2 = 2.75$, pero esto realizamos cuando los datos no son agrupados.



En el gráfico se presenta una distribución normal estándar y la mediana y la moda coinciden con el valor 0 y el área sombreada es igual a 0.5 o 50%

La mediana para datos agrupados se calcula por el valor que divide en dos partes iguales la distribución. La fórmula computacional para la mediana de los datos agrupados, está dada por:

$$Mediana = L + \frac{\frac{n}{2} - \sum f}{f_m} c$$

RECORDANDO

Donde los valores para nuestra Actividad de Aprendizaje desarrollada 3 son:

- L es el límite inferior de la clase mediana (la clase que contiene la mediana de los datos no agrupados) y es 2.
- n es el número total de datos y es 50.
- Σf es la suma de las frecuencias de todas las clases por debajo de la clase mediana y es 20.
- f_m es la frecuencia de la clase mediana y es 8.
- c es la longitud del intervalo de la clase mediana y es 1

Por tanto, aplicando la fórmula computacional de la mediana tenemos que Mediana = 2.625 ≅ 2.6.

Media: La media de n observaciones $x_1, \ldots, x_n$ es el promedio aritmético de los mismos (suma de todos los valores divididos para el total n) y denotamos por $\bar{X}$, donde

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

Observación. El símbolo $\bar{x}$ se referirá al valor de la media de una muestra. Es decir, es la ***media muestral***. La media de todas las observaciones de una población se representará con el símbolo μ que se lee "mu". Obsérvese que en general no es posible medir μ más bien μ es un parámetro es una cantidad desconocida que se desea estimar a partir de la información de una muestra, en general se estima con $\bar{x}$. La media de un conjunto de observaciones solamente localiza el centro de la distribución de los datos; por sí misma no ofrece una descripción adecuada de un conjunto de observaciones.

En general si las observaciones x_i tienen una frecuencia absoluta n_i con $i = 1,\ldots,m$ y $\Sigma n_i = n$, entonces la media viene calculada por las fórmulas:

Para el caso discreto

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

Para el caso continuo

$$\bar{x} = \frac{\sum_{i=1}^{k} n_i c_i}{n}$$

donde c_i es el punto medio de la clase i-ésima.

RECORDANDO

Para el caso discreto: Se puede ver que m = 9. En la Actividad de Aprendizaje desarrollada 4, luego tomando los resultados de la tabla siguiente tenemos.

$$\bar{X} = \frac{(n_1 x_1 + \ldots + n_9 x_9)}{229} = \frac{677}{229} = 2.9 \cong 3$$

En efecto los cálculos se exponen en la siguiente tabla.

i	n_i	x_i	$n_i x_i$
1	6	0	0
2	42	1	42
3	65	2	130
4	44	3	132
5	27	4	108
6	19	5	95
7	15	6	90
8	8	7	56
9	3	8	24
Total	229		677

Por tanto interpretando este valor se tiene que el número medio de hijos por familia en la ciudad de Riobamba es aproximadamente 3.

Para el caso continuo: Tomando los datos de frecuencias del literal a) de la Actividad de Aprendizaje desarrollada 3, donde k = 10, tenemos la tabla siguiente

i	n_i	c_i	$n_i c_i$
1	9	0.5	4.5
2	11	1.5	16.5
3	8	2.5	20
4	6	3.5	21
5	4	4.5	18
6	4	5.5	22
7	3	6.5	19.5
8	3	7.5	22.5
9	0	8.5	0
10	2	9.5	19
Total	n=50		163

El valor de la media muestral es más preciso que con cada observación. Si consideramos los datos:

-10, 2, 3, 4, 5 y 17

Donde la mediana y la media es 3.5, esto se debe a que las observaciones son simétricas. Pero note que pasa con los datos 1, 2, 3, 4, 5, 17 nuevamente la mediana es 3.5 y la media es 5.33.

En este caso, es evidente que la media muestral no dice mucho con respecto a la tendencia central de la mayor parte de los datos. Sin embargo, la mediana sigue siendo 3.5 y ésta es, probablemente, una medida de tendencia central más significativa para la mayor parte de los datos.

En conclusión si los datos son simétricos la media y la mediana coinciden. Si además, los datos tienen una sola moda (esto es, son unimodales), entonces la moda, la mediana y la media coinciden. Si los datos están sesgados (esto es, son asimétricos, con una larga cola en uno de los extremos) entonces se tiene que

moda < mediana < media

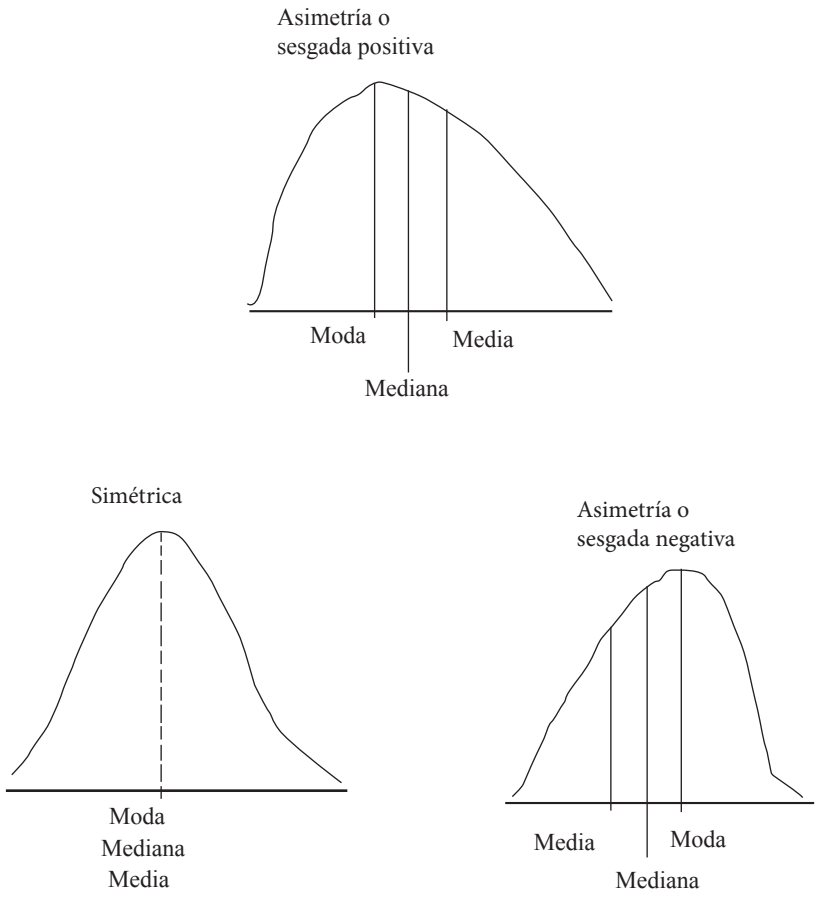
Si la distribución está sesgada a la derecha.

Mientras que

media < mediana < moda

Si la distribución está sesgada a la izquierda.

Figura 4: Asimetría de los datos



Relación entre la moda, mediana y media

Generalmente la media muestral es más estable que la mediana, en el sentido que ésta no cambia mucho de una muestra a otra. En consecuencia, muchas técnicas estadísticas analíticas utilizan la media muestral. Sin embargo, la moda y la mediana se utilizan mucho como medidas descriptivas de los datos.

2.2.2 MEDIDAS DE DISPERSIÓN

Las medidas de tendencia central no necesariamente proporcionan información suficiente para describir datos de manera adecuada. Por ejemplo, considérense las tres muestras de datos:

Muestra 1: 1, 2, 3, 4, 5, 6;
Muestra 2: 1, 1, 1, 6, 6, 6;
Muestra 3: -13, 2, 3, 4, 5, 20

ESTADISTICAS DESCRIPTIVAS	*Muestra 1*	*Muestra 2*	*Muestra 3*
Media	3,5	3,5	3,5
Error típico	0,76	1,12	4,28
Mediana	3,5	3,5	3,5
Moda	#N/A	1	#N/A
Desviación estándar	1,87	2,74	10,48
Varianza de la muestra	3,5	7,5	109,9
Curtosis	-1,2	-3,33	2,34
Coeficiente de asimetría	0	-6,7E-17	0
Rango	5	5	33
Mínimo	1	1	-13
Máximo	6	6	20
Suma	21	21	21
CV	53%	78%	300%
Cuenta	6	6	6

Al realizar cálculos en Excel tenemos que la mediana y la media coinciden con el valor 3,5; dicho valor se creería que es representativo para los tres grupos, sin embargo, se observa a simple vista que la dispersión o variabilidad de la muestra 3 es mucho mayor que la muestra 2 y ésta última es mayor que la muestra 1.Veremos más adelante que una medida para comparar grupos de datos sin tomar en cuenta sus dimensiones es el coeficiente de variación o simplemente CV. La muestra 3 tiene el mayor CV, 300%, de los tres grupos de datos. Por lo tanto la media y la mediana no son suficientes para describir un conjunto de datos por lo que éstas medidas no describen de manera adecuada a los tres muestras.

En una tarea donde se manejen datos estadísticos (calificaciones de exámenes, edades o estaturas de los alumnos, etc.) es necesario saber la variación de los datos o saber que tan dispersos están entre ellos o respecto a una medida de tendencia central. Ahora daremos algunas definiciones que realizan esta actividad como por ejemplo el rango o recorrido, el rango intercuartílico, el rango interdecílico, la varianza, la desviación estándar y el coeficiente de variación o simplemente CV, etc.

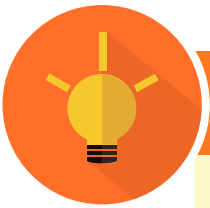
Una de las medidas de dispersión más elemental es el rango de una muestra.

Rango o Recorrido r: Es la diferencia entre el valor máximo y el valor mínimo de las observaciones. Entonces, el rango es

$$r = máx(x_i) - min(x_i)$$

Para las tres muestras dadas anteriormente, el recorrido de la muestra 3 es $r_3 = 20 - (-13) = 33$, mientras que de la muestra 2 es $r_2 = 6 - 1 = 5$ y de la muestra 1 es también 5 ($r_1 = 6 - 1 = 5$). De estos resultados es claro que entre más grande sea el rango, mayor será la variabilidad en los datos. Sin embargo no es suficiente esta medida, pues la variabilidad de las muestras 1 y 2 es notoria y su valor es el mismo, 5 y es necesario entonces definir otras medidas de variabilidad como la varianza, la desviación estándar, entre otras.

Nota. *Como regla general se debe evitar el uso del recorrido como medida de variabilidad, cuando el número de observaciones en un conjunto es demasiado grande o cuando éste contenga algunas observaciones cuyo valor sea relativamente grande. Para muchos problemas tiene una mayor utilidad determinar el recorrido entre dos valores cuantiles, que entre dos valores extremos, es decir, se considera un porcentaje de ellos; tomándose generalmente un 50% o un 80% del total de los datos se calcula a través de la gráfica de la ojiva o de los percentiles, cuartiles y deciles. Por tanto se hace necesario definirlos.*



RECORDANDO

Para el caso discreto: De los datos de la Actividad de Aprendizaje desarrollada 4 se tiene que el máximo valor es 8 y el mínimo es 0, por tanto el recorrido o rango es $r = 8 - 0 = 8$.

Para el caso continuo: De los datos no agrupados de la Actividad de Aprendizaje desarrollada 3 se observe que el valor máximo es 9.7 y el valor mínimo es 0.2, por tanto el recorrido o rango es
$r = 9.7 - 0.2 = 9.5$.

Percentiles, Cuartiles y Deciles

Se ha definido la mediana como el valor que divide los datos ordenados en dos partes iguales. Es posible también dividir los datos ordenados en más partes iguales como por ejemplo, dividir en cuatro o en diez partes, los tres o nueve puntos de división se conocen como *cuartiles o deciles* respectivamente. Generalmente se denotan los tres cuartiles por Q_1, Q_2 y Q_3 y a los nueve deciles por D_1, D_2, D_3, D_4, D_5, D_6, D_7, D_8 y D_9.

Nota. *Nótese que el primer cuartil es el 25 percentil, es decir,* $Q_1 = p_{0.25}$, *el segundo cuartil es el 50 percentil, es decir,* $Q_2 = D_5 = p_{0.50}$ *que es la mediana y el tercer cuartil es el 75 percentil, es decir,* $Q_3 = p_{0.75}$.

Si un conjunto ordenado de datos se divide en cien partes iguales, los 99 puntos de división se denominan percentiles. De manera general, se denota el 100k-ésimo percentil por p_k y se define como el valor tal que al menos el k*100% de las observaciones están en el valor o por debajo de él, y al menos el 100(1-k)% están en el valor o por encima de él. A continuación presentamos el procedimiento para calcular los percentiles p_k a partir de los datos ordenados.

Procedimiento para calcular el k-ésimo percentil de un conjunto ordenado de datos:

Paso 1. Encontrar el número de la posición i del percentil mediante el cálculo de n_k

a) Si n_k no es un entero, entonces i es el entero inmediato superior.
b) Si n_k es entero, entonces i es igual a $n_k + 0.5$

Paso 2.

a) Si i es un entero, cuéntese desde la observación más pequeña hasta hallar el i-ésimo valor.
b) Si i no es un entero, entonces contiene una fracción igual a un medio, con lo que el valor de p_k es el promedio de las observaciones ordenadas n_k y $(n_k + 1)$

RECORDANDO

1. Se desea encontrar los percentiles 25 y 92 de los datos de la Actividad de Aprendizaje desarrollada 3, calculamos primero $p_{0.25}$. Puesto que k = 0.25, n_k = 50(0.25) =12.5, que no es un entero, entonces el número de la posición es i = 13. Por tanto, el percentil 25 o el primer cuartil es la observación ordenada número 13, esto es $p_{0.25}$ = 1.3. El percentil 92 se encuentra de manera parecida. Puesto que ahora k = 0.92, n_k = 50(0.92) = 46 es un entero, el número de la posición es i = 46 + 0.5, el cual es el promedio de las observaciones cuarentaiseisava y cuarentaisieteava. Luego, el percentil 92 es $p_{0.92}$= (7.2 + 7.6)/2 = 7.4.

2. Compruebe con el mismo procedimiento el valor de los percentiles siguientes para los datos de la Actividad de Aprendizaje desarrollada 3: $p_{0.10}$ = 0.5, $p_{0.50}$ = 2.75, $p_{0.75}$ = 4.7, $p_{0.90}$ = 7.

Definición.- La diferencia entre los percentiles 75 avo (tercer cuartil Q_3) y 25 avo (primer cuartil Q_1) recibe el nombre de recorrido intercuartil y denotamos por

$$R.Ic. = Q_3 - Q_1$$

El recorrido intercuartil para la Actividad de Aprendizaje desarrollada 3 es: $R.Ic.$= 4.7 - 1.3 = 3.4

Definición.- La diferencia entre los percentiles 90 avo (noveno decil D_9) y 10 avo (primer decil D_1) recibe el nombre de recorrido interdecil y denotamos por

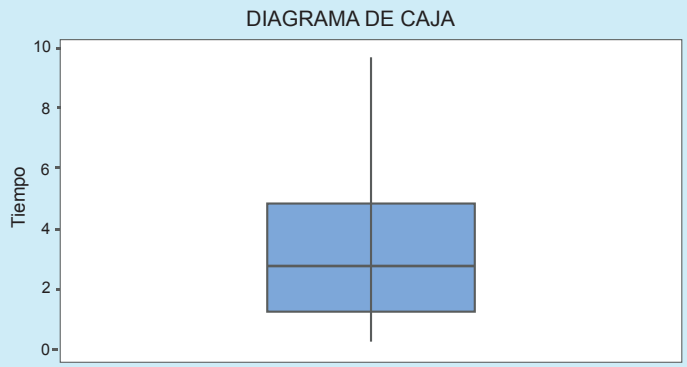
$$R.Id. = D_9 - D_1$$

El recorrido interdecil para la Actividad de Aprendizaje desarrollada 4 es: $R.Id.$= 7.0 - 0.5 = 6.5

Observando las gráficas de las ojivas para los dos casos discreto y continuo, se puede determinar valores aproximados de percentiles y calcular los recorridos. Tanto el recorrido intercuartil como el recorrido interdecil nos ayudan en muchas investigaciones estadísticas a separar datos que alteran las medidas descriptivas reales y por ende modifican las conclusiones respecto a la población. Un gráfico para observar datos contenidos (en un rectángulo) el recorrido intercuartil es el llamado Box-and-Whisker Plot traduciendo sería ***diagrama de caja y bigotes*** ó simplemente ***diagrama de caja.***

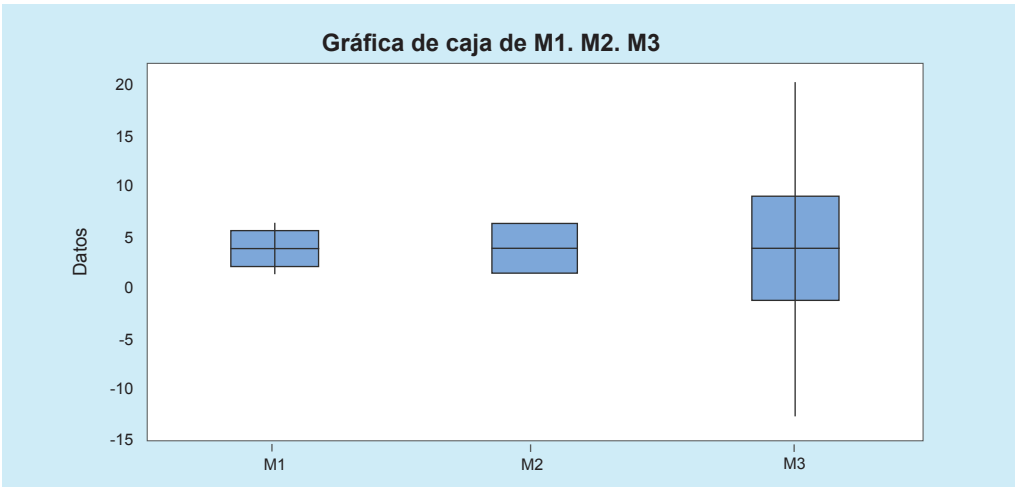
El diagrama de caja presenta los tres cuartiles y los valores mínimo y máximo de los datos sobre un rectángulo, alineado horizontal o verticalmente. El rectángulo delimita el rango intercuartílico con la arista izquierda (o inferior) ubicada en el primer cuartil, Q_1 y la arista derecha (o superior) en el tercer cuartil, Q_3. Se dibuja una línea a través del rectángulo en la posición que corresponde al segundo cuartil es decir la mediana. De cualquiera de las aristas del rectángulo se extiende una línea o bigote que va hacia los valores extremos. Estas son observaciones que se encuentran entre 0 y 1.5 veces el rango intercuartílico a partir de las aristas del rectángulo o que están más allá de tres veces.

El diagrama de caja es una presentación visual que describe al mismo tiempo varías características importantes de un conjunto de datos, tales como el centro, la dispersión, la desviación de la simetría y la identificación de observaciones que se alejan de manera poco usual del resto de los datos. Este tipo de observaciones se conocen como valores ***atípicos o anómalos.***

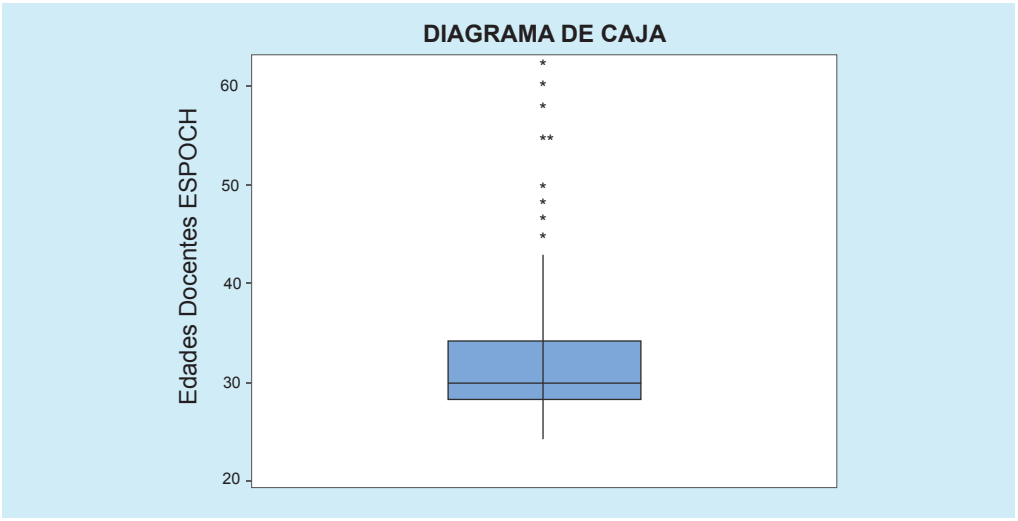


El diagrama de caja de la Actividad de Aprendizaje desarrollada 3 indica que la distribución de los tiempos de transacción es asimétrica a la derecha.

Los diagramas de caja son muy útiles al hacer comparaciones gráficas entre conjunto de datos, ya que tienen un gran impacto visual y son fáciles de comprender, para el ejemplo de las tres muestras M1, M2 y M3, se observa que la mediana de las tres cajas coincide en su valor 3.5 y la caja de la muestra M3 es más grande y más largos los bigotes lo que nos indica una mayor dispersión, se note que el diagrama de caja de la muestra M2 no tiene bigotes ¿por qué?



Para visualizar si se presenta o no datos anómalos (atípicos) se ha recopilado las edades de 83 docentes de la Escuela Superior Politécnica de Chimborazo y en MINITAB se ha elaborado:



Obervación. Se observe que los datos anómalos vienen representados en la gráfica anterior por * y aparecen sobre el bigote superior. ¿Cuántos datos anómalos observa? Verdad ¿que no es posible distinguir éstos? Para responder afirmativamente realice un diagrama de tallo y hoja.

Varianza: La varianza de las observaciones $x_1,...,x_n$ es el promedio del cuadrado de las distancias entre cada observación y la media del conjunto de observaciones. La ***varianza muestral*** de las observaciones se denota por S'^2 y es:

$$s'^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^2$$

Generalmente se utiliza la varianza empírica nuevamente llamada varianza en los cálculos estadísticos muestrales (por ser un buen estimador de la varianza poblacional σ^2), que se define y denota por:

$$s^2 = \frac{1}{n\text{-}1}\sum_{i=1}^{n}(x_i - \bar{X})^2 \qquad (1)$$

Nota. *A mayor varianza dentro del conjunto de observaciones corresponde una mayor dispersión dentro del mismo conjunto. La varianza es útil en la comparación de la variación relativa de dos conjuntos de observaciones, pero sólo aporta información con respecto a la variación en un sólo conjunto de datos cuando se interpreta en términos de la desviación estándar.*

Desviación estándar. La desviación estándar de un conjunto de observaciones es la raíz cuadrada positiva de la varianza, es decir:

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

Para el caso discreto: Cuando los datos x_i tienen frecuencia absoluta n_i la varianza se expresa por la fórmula:

$$s^2 = \sum_{i=1}^{m}\frac{n_i(x_i - \bar{X})^2}{n\text{-}1}$$

Tomando los datos de la Actividad de Aprendizaje desarrollada 4 donde $\bar{X} = 2.96$ se tiene que $S^2 = 3{,}193$ entonces $S = 1.79$

El uso de la ecuación (1) puede dar origen de errores grandes de redondeo. Obtengamos una fórmula más computacional.

$$s^2 = \sum_{i=1}^{n}\frac{(x_i - \bar{X})^2}{n\text{-}1} = \sum_{i=1}^{n}\frac{(x_i^2 - 2\bar{X}x_i + \bar{X}^2)}{n\text{-}1} = \frac{\Sigma x_{i=1}^{n} x_i^2 - 2\bar{X}\Sigma_{i=1}^{n} x_i + \Sigma_{i=1}^{n}\bar{X}^2}{n\text{-}1} = \frac{\Sigma_{i=1}^{n} x_i^2 - \frac{(\Sigma_{i=1}^{n} x_i)^2}{n}}{n\text{-}1}$$

Luego la desviación estándar está dada por:

$$s = \sqrt{\frac{\Sigma_{i=1}^{n} x_i^2 - \frac{(\Sigma_{i=1}^{n} x_i)^2}{n}}{n-1}}$$

Para datos agrupados en k grupos, la varianza viene dada por la expresión:

$$s^2 = \frac{\sum_{i=1} n_i x_i^2 - \frac{(\sum_{i=1}^{k} n_i x_i)^2}{n}}{n\text{-}1}$$

Para el caso continuo: Para datos agrupados, puede calcularse el valor aproximado de la varianza mediante el uso de la fórmula:

$$s^2 = \frac{\sum_{i=1}^{k} n_i (c_i - \bar{X})^2}{n - 1} = \frac{\sum_{i=1}^{k} n_i c_i^2 - \frac{(\sum_{i=1}^{k} n_i c_i)^2}{n}}{n - 1}$$

Donde k es el número de clases, n_i es la frecuencia de la clase i-ésima, c_i es el centro o punto de la clase i-ésima y $\sum_{i=1}^{k} n_i = n$

La fórmula para la desviación estándar es

$$s = \sqrt{\frac{\sum_{i=1}^{k} n_i (c_i - \bar{X})^2}{n - 1}}$$

Para nuestra Actividad de Aprendizaje desarrollada 3 tenemos:

c_i	c_i^2	n_i	$n_i c_i$	$n_i c_i^2$
0.5	0.25	9	4.5	2.25
1.5	2.25	11	16.5	24.75
2.5	6.25	8	20.0	50.00
3.5	12.25	6	21.0	73.50
4.5	20.25	4	18.0	81.00
5.5	30.25	4	22.0	121.00
6.5	42.25	3	19.5	126.75
7.5	56.25	3	22.5	168.75
8.5	72.25	0	0.0	0.00
9.5	90.25	2	19.0	180.50
Total		50	163.0	828.50

Luego por la fórmula computacional, $s^2 = \frac{(828.5 - 531.38)}{49} = 6.064$, entonces $S = 2.4625$

Es útil comparar la variabilidad de dos o más conjuntos de datos que difieren de manera considerable en la magnitud de las observaciones, por ejemplo si tenemos los pesos en kilogramos y las estaturas en centímetros, de los estudiantes de la ESPOCH, para hacer esto, se utiliza una medida adimensional de variación relativa, llamada ***coeficiente de variación*** y se denota por CV.

Coeficiente de variación, CV, es una medida de dispersión relativa y adimensional de un conjunto de datos, que se obtiene dividiendo la desviación estándar entre la media, es decir

$$CV = \frac{S}{\bar{X}} * 100\%$$

Este valor nos ayuda a comparar la dispersión entre grupos de datos.

ACTIVIDAD DE APRENDIZAJE DESARROLLADA 05

Comparar los resultados del número de hijos por familia en la ciudad de Riobamba (X) y el tiempo de transacción bancaria (Y) que tienen las siguientes medias y desviaciones estándar:

$\bar{X} = 2.96 \cong 3$ hijos $\quad S_X = 1.79$ hijos
$\bar{Y} = 3.26$ minutos $\quad S_Y = 2.48$ minutos

Solución

Los coeficientes de variación respectivos son:

$$CV(X) = \frac{1.79}{2,96} = 0.6047 \quad \text{y} \quad CV(Y) = \frac{2,48}{3,26} = 0.7607$$

O, expresando en porcentajes, que es la forma más común, se tiene:

$$CV(X) = 60.47\ \% \quad \text{y} \quad CV(Y) = 76.07\ \%$$

Comparando estos valores, (aunque no tiene mucho sentido hacerlo) se dice que los tiempos de transacción tienen mayor variabilidad que el número de hijos por familia.

RECORDANDO

De la Actividad de Aprendizaje desarrollada 1. (Matriz de datos) se determina:

Estadísticas	*Matemáticas*	*Castellano*	*Ciencias Sociales*	*Inglés*	*Ciencias Naturales*
Promedio	15,87	16,85	16,47	15,74	16,26
Desviación estándar	3,72	2,53	1,53	3,09	2,68
CV en %	23,46%	15,02%	9,28%	19,62%	16,46%

De acuerdo al estadístico CV, la asignatura que presenta menor variabilidad por lo tanto es más homogéneo en el rendimiento académico de las 5 asignaturas es Ciencias Sociales.

ACTIVIDAD DE APRENDIZAJE DESARROLLADA 06

Se ha realizado la recopilación de las edades de 83 docentes de la ESPOCH.

29	31	30	28	26	32	33	27	59	31	30
44	24	28	28	29	31	37	28	29	27	29
24	32	26	32	26	46	33	28	28	27	31
27	29	27	28	33	26	31	47	36	28	
26	33	55	30	57	27	30	31	55	43	
25	29	30	35	28	39	28	34	27	40	
27	28	30	28	26	34	39	37	61	36	
35	28	30	31	24	33	31	45	34	30	

Con estos datos queremos:

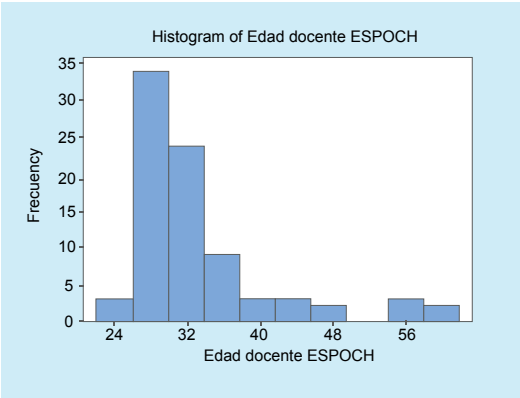
a) Ordenar los datos en una tabla de dos encabezados (Edad y frecuencias).

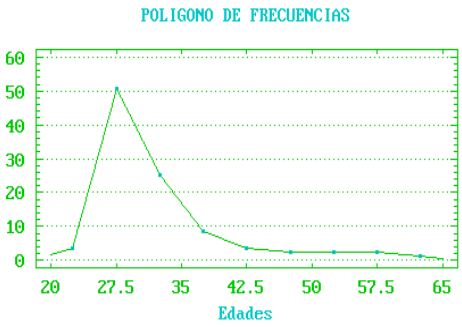
Edad	Frecuencia	Edad	Frecuencia	Edad	Frecuencia
24	2	38	0	52	0
25	1	39	2	53	0
26	6	40	1	54	0
27	9	41	0	55	2
28	13	42	0	56	0
29	6	43	1	57	1
30	8	44	1	58	0
31	8	45	1	59	1
32	3	46	1	60	0
33	5	47	1	61	1
34	3	48	0		
35	2	49	0		
36	2	50	0		
37	2	51	0		

b) Construya una tabla de frecuencias con los valores de clase y sus respectivas frecuencias absolutas, acumuladas y relativas. Utilice 9 clases:

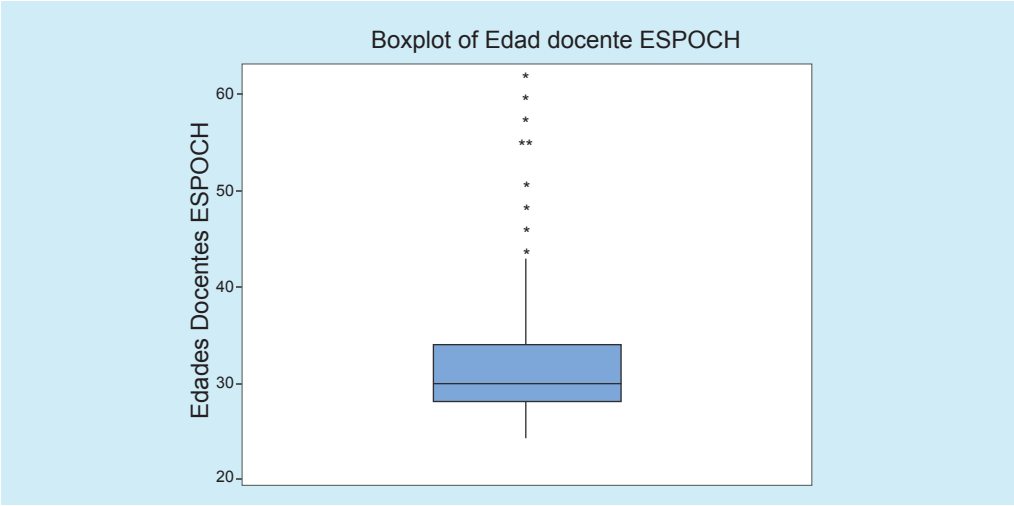
Clase i	Intervalo	c_i	n_i	f_i	F_i	F_i
1	(20-25]	22.5	3	0.036	3	0.0361
2	(20-30]	27.5	42	0.506	45	0.5422
3	(30-35]	32.5	21	0.253	66	0.7952
4	(35-40]	37.5	7	0.084	73	0.8795
5	(40-45]	42.5	3	0.036	76	0.9157
6	(45-50]	47.5	2	0.024	78	0.9398
7	(50-55]	52.5	2	0.241	80	0.9639
8	(55-60]	57.5	2	0.241	82	0.9880
9	(60-65]	62.5	1	0.012	83	1.0000
	Total		83	1.000		

Represente además gráficamente la distribución de frecuencias obtenida en el literal b) (Seleccione 2 de los cuatro gráficos tratados en el texto).





Se observe que el histograma de las edades de los docentes de la ESPOCH presenta asimetría a la derecha y el diagrama de caja y bigotes es:



c) Calcule las medidas de tendencia central

i	c_i	n_i	$c_i n_i$
1	22,5	3	67.5
2	27,5	42	1155.0
3	32,5	21	682.5
4	37,5	7	262..5
5	42,5	3	127.5
6	47,5	2	95.0
7	52,5	2	105.0
8	57,5	2	115.0
9	62,5	1	62.5
Total		83	2673.0

Luego, la ***media*** de los datos es: $\bar{x} = \frac{2673}{83} = 32.2$ a continuación se detalla el ***cálculo de la mediana***

i	C_i	n_i
1	22,5	3
2	27,5	42
3	32,5	21
4	37,5	7
5	42,5	3
6	47,5	2
7	52,5	2
8	57,5	2
9	62,5	1
Total		83

Clase de la mediana: (25-30]

Datos:
L = 25 n = 83; c = 5 $\sum f$ =3 fm =4

$$Mediana = Me = 25 + ((83/2) - 3)/42)*5 = 29.58$$

Cálculo de la moda: La clase modal coincide en este caso con la clase mediana, por lo tanto el punto medio de esta clase representa la moda y es 27,5.

Finalmente las medidas de tendencia central son:

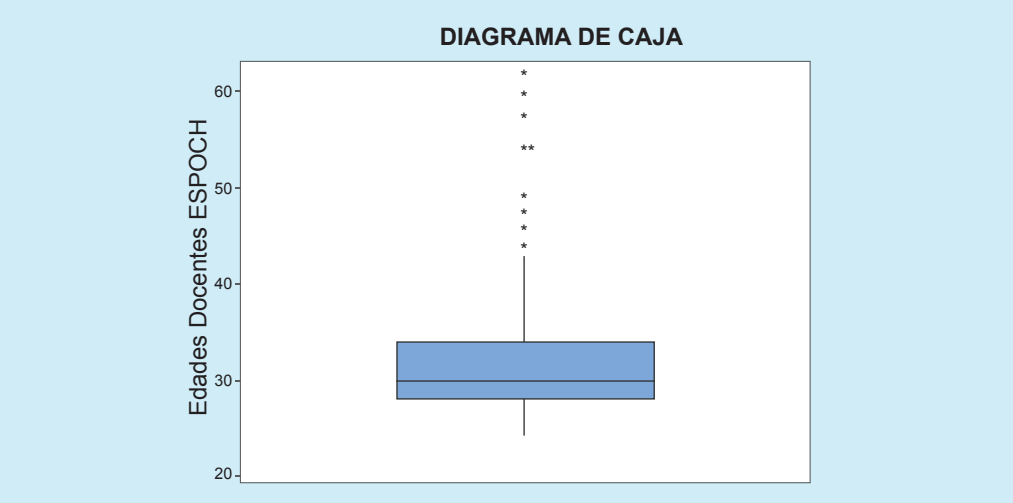
Media = 32.2
Mediana = 29.58
Moda = 27.5

d) Determine las medidas de dispersión:

Varianza = 64.34
Desviación estándar = 8.04

e) Calcule el coeficiente de variación y realice un diagrama de caja

$$CV = \left(\frac{8.0}{32.2}\right)*100\% = 24.6\ \%$$



De los resultados obtenidos emita su interpretación.

Tanto el histograma como el polígono de frecuencias de los datos de las edades de los docentes politécnicos demuestran una tendencia hacia la derecha, la cola de la curva está a la derecha efectivamente se comprueba al ver que se cumple la condición:

Moda < Mediana < $\bar{X}$.

Al ser CV < 33% (su valor es de 24.6%) se puede decir que las edades de los docentes politécnicos son homogéneas. La edad promedio de los docentes politécnicos es 32.2 años, es bastante joven, el 50.6% fluctúa entre 25 y 30 años, una población de recién graduados.

ACTIVIDAD DE APRENDIZAJE DESARROLLADA 07

Las siguientes actividades de aprendizaje desarrolladas tienen el propósito de poner en práctica los conocimientos definidos en los capítulos vistos y que el estudiante adquiera destrezas en el manejo de técnicas de la Estadística descriptiva.

1. Determine si cada una de las siguientes variables es cualitativa o cuantitativa. Si es cuantitativa determine si es discreta o continua:

a) Número de teléfonos por casa. ***CUANTITATIVA DISCRETA***
b) Tipo de teléfonos usados principalmente. ***CUALITATIVA***
c) Número de llamadas, de larga distancia, hechas. ***CUANTITATIVA DISCRETA***
d) Duración (en minutos) de la llamada de larga distancia más larga por mes. ***CUANTITATIVA CONTINUA***
e) Color de teléfono usado principalmente. ***CUALITATIVA***
f) Costo mensual (en dólares y centavos) de las llamadas de larga distancia. ***CUANTITATIVA CONTINUA***
g) Propiedad de un teléfono celular. ***CUALITATIVA***
h) Número de llamadas locales hechas. ***CUANTITATIVA DISCRETA***
i) Si existe una línea telefónica conectada a un MODEM de computadora en la casa. ***CUALITATIVA***
j) Determinar si existe una máquina de Fax en la casa. ***CUALITATIVA***

2. Los datos sin procesar mostrados a continuación son los cobros por elec tricidad durante el mes de Julio del 2014, para una muestra aleatoria de 50 apartamentos de tres recamaras en Quito.

96	171	202	178	147	102	153	197	127	82
157	185	90	116	172	111	148	213	130	165
141	149	206	175	1232	128	144	168	109	167
95	163	150	154	130	143	187	166	139	149
108	119	183	151	114	135	191	127	129	158

a) Forme una distribución de frecuencias con 7 intervalos de clase, con los siguientes límites de clase: $80 pero menos de $100; $100 pero menos de $120, etc.

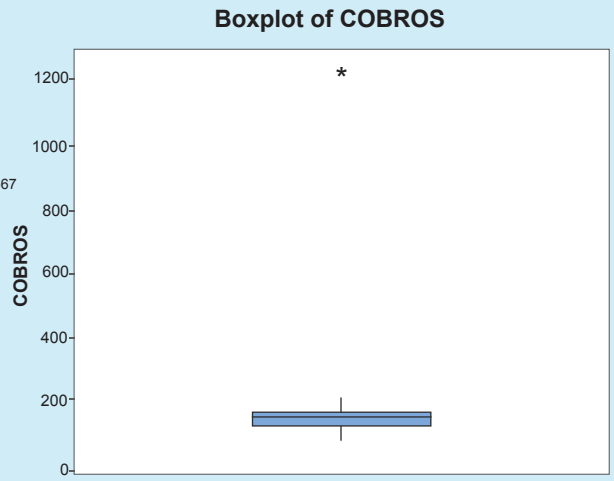
Clase	Límite Inferior	Límite Superior	Frecuencia	Frecuencia relativa
1	80	100	4	0.08
2	100	120	7	0.14
3	120	140	8	0.16
4	140	160	13	0.26
5	160	180	9	0.18
6	180	200	5	0.10
7	200	220	3	0.06
Suma	49	0.98		

Observación.- El valor 1232 es atípico y corresponde al 2% de los datos y se lo separa para el manejo estadístico de datos. En los diagramas de tallo y hojas así como de caja y bigotes se observa este aspecto en inglés steam and leaf y boxplot respectivamente.

Stem-and-Leaf Display: COBROS

Stem-and-leaf of COBROS N = 50
Leaf Unit = 10

```
4       0     8999
(42)    1     00011112222333344444445555556666667
4       2     001
1       3
1       4
1       5
1       6
1       7
1       8
1       9
1       10
1       11
1       12    3
```



b) Construya una distribución completa con 6 intervalos de clase (puntos medios, frecuencias absolutas, relativas parciales y acumuladas).

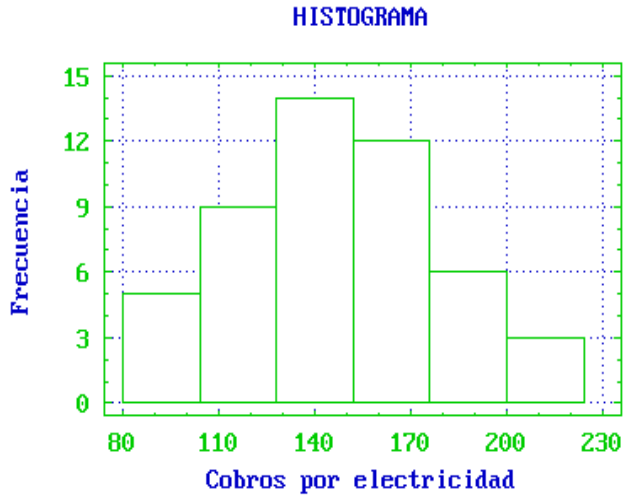
Clase	Límite Inferior	Límite Superior	Punto medio	Frecuencia Absoluta	Frecuencia Relativa	Frecuencia Acumulada
1	80	104	92	5	0.10	0.10
2	104	128	116	9	0.18	0.28
3	128	152	140	14	0.28	0.56
4	152	176	164	12	0.24	0.80
5	176	200	188	6	0.12	0.92
6	200	224	212	3	0.06	0.98
Suma				49	0.98	

Observación.- El valor 1232 es atípico y corresponde al 2% de los datos el cual no se le ha tomado en cuenta para el análisis estadístico.

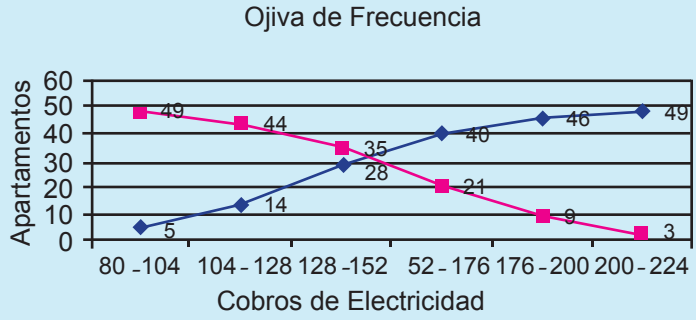
c) Construya un diagrama de tallo y hoja, excluyendo el dato atípico.

Tallo	Hoja	Frecuencia
8	2	1
9	056	3
10	289	3
11	1469	4
12	7789	4
13	0059	4
14	1347899	7
15	013478	6
16	35678	5
17	1258	4
18	357	3
19	17	2
20	26	2
21	3	1

d) Construya un histograma.



e) Construya las curvas de frecuencias acumuladas “menor que” y “mayor que”, en un mismo gráfico.



La curva de los cuadrados representa la de las frecuencias acumuladas “menores que” y la curva de los rombos representa la de las frecuencias “mayores que”, en el mismo gráfico elaborado en EXCEL.

3) Dada las series de datos basadas en el precio de cierre de acciones de muestras aleatorias de 25 artículos negociados en la Bolsa Norteamérica y 50 artículos negociados en la Bolsa de Nueva York:

Bolsa Norteamericana

6.88	15.88	5.38	33.62	14.25
0.75	16.50	14.38	4.88	4.00
3.88	8.75	2.50	9.00	15.25
4.12	9.25	4.88	2.00	2.38
11.88	7.50	6.38	20.00	49.50

Bolsa de Nueva York

36.50	3.75	9.12	5.75	24.00	26.00	12.88	35.25	3.75	25.00
23.50	25.00	33.38	21.88	10.88	19.00	5.50	20.62	64.75	35.00
8.25	15.50	22.50	6.12	18.75	46.00	37.50	24.00	14.25	9.00
57.50	36.12	8.75	25.00	53.88	23.50	8.88	80.50	46.38	12.38
27.12	6.00	8.62	15.88	20.38	22.62	59.12	29.38	4.75	31.00

a) Usando anchos de intervalo de $10, forme una distribución de frecuencias, absolutas y relativas, para cada serie.

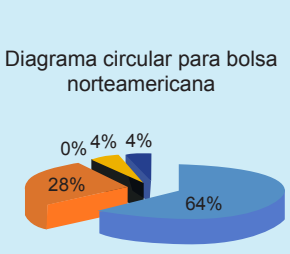
Clase	Límite Inferior	Límite Superior	Punto Medio	Frecuencia absoluta	Frecuencia relativa
1	0	10	5	16	0.64
2	10	20	15	7	0.28
3	20	30	25	0	0.00
4	30	40	35	1	0.04
5	40	50	45	1	0.04
		Suma		25	1.00

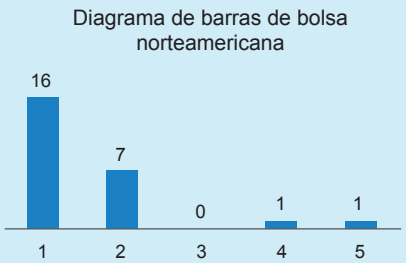
Tabla de frecuencias de la Bolsa Norteamericana

Clase	Límite Inferior	Límite Superior	Punto Medio	Frecuencia Absoluta	Frecuencia relativa
1	0	10	5	13	0.26
2	10	20	15	8	0.16
3	20	30	25	15	0.30
4	30	40	35	7	0.14
5	40	50	45	2	0.04
6	50	60	55	3	0.06
7	60	70	65	1	0.02
8	70	80	75	0	0.00
9	80	90	85	1	0.02
Suma				50	1.00

Tabla de frecuencias de la Bolsa de Nueva York

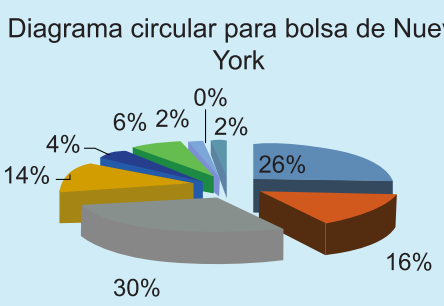
b) Para cada serie de datos, construya un diagrama circular y un diagrama de barras para el precio de cierre. Comente los resultados.

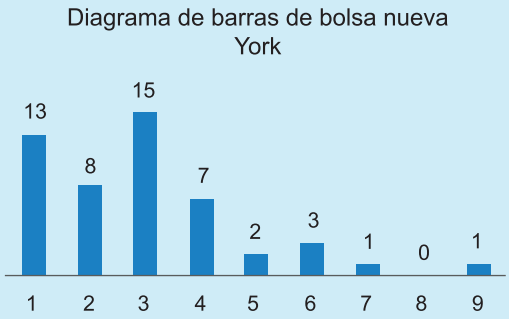




Comentario 1. Se observa en ambos diagramas que la primera clase (0-10] tiene un número grande de valores de la Bolsa Americana y la tercera clase no tiene valores, es decir no existe datos, frecuencia absoluta es 0.

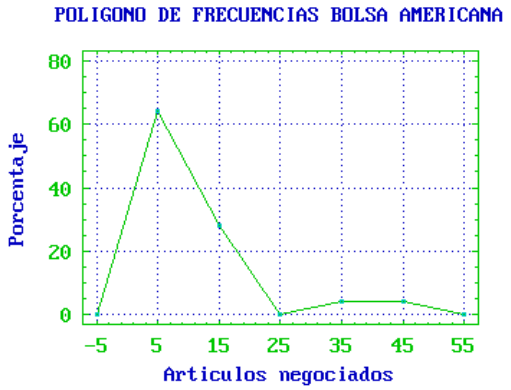
Comentario 2. Ahora se observa que la tercera clase de la Bolsa de Nueva York tiene la mayor cantidad de valores y se aprecia que la clase octava, es decir, la clase (70-80] no tiene valores o que su frecuencia absoluta es cero.

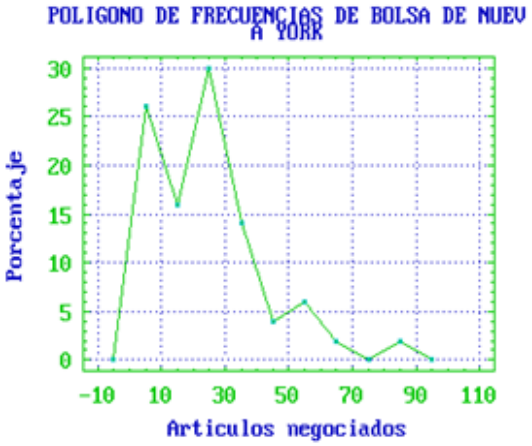




c) Construya un polígono de frecuencias para cada serie

Los polígonos de frecuencia se los desarrolla en el software estadístico STATGRAPHIC





Comentario. Se observan que los dos gráficos presentan asimetría positiva, es decir, la cola más grande está a la derecha.

4. Conteste correctamente a las siguientes preguntas, sea claro, conciso y explícito.

¿Qué medida de tendencia central se define como el valor del elemento que aparece con más frecuencia?

- *La moda es la medida que aparece con más frecuencia.*

¿Cuáles son las dos medidas de tendencia central que se ven afectadas por valores extremos?

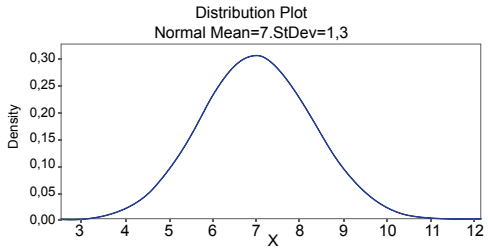
- *La media y la moda*

Investigación. ¿Qué medida debe utilizarse para determinar el incremento porcentual anual promedio?

- *La media geométrica.*

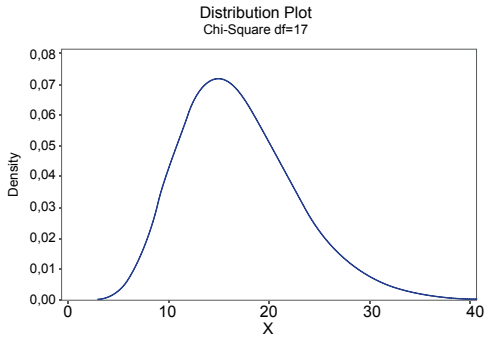
¿Cómo se describe la forma de una distribución de frecuencias si las tres medidas de tendencia central principales son iguales?

- *Se describe en forma simétrica y además se parece a una campana como la siguiente y se llama campana de Gauss.*



¿Cómo se describe la forma de una distribución de frecuencias si la media es mayor que las otras medidas?

- *Se describe con sesgo asimetría positivo, es decir de la forma:*



¿En una distribución de frecuencias con sesgo asimetría negativo, ¿qué medida de tendencia central es menor?

- *La moda por encontrarse más a la izquierda.*

¿Cuál es la ecuación para calcular la media aritmética de datos no agrupados?

$$Media\ de\ una\ muestra \Rightarrow \bar{x} = \frac{\sum x}{n}$$

¿Cuál es la ecuación para calcular la media aritmética cuando los datos se han agrupado en una distribución de frecuencias?

- *La ecuación para datos agrupados en una distribución de frecuencias es* $\bar{x} = \frac{\sum fx}{n}$ *donde f representa las frecuencias absolutas para cada valor central X de clase i.*

¿Cuál es la medida de tendencia central que no debe utilizarse cuando se tiene una distribución sesgada?

- *La media.*

ACTIVIDAD DE APRENDIZAJE DESARROLLADA 08

Dada la siguiente muestra de 24 observaciones diarias del número de kilómetros redondeadas a la décima más próxima, que recorrió Rolando Viera en su trabajo como agente vendedor:

100.3	122.7	93.4	112.0	129.7	101.3
117.7	98.9	127.3	119.1	120.1	97.3
121.9	130.7	115.3	105.5	99.4	109.1
101.1	125.7	122.3	98.1	97.2	95.3

Construya una tabla de frecuencias con seis clases y calcule las medidas de tendencia central y de dispersión:

Solución:

Clase i	Límites	Punto medio X	Frecuencia absoluta f	Frecuencia relativa	Frecuencia acumulada	Frecuencia Relativa acumulada
1	(92 - 99]	95.5	6	0.250	6	0.250
2	(99 - 106]	102.5	5	0.208	11	0.458
3	(106 - 113]	109.5	2	0.082	13	0.542
4	(113 - 120]	116.5	3	0.125	16	0.667
5	(120 - 127]	123.5	5	0.208	21	0.875
6	(127 - 134]	130.5	3	0.125	24	1.000
	Total		24	1.000		

X	f	Xf
95.5	6	573
102.5	5	512
109.5	2	219
116.5	3	349.5
123.5	5	617.5
130.5	3	391.5
Total	24	2663

De las tablas anteriores tenemos:

Medidas de Tendencia Central

La media es:

$$\bar{X} = \frac{\sum fX}{n} = \frac{2663}{24} = 110.94$$

La mediana es:

$$Me = L + \frac{\frac{n}{2} - FA}{f} i = 106 + \frac{\frac{24}{2} - 11}{2} 7 = 109.5$$

La moda es el punto de la clase modal, es decir, Moda = 95,5

Medidas de Dispersión

La Desviación estándar es:

$$s=\sqrt{\frac{\sum fX^2-\frac{(\sum fX)^2}{n}}{n-1}}=\sqrt{\frac{299302-\frac{(2663)^2}{24}}{24-1}}=12.89$$

El coeficiente de variación

$$CV=\frac{S}{\bar{X}}(100\%)=\frac{12.89}{110.96}(100\%)=11.62\%$$

El coeficiente de asimetría (Investigación).

$$CA=\frac{3(media\ -\ mediana)}{Desviación\ -\ estandar}=\frac{3(110.96\text{-}109.5)}{12.89}=0.34$$

La distribución tiene sesgo positivo.

Investigación.

El Segundo y tercer cuartil

$$L_{50}=(24+1)\frac{50}{100}=12.5$$ Que corresponde al valor 110.55

$$L_{75}=(24+1)\frac{75}{100}=18.75$$ Que corresponde al valor 122.1

El cuarto y octavo decil

$$L_{50}=(24+1)\frac{40}{100}=10$$ Que corresponde al valor 101.3

$$L_{80}=(24+1)\frac{80}{100}=20$$ Que corresponde al valor 122.7

El percentil 25° y el 70°

$$L_{25}=(24+1)\frac{25}{100}=6.25$$ Que corresponde al valor 99.15

$$L_{70}=(24+1)\frac{70}{100}=17.5$$ Que corresponde al valor 120.1

Investigación.

¿Entre qué valores espera usted que se encuentre el 60% de los valores? (Emplee el teorema de Chebyshev).

- Tomando en cuenta el resultado de este teorema calculamos:

$1 - 1/k^2 = 0.60$, es decir $K=\sqrt{\frac{1}{0.40}}=1.58$

Luego los valores se calculan por: $\bar{X}\pm(1{,}58)S$ esto es, 110.89 - (1.58)(12.40) = 91.30 y 110.89 + (1.58)(12.40) = 130.48 con media y desviación estándar de datos no agrupados 110.89 y 12.40 respectivamente.

ACTIVIDAD DE APRENDIZAJE DESARROLLADA 09

El número de Junio de 2003 de la revista PREVENTION medía la disminución de los niveles de estrés de las personas que utilizan un tapiz rodante durante 30 minutos al día y cuatro veces a la semana como mínimo. Las disminuciones se indican aquí en una tabla de frecuencias:

Disminución del nivel de tensión	FRECUENCIA (Número de Ejercicios)
10 y menos de 15	3
15 y menos de 20	4
20 y menos de 25	7
25 y menos de 30	10
30 y menos de 35	15
35 y menos de 40	8
40 y menos de 45	5

¿Puede una persona media esperar que disminuya su nivel de estrés en 25 puntos?

Solución:

Determinamos los puntos medios da cada clase para calcular la media

X	f	Xf	X^2	X^2f
12,5	3	37,5	156,25	468,75
17,5	4	70	306,25	1225
22,5	7	157,5	506,25	3543,75
27,5	10	275	756,25	7562,75
32,5	15	487,5	1056,25	15843,75
37,5	8	300	1406,25	11250
42,5	5	212,5	1806,25	9031,25
Total	52	1540	5993,75	48925

Luego la media es 1540/52 = 29,62. Con este valor no se puede esperar que disminuya el nivel de estrés en 25 puntos. La quinta clase es la clase de la mediana, entonces,

$$mediana = 30 + \frac{26-24}{15}5 = 30,67$$

éste valor tampoco nos indica que el estrés ha disminuido en 25 puntos.

- ¿Qué variación de la disminución de estrés podría existir de una persona a la siguiente?

$$S = \sqrt{\frac{\sum fX^2 - \frac{(\sum fX)^2}{n}}{n-1}} = \sqrt{\frac{48925 - \frac{(1540)^2}{52}}{51}} = 8,065$$

Por tanto la variación de la disminución del nivel de tensión es de 8.07 puntos. Con estos valores se puede decir ¿que la disminución del nivel de tensión es homogéneo? Para responder esta pregunta calculamos el coeficiente de variación CV.

- Determine el coeficiente de variación (CV)

$$CV = (\frac{S}{\bar{X}})*100\% = (\frac{8,06}{29,62})*100\% = 27,23\%$$

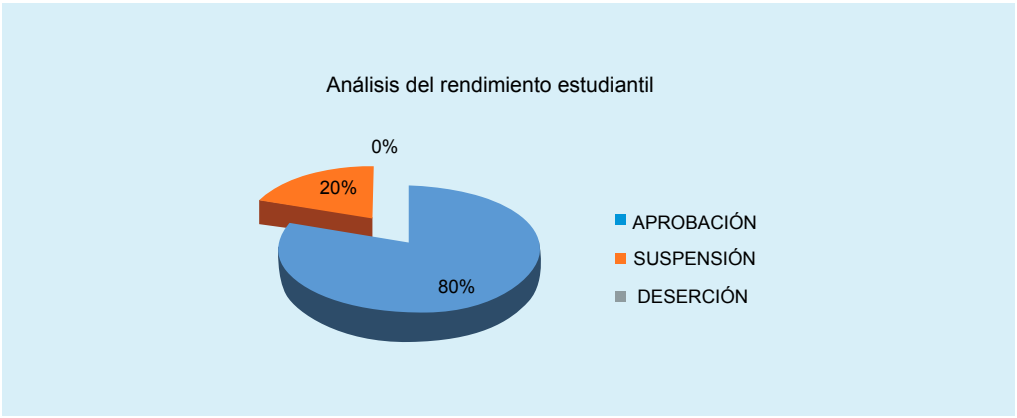
Puesto que, CV < 33% entonces los datos del nivel de tensión es homogéneo.

ACTIVIDAD DE APRENDIZAJE DESARROLLADA 10

Análisis y comentarios sobre el rendimiento estudiantil de los estudiantes del cuarto nivel de la carrera de Ingeniería en Estadística Informática de la Escuela de Física y Matemática de la ESPOCH periodo lectivo octubre 2014 - febrero 2015 (Estadística estudiantil: aprobación, repitencia, deserción), se detalla a continuación la lista de estudiantes matriculados en el periodo de referencia en la asignatura Estadística Inferencial al termino de las tres pruebas parciales: primera prueba evaluada sobre 8, segunda y tercera pruebas evaluadas sobre 10, obtenemos los siguientes datos:

N°	Código	Ev 1.	Ev 2.	Ev 3.	Total/28	% Asistencia	Observaciones
1	403	5	7	9	21	85	
2	462	5	5	4	14	80	
3	474	4	6	5	15	85	
4	475	7	9	9	25	98	EXONERADA
5	478	7	9	9	25	98	EXONERADO
6	481	8	8	10	26	100	EXONERADA
7	488	7	9	9	25	100	EXONERADA
8	516	5	7	7	19	90	
9	517	6	9	9	24	90	
10	518	7	9	9	25	100	EXONERADO

Solución

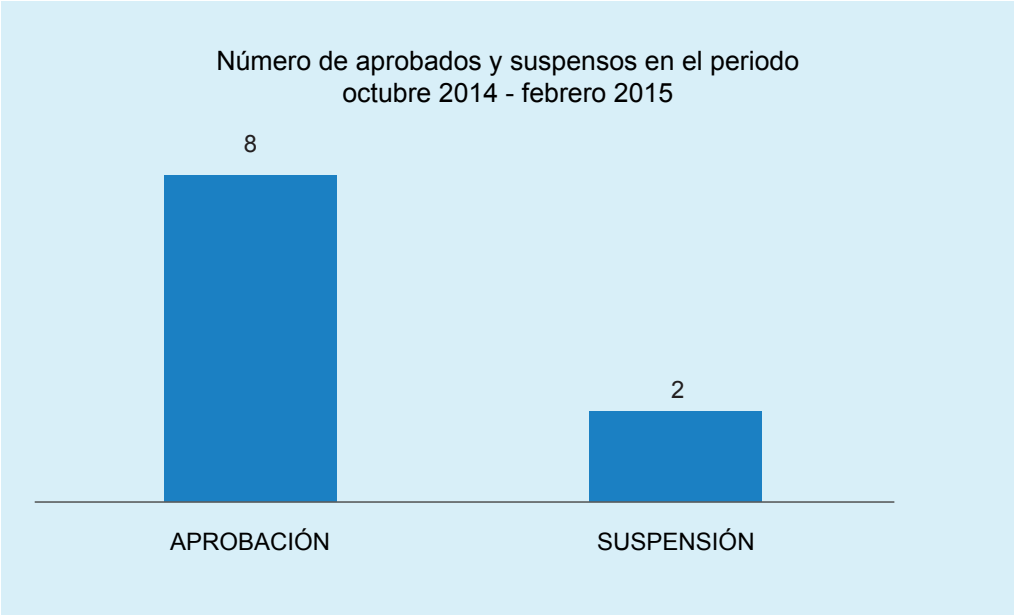


Estadísticas descriptivas del rendimiento académico y asistencia sin estudiantes desertados.

	Parcial 1	*Parcial 2*	*Parcial 3*	*Total/28*	*% Asist.*
Media	6,1	7,8	8	21,9	92,6
Error típico	0,4	0,5	0,6	1,4	2,4
Mediana	6,5	8,5	9	24,5	94
Promedio en %	76%	78%	80%	78%	93%
Desviación estándar	1,3	1,5	2	4,5	7,5
Varianza de la muestra	1,7	2,2	4	19,9	56,7
Curtosis	-1,2	-0,5	0,6	-0,5	-1,4
Coeficiente de asimetría	-0,2	-0,9	-1,4	-1,0	-0,4
Rango	4	4	6	12	20
Mínimo	4	5	4	14	80
Máximo	8	9	10	26	100
Suma	61	78	80	219	926
CV	21%	19%	25%	20%	8%

Estadística estudiantil del total de matriculados

Estadísticas del total de matriculados		
	NO.	%
APROBACIÓN	8	80%
SUSPENSIÓN	2	20%
DESERCIÓN	0	0%
TOTAL	10	100%



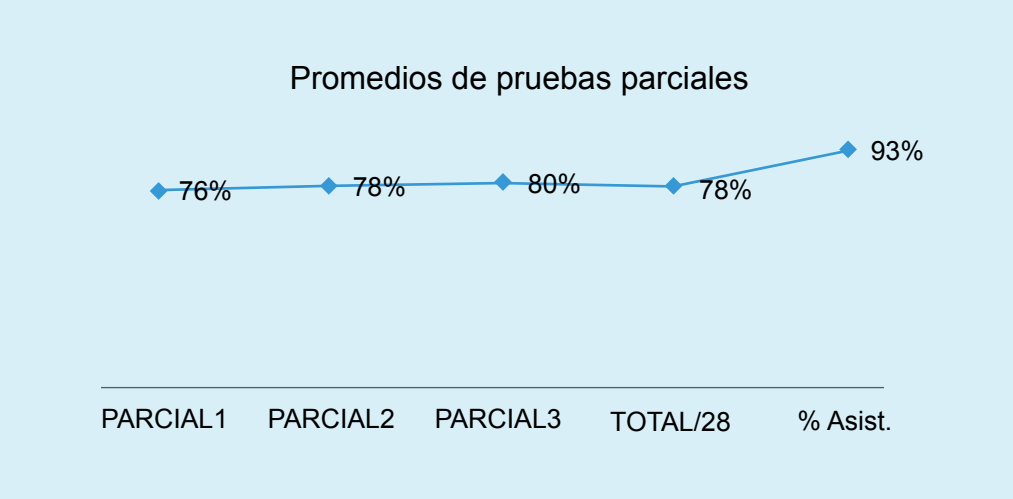
Estadística estudiantil luego de dar el examen de suspensión

Estadística estudiantil después de dar prueba de suspensión		
	NO.	%
APROBACIÓN	9	90%
REPITENCIA	1	10%
TOTAL	10	0%

Resumen estadístico

De los 10 estudiantes matriculados en la asignatura de Estadistica Inferencial de la carrera de: INGENIERIA EN ESTADISTICA INFORMATICA de la Facultad de Ciencias para el periodo Octubre 2014 – Febrero 2015 sobre ellos realizamos el análisis estadistico donde observamos que la asistencia del curso tiene un promedio de 93%, que en aprovechamiento superan el 76%, especificamente sobre el total de 28, tenemos 78%, consecuentemente este resultado se refleja al terminar el curso.

	EV.1	EV.2	EV.3
Media o promedio	6.1	7.8	8
% DEL PROMEDIO	76,0%	78%	80%



Cabe indicar que de los dos suspensos: un estudiante aprueba y el otro reprueba, es decir, en general en el cuarto nivel de la asignatura de Estadística Inferencial aprueban el 90% y repiten el 10%

ACTIVIDAD DE APRENDIZAJE DESARROLLADA **11**

La educación no solamente es instruir, sino también educar; de ahí que consideramos la siguiente aplicación estadística y con los resultados de las encuestas realizadas a los estudiantes de la ESPOCH, si se puede o no se puede crear un gimnasio particular para la preparación física en la ESPOCH en el periodo lectivo Marzo-Julio 2011.

2.3 *APLICACIÓN DE LA INVESTIGACIÓN ESTADÍSTICA*

¿Se puede crear un gimnasio particular para la preparación física en la ESPOCH?[1]

RESUMEN

De la encuesta aplicada a 200 estudiantes en las diferentes facultades de la ESPOCH para conocer aspectos inherentes a la implementación de un gimnasio específicamente en las áreas: tonificación, instrucción personalizada, asesoría nutricional, aumento y reducción de peso, tabeo, aeróbicos, físico culturismo, que se quiere implementar en el gimnasio de la ESPOCH se realizó primeramente un muestreo estratificado y luego se procedió a realizar el estudio estadístico para representar, analizar e interpretar los datos obtenidos en la encuesta y aplicando un itinerario básico para un proyecto de investigación estadística consiguiendo resultados que ayudarán a tomar buenas decisiones a nuestras autoridades.

[1] Trabajo elaborado por Jorge Congacha A., Nelson Rea, Carlos Miranda, Jaime Gualli, Franklin Cazorla, Laura Rochina.

SUMMARY

A survey was done of 200 students in different faculties of ESPOCH to determine various inherent aspects of the necessity of an enlarged gymnasium. This facility would serve in areas of toning, personalized instruction, nutritional orientation personalized instruction, nutritional orientation, weight loss and weight gain, tabeo, aerobics, body building that we would like to implement in the ESPOCH gymnasium. First a group sample was done, then a statistical study to represent, analyze and interpret the information gathered in the survey and to make a basic schedule for the project of statistical investigation to find results that will help our authorities to make better decisions.

INTRODUCCION

LA ESPOCH es una institución de educación superior que desde 1972, en su campus ecológico amplio, viene formando profesionales éticos y competitivos en diferentes áreas técnicas, que ayudan al desarrollo científico, social, investigativo de la provincia de Chimborazo y del País, tomando en cuenta la cultura y el deporte. En lo que respecta al deporte es necesario implementar más áreas en un gimnasio para la tonificación, instrucción personalizada, asesoría nutricional, aumento y reducción de peso, tabeo, aeróbicos, físico culturismo, en vista de que el número de estudiantes va creciendo cada semestre. Por ello la necesidad de la implementación de un gimnasio particular para el fisicoculturismo teniendo un horario acorde a las necesidades de los estudiantes de cada facultad, ya que la gran mayoría de estudiantes tiene diferentes horarios de estudio.

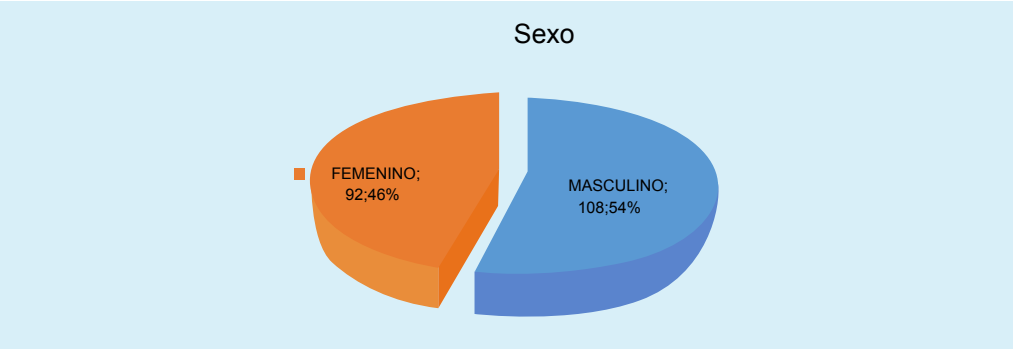
METODOS Y MATERIALES

La recopilación de la información requerida la planteamos en la encuesta que se indica a continuación. El software estadístico: MINITAB, SPSS y la Hoja electrónica EXCEL nos ayudaron a la representación y análisis de la información y para conseguir los resultados esperados correctamente aplicamos la metodología de la investigación estadística siguiendo los pasos:

- Planeación de la investigación
- Elaboración de los instrumentos de análisis
- Prueba piloto
- Selección de la muestra piloto
- Elaboración definitiva de los instrumentos de análisis.
- Selección y entrenamiento de los encuestadores.
- Recolección de datos
- Análisis estadístico
- Informe de la investigación

RESULTADOS Y DISCUSIÓN

CUESTIONES INFORMATIVAS



Del total de encuestados 54% son de sexo masculino: 108 y el 46% son de sexo femenino: 92

Edad (años)	
Media	21
Moda	20
Mínimo	17
Máximo	28
Cuenta	200

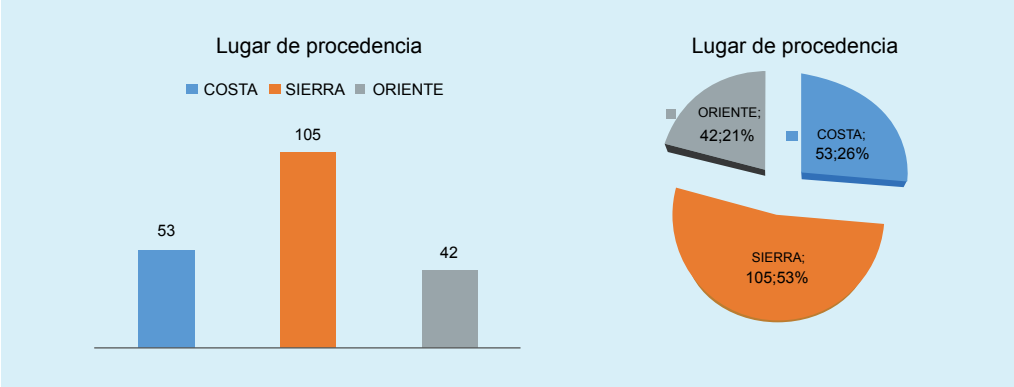
Aquí podemos observar que la edad promedio de los encuestados es de 21 años, de los cuales tenemos más encuestados de 20 años. De los mismos tenemos un encuestado de 17 años que es el menor de todos y un individuo de 28 años que es el mayor de un total de 200 encuestados en las diferentes facultades de la ESPOCH.

Estatura (metros)	
Media	1,6282
Mediana	1,63
Moda	1,6
Mínimo	1,39
Máximo	1,85
Cuenta	200

La estatura promedio de los encuestados es de 1,62 metros siendo la más repetitiva de 1,60 metros. La estatura más pequeña es de 1,39 metros y la máxima estatura es de 1,85 metros.

Peso (libras)	
Media	128,98
Mediana	121
Moda	120
Desviación estándar	21,88
Mínimo	86
Máximo	209,48
Cuenta	200

El peso promedio de los encuestados es de 128,98 libras; el peso de más frecuencia en los encuestados es de 120 libras con un valor mínimo en peso de 86 libras y un máximo de 209.47 libras



Del total de encuestados tenemos que de la costa son 53 individuos y corresponde al 26%, de la sierra son 105 individuos y corresponde al 53% y del oriente ecuatoriano son 42 individuos y corresponde al 21%. Como se ve en el diagrama circular.

CUESTIONES REQUERIDAS

El 82% de encuestados dieron una respuesta afirmativa a la preparación física dentro de la ESPOCH frente a un 18% que no le interesa la propuesta.

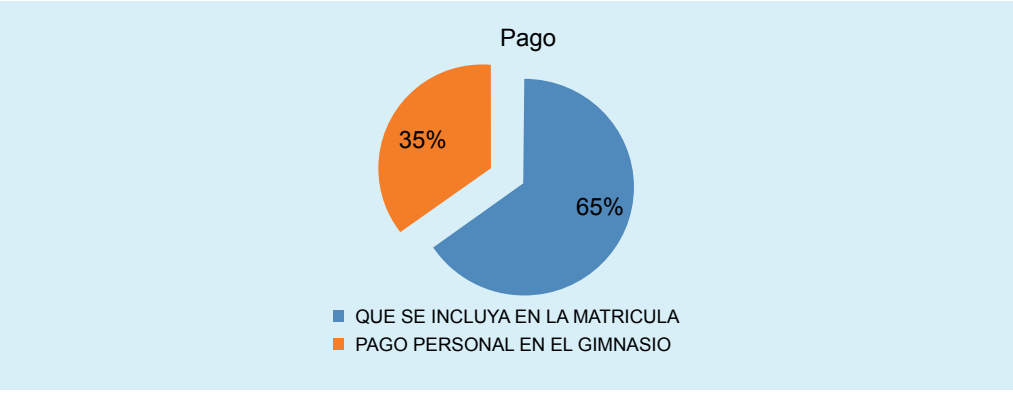
¿En qué áreas le gustaría prepararse físicamente?

Áreas	Individuos
tonificación	78
instrucción personalizada	41
asesoría nutricional	66
aumento y reducción de peso	64
tabeo	24
aeróbicos	49
físico culturismo	38
Total	360*

Cabe anotar que los encuestados respondieron al menos por una opción de la encuesta que al final presentamos

Se pude observar que los estudiantes desean prepararse físicamente con un alto porcentaje, 22%, en el área de tonificación muscular, seguida de aumento y reducción de peso, 18 %, empatado con asesoría nutricional, y con porcentajes menores los estudiantes desean prepararse en las áreas de aeróbicos, 14%, instrucción personalizada, 11%, físico culturismo, 10%, y tabeo, 7%,

¿Cómo desea que se realice el pago por adquirir este servicio?



Con los datos obtenidos en la encuesta pudimos observar que los estudiantes quieren que el pago se realice en la matrícula, 65%. Se realizó también diagramas de barras como el circular para observar que los estudiantes de la ESPOCH en este periodo escogieron un horario de preferencia de los días sábados de 8.00-12:00, 40%, seguido de un porcentaje menor, 26% desean un horario de lunes a viernes de 14:00-18:00.

CONCLUSIONES

- La mayoría de los estudiantes encuestados están de acuerdo con implementación de un gimnasio en la ESPOCH.
- Los estudiantes están de acuerdo con el horario establecido para los días sábados
- La mayoría de los estudiantes desean prepararse en el área de tonifi cación muscular.
- Con respecto al pago los estudiantes en su mayoría desean que se incluya en la matricula.

RECOMENDACIONES

- El horario de atención debe ser de acuerdo a los horarios establecidos por los estudiantes.
- El gimnasio debe contar con suficientes máquinas para la tonificación muscular y en las demás aéreas presentadas.
- El cobro debe ser en el gimnasio personalmente ya que si se incluye en la matricula, el gimnasio deberá realizar un trámite muy extenso.
- Que la politécnica tenga un gimnasio con todas las necesidades del estudiante.
- Establecer si está de acuerdo con el costo propuesto del pago $1extra matrícula para la utilización del gimnasio en la ESPOCH.
- Realizar un estudio del análisis inferencial para estimar parámetros y comprobar hipótesis

REFERENCIAS BIBLIOGRÁFICAS

1. CONGACHA, J.; ORTEGA, M. 2001. Introducción a la Estadística y teoría de las probabilidades. ESPOCH. Riobamba. 142 p.
2. LEVIN, R. 1996. Estadística para Administradores. PRENTICE-HALL. HISPANOAMERICANA, S.A MEXICO. 1018 p.
3. LOPES. P. 2000. Probabilidad & Estadística Conceptos, Modelos, Aplicaciones en Excel. Pearson. Colombia. 298 p.
4. ZURITA, G. 2008. Probabilidad y Estadística Fundamentos y Aplicaciones. ESPOL. Guayaquil. 802 p.

ANEXOS

La encuesta que se aplicó para la recopilación de la información a los estudiantes de la ESPOCH, respecto a si se puede o no se puede crear un gimnasio particular para la preparación física en la ESPOCH en el periodo lectivo Marzo-Julio 2011 y que la representamos gráfica y numéricamente exponemos a continuación.

ESCUELA SUPERIOR POLITECNICA DE CHIMBORAZO

FACULTAD DE CIENCIAS
ESCUELA DE FISICA Y MATEMATICA
ENCUESTA A ESTUDIANTES ESPOCH

La presente encuesta tiene como objetivo conocer aspectos inherentes para la implementación de un gimnasio específicamente en las áreas: tonificación, instrucción personalizada, asesoría nutricional, aumento y reducción de peso, tabeo, aeróbicos, físico culturismo. Por lo tanto le solicitamos consigne sus respuestas con la honradez que le caracteriza y le distingue.

Reciba nuestro agradecimiento por tan significativa colaboración.

Cuestiones informativas:

1. **SEXO:** Masculino ☐ Femenino ☐

2. Edad:_____ Estatura:_____ Peso:_____

3. Lugar de procedencia_____

Cuestiones requeridas:

4. **¿ DESEARIA PREPARARSE FISICAMENTE ?**

Si ☐ No ☐

5. **¿ EN QUE AREAS LE GUSTARIA PREPARARSE FISICAMENTE ?**

Tonificación instrucción personalizada asesoría nutricional
Aumento y reducción de peso tabeo aeróbicos físico culturismo

6. **¿ EN QUÉ HORARIO LE GUSTARÍA PREPARARSE FÍSICAMENTE ?**

Lunes a viernes 8:00 – 12:00 ☐
Lunes a viernes 14:00 – 18:00 ☐
Lunes a viernes 20:00 – 22:00 ☐
Sábado 8:00 – 12:00 ☐

2.4 ACTIVIDADES DE APRENDIZAJE 2

1. Realice siguientes actividades de aprendizaje

a. Ilustre con ejemplos lo que se entiende por población, muestra, variables cualitativas y variables cuantitativas.

.....

b. ¿Por qué es útil la estadística en el campo para el cual se está preparando?

.....

c. "La estadística estudia el comportamiento de fenómenos colectivos y nunca de una observación individual". Comentar este principio.

.....

d. Según la fórmula de Sturgues, ¿cuántas clases o intervalos se obtendrían para una muestra que contiene: 50, 90, 1200 y 5000 observaciones?

.....

2. De la asignatura que usted imparte clases, recopile los promedios finales de un curso. Con estos datos construya:

a. Una tabla de frecuencias (las clases tome determinando la raíz cuadrada del número de datos)
b. Un histograma y sobre este indique la moda
c. Un polígono de frecuencias en porcentaje y una ojiva.
d. Interprete los incisos anteriores

3. Conteste el siguiente cuestionario. Ponga una X en la alternativa que crea correcta.

a. En una distribución asimétrica se tiene que la Media = 20; Mediana = 24. El valor de la moda deberá ser:

- *Mayor que la media y menor que la mediana*
- *Mayor que la mediana*
- *Menor que la mediana*
- *Menor que la media*

b. La moda generalmente se define como aquel valor de la variable que:

- *Se ve afectada por valores extremos*
- *Mas se repite*
- *Tiene la menor frecuencia*
- *Supera a la mitad de las observaciones*
- *Tiene mayor grado de variabilidad*

c. En una distribución simétrica la media Me, la mediana Md y moda Mo, debe suceder que:

- $M_o < M_d < M_e$
- $M_d = M_e = M_o$
- $M_d < M_e < M_o$
- $M_d > M_e > M_o$

d. Con los siguientes datos correspondientes a una tabla de frecuencias, de una variable discreta. Calcule:

La media aritmética: 6.825; 7.253; 7.54; 8.12; 10.16
La mediana es: 3; 5; 6; 9; 12
La moda es: 3; 5; 6; 9; 1

i	x_i	n_i
1	3	8
2	6	20
3	9	7
4	12	3
5	15	2
Total		40

4. ***Considere el siguiente par de muestras, calificaciones sobre10 de ocho estudiantes de dos quimestres:***

Muestra1: 10, 9, 8, 7, 8, 6, 10, 6
Muestra2: 6, 10, 10, 6, 8, 10, 8, 6

a. Calcule el rango de ambas muestras. Es posible concluir que las dos muestras exhiben la misma variabilidad?

b. Calcule la desviación estándar de cada una de las muestras. ¿Estas cantidades indican que las dos muestras tienen la misma variabilidad?

c. Calcule las medias muestrales y los coeficientes de variación de las dos muestras. Comente estos resultados.

5. ***Las siguientes puntuaciones representan la calificación del examen principal para un curso de Econometría evaluadas sobre 100:***

23 60 79 32 57 74 52 70 82
36 80 77 81 95 41 65 92 85
55 76 52 10 64 75 78 25 80
98 81 67 41 71 83 54 64 72
88 62 74 43 60 78 89 76 84
48 84 90 15 79 34 67 17 82
69 74 63 80 85 61

a. Elabore un diagrama de tallo y hojas para las calificaciones del examen, donde los tallos sean 1, 2, 3,…, 9.

b. Elabore un histograma de frecuencias relativas, trace un estimado de la gráfica de la distribución y analice la asimetría de la distribución.

c. Calcule la media, la mediana y la desviación estándar de la muestra.

d. Elabore un informe con los tres incisos anteriores, interpretando los resultados obtenidos.

3 CAPÍTULO TEORÍA DE LAS PROBABILIDADES

OBJETIVOS

- Explicar de diferentes maneras en que surge la probabilidad.
- Examinar el uso de la Teoría de Probabilidades en la toma de decisiones.
- Desarrollar habilidades y destrezas para el cálculo de diferentes tipos de probabilidad.
- Aplicar argumentos de la Teoría de las Probabilidades en problemas de la vida real.

CONTENIDOS

3.1 CONCEPTOS DE PROBABILIDAD

En este capítulo introducimos el vocabulario básico de la Teoría de las Probabilidades, los términos que se introduzcan constituirán el lenguaje común y el lector debe familiarizarse con ellos.

La noción básica en teoría de las probabilidades es la de experimento aleatorio, pero antes tengamos presente la siguiente definición de ***experimento***.

Definición.- Un ***experimento*** es el proceso por medio del cual se obtiene una observación.

Un ***experimento aleatorio*** es aquel cuyos resultados no pueden ser determinados. Algunas actividades de aprendizaje de este tipo de experimento aclaran lo dicho.

Actividades de aprendizaje. Definimos los siguientes experimentos aleatorios:

1. Lanzamiento de un dado.
2. Lanzamiento de una moneda.
3. En una fábrica en la que se desea detectar artículos defectuosos de un lote de 100 artículos.
4. El número de bachilleres de la última promoción (suponemos 210) que aprobarán los cursos pre-politécnicos de un determinado colegio.
5. Cierto día usted decide tomar lección oral a un alumno, eligiéndolo al azar.

Una forma de definir de modo preciso la esencia de un experimento aleatorio es describiendo el conjunto de todos los resultados posibles del experimento.

Este conjunto se llama ***espacio muestral.***

Un ***espacio muestral*** se denota por Ω o por S. En esta introducción a la probabilidad, se usarán los conceptos básicos de conjuntos y las operaciones entre conjuntos y tomaremos el símbolo Ω como espacio muestral de un experimento aleatorio. Se supone que el lector está familiarizado con estos conceptos.

Un evento es un subconjunto de un espacio muestral. Un evento se indica con letras mayúsculas del alfabeto.

Actividades de aprendizaje de espacios muestrales y eventos

Para el experimento 1: Ω = {1, 2, 3, 4, 5, 6}

Para el experimento 2: Ω = {C, S}; se observa la cara C o el sello S.

Para el experimento 3: Ω = {0, 1, 2,..., 100}

Para el experimento 4: Ω= {0, 1, 2,...,210}

Para el experimento 5: Ω= {estudiantes de su curso}

Al observar un número par en el experimento 1 se define el evento A = {2, 4, 6}.

En el experimento 2 tenemos los eventos B = {C} y D = {S}.
Describa los espacios muestrales de los siguientes experimentos

A. Lanzamiento de dos monedas:

$\Omega = \{CC, CS, SC, SS\}$

B. Lanzamiento de dos dados:

$\Omega = \{(i, j) / \quad i,j = 1,2,3,4,5,6\}$

C. Un estudiante lleva en su bolsillo cuatro monedas de 100$, 500$ y 1000$; si son todas iguales y saca dos sucesivamente y si m_1 = 100$, m_2 = 500$, m_3 = 1000$, entonces

$\Omega = \{m_1m_2, m_1m_3, m_2m_3\}$

Ahora láncese una moneda hasta que aparezca una cara y luego cuéntese el número de veces que se lanzó la moneda; ¿cual es el espacio muestral Ω?

El espacio muestral de este experimento es

$\Omega = \{1, 2, 3,...\}$.

Observación.- Una de las características básicas del concepto de experimento es que no sabemos qué resultado particular posible se obtendrá al realizar el experimento. En otras palabras, si A es un evento no podemos indicar con certeza que A ocurrirá o no.

Por lo tanto llega a ser muy importante tratar de asociar un número con el evento A, que medirá de alguna manera, la posibilidad de que A ocurra. Esta tarea nos conduce a la Teoría de las Probabilidades.

¿Cómo asignar un número a cada evento A que medirá la posibilidad de que A ocurra cuando el experimento se realiza? Al respecto, un enfoque podría ser el siguiente: repetir el experimento un gran número de veces, calcular la frecuencia relativa f_A correspondiente al evento A y usar este número como medida.

Aún más como sabemos que el experimento se repite más y más veces, la frecuencia relativa se estabiliza cerca de un número, digamos p, pero lo que queremos es un medio de obtener tal número sin recurrir a la experimentación.

Con esta observación damos los siguientes conceptos de probabilidad.

Nota. *Las actividades de aprendizaje propuestas y resueltas anteriormente de espacios muestrales ponen de manifiesto que el conjunto de todos los posibles resultados puede ser finito o infinito numerable o infinito continuo, como el dar en el blanco sobre un circulo de diferentes diámetros con un dardo.*

Respecto a esta descripción los espacios muestrales pueden ser discretos finitos o infinitos numerables o infinitos continuos.

3.1.1 CONCEPTO CLÁSICO (SEGÚN LAPLACÉ)

Sea Ω un espacio muestral de n elementos y el evento A con cardinalidad m, con 0 m n, los puntos muestrales de Ω tienen la misma probabilidad, es decir, son equiprobables e iguales a 1/n. La probabilidad del evento A se denota y calcula por:

$$P(A) = \frac{\#\ de\ eventos\ simples\ favorables}{\#\ de\ eventos\ simples\ posibles} = \frac{Card(A)}{Card(\Omega)} = \frac{m}{n}$$

Conclusión.- Al analizar el concepto de frecuencia de probabilidad ésta debe cumplir:

1) Para todo evento simple (un evento que no puede descomponerse e_i)

$$e_i \in \Omega \quad 0 \leq P(\{e_i\}) \leq 1$$

2) $\sum_{i=1}^{n} P(\{e_i\}) = 1$ luego $P(\Omega) = 1$

3) Si tenemos 2 eventos mutuamente excluyentes A y B (A ∩ B = Ø) no hay puntos muestrales comunes (elementos del espacio muestral) y sean las frecuencias f_A y f_B respectivamente de los eventos A y B entonces:

$$f_{A \cup B} = f_A + f_B$$

3.1.2 CONCEPTO AXIOMÁTICO DE PROBABILIDAD

Siendo nuestro objetivo el estudio de los eventos considerados como conjuntos damos a continuación las notaciones conjuntistas y el significado correspondiente en la siguiente tabla.

SÍMBOLO	SIGNIFICADO
Ω	*Evento seguro*.
A^C	*Evento contrario*. Evento que ocurre cuando no ocurre A o viceversa.
A∪B	*Unión de eventos* A y B. Evento que ocurre cuando ocurre uno al menos de los posibles resultados de A o de B.
A∩B (A-B=A∩CB)	*Conjunción de eventos* A y B. Evento que ocurre cuando ocurre simultáneamente los posibles resultados de A y B (A y CB)
A⊆ B	*A implica B*. Si ocurre A, necesariamente ocurre B
A∩B= Ø	*Eventos mutuamente excluyentes o incompatibles*.
GENERALIZACIONES	
$\bigcup_{i=1}^{n} A_i = B$	B es el evento que ocurre ssi al menos uno de los eventos A_i ocurren.
$\bigcap_{i=1}^{n} A_i = C$	C es el evento que ocurre ssi todos los eventos A_i ocurren.
$\bigcup_{n \in N} A_n = D$	D es el evento que ocurre ssi al menos uno de los eventos A_n ocurren.
$\bigcap_{n \in N} A_n = E$	E es el evento que ocurre ssi todos los eventos A_n ocurren.
$\bigcap_{i=1}^{n} A_i = \emptyset$	Los eventos $A_1, A_2, \ldots, A_n$ son mutuamente excluyentes.
$\bigcap_{n \in N} A_n = \emptyset$	La sucesión de eventos $A_1, A_2, \ldots, A_n, \ldots$ son mutuamente excluyentes.

Observación.- La tabla nos pone de manifiesto que la unión, intersección tanto finita como infinita numerable de eventos es todavía un evento. Al igual que el complemento de un evento es un evento o el mismo espacio muestral es un evento (todo conjunto es subconjunto de si mismo).

Definición.- Sea Ω un espacio muestral correspondiente a un experimento que se estudia y F la familia de eventos de Ω, introducimos una "medida generalizada" que atribuya a cada evento un número que llamaremos probabilidad y satisface los siguientes axiomas:

A1.- A cada evento de F se asigna un número real no negativo, o sea

$$\forall n \in N; A_n \in F \Rightarrow P(A_n) \geq 0.$$

A2.- P(Ω) =1. La probabilidad del evento seguro es 1

A3.- $\forall n \in N; A_n \in F / A_i \cap A_j = \varnothing, i \neq j \Rightarrow P\left(\bigcup_{n \in N} A_n\right) = \sum_{n \in N} P(A_n)$

Nota.

$A_n \in F \; o \; A_i \in F \; o \; A_j \in F$ *indica que* A_n *ó* A_i *ó* A_j *son eventos.*

Comentario de la definición:

A1 establece la unicidad de la probabilidad de un evento, A2 asigna al evento seguro el máximo valor posible y A3 es el principio de la probabilidad total respecto a eventos mutuamente excluyentes.

Definición.- Llamamos espacio de probabilidades a la terna (Ω, F, p).
Las siguientes propiedades nos dan como consecuencia inmediata de la definición anterior:

1. $P(\emptyset)= 0$ ($\emptyset = \Omega^c$; es el ***evento imposible***)

2. ***Aditividad finita***.

$$i;\ 1\le i\le n;\ A_i \in F,\ A_i \cap A_k = \emptyset,\ j\neq k\ \ 1\le j,\ k\le n \Rightarrow P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1} P(A_i)$$

Si n=2 y sean A y B eventos mutuamente excluyentes, entonces

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

3. $P(A^c) = 1 - P(A)$. ***Probabilidad del evento contrario de A***

4. Teorema de la **probabilidad total**

Si A y B son eventos cualesquiera, entonces

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

5. Propiedad de **monotonía de la probabilidad**

Si A y B son eventos cualesquiera tales que $A \subseteq B$, entonces $P(A) \le P(B)$

5.1. si $B = \Omega$, entonces $P(A) \le 1$; $A \in F$

Por tanto, para cualquier evento $A \in F$ por A1 y 5.1 se tiene:

$$0 \le P(A) \le 1$$

Entonces, siempre la probabilidad de un evento A es un número definido entre 0 y 1 incluidos ò diremos también en términos de porcentaje que $0\% \le P(A) \le 100\%$

6. **Independencia de eventos.** Dos eventos A y B son independientes si:

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

Nota. *La propiedad de independencia de eventos significa que el conocimiento de la realización del evento B no aporta en nada al conocimiento de la realización del evento A y viceversa.*

ACTIVIDAD DE APRENDIZAJE PROPUESTA

1. Supongamos el experimento aleatorio “se lanza una moneda tres veces”.

a) Definir el espacio muestral Ω y la probabilidad de los elementos muestrales
b) Escriba explícitamente los elementos muestrales de los siguientes eventos y calcule su probabilidad.
b1) Sale dos caras y un sello
b2) Sale al menos una cara
b3) Sale al menos un sello
b4) Sale al menos un sello y al menos una cara.

2. Cada uno de los cinco posibles resultados de un experimento aleatorio es igualmente probable. El espacio muestral Ω = {a, b, c, d, e}. Sean A: el evento {a, b}. y B: el evento {c, d, e}. Determine lo siguiente:

I. P(A)
II. P(B)
III. P(CA)
IV. P(A∪B)
V. P(A∩B)

3. Si P(A) = 0.3, P(B) = 0.2 Y P(A∩B) = 0.1, determine las siguientes probabilidades:

I. P(CA); CA = A^c
II. P(A∪B)
III. P(A^c∩B)
IV. P(A∩B).

4. Demuestre que, si A y B son eventos independientes de Ω entonces:

a) A y B^c son independientes
b) A^c y B son independientes
c) A^c y B^c son independientes

Observación.- Se observe que dos eventos mutuamente excluyentes son independientes si uno de ellos tiene probabilidad nula, es decir, dado que

$$A \cap B = \emptyset \text{ se tiene } P(A \cap B) = P(A)P(B) \text{ si } P(A) = 0 \text{ ó } P(B) = 0.$$

No se debe confundir independencia con exclusión mutua.

ACTIVIDAD DE APRENDIZAJE DESARROLLADA 01

Consideremos el lanzamiento de un dado y los eventos siguientes:
A: “Se obtiene un número par”
B: “Se obtiene un número impar”
El espacio muestral es Ω = {1, 2, 3, 4, 5, 6} y los eventos son A = {2, 4, 6}. B = {1, 3, 5}

¿ Son A y B eventos independientes ?

Solución:
En efecto A∩B = Ø, implica que los dos eventos son mutuamente excluyentes, pero A y B no son independientes, puesto que si conocemos que el evento A (B) se realizó no podemos esperar que el evento B (A) también se realice o sea,

$$P(A)P(B) = (½)*(½) = 1/4, \text{ en tanto que, } P(A \cap B) = P(\emptyset) = 0$$

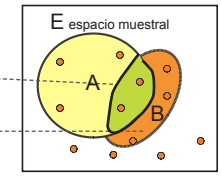
Luego, $P(A \cap B) \neq P(A)P(B)$, puesto que

$$0 \neq 1/4$$

Definición. Se llama probabilidad de A condicionada a B, o probabilidad de A sabiendo que pasa B, se denota por P(AIB) y se lee "la probabilidad del evento A dado el evento B".

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

"tamaño" de uno respecto al otro



Si tomamos como ejemplo repetir en 1000 ocasiones el experimento de elegir a una mujer de una población muy grande. El resultado está en la tabla.

Recuento

		MENOPAUSIA		Total
		NO	SI	
CLASIFICACIÓN OMS	NORMAL	189	280	469
	OSTEOPENIA	108	359	467
	OSTEOPOROSIS	6	58	64
Total		303	697	1000

¿Cuál es la probabilidad de que una mujer tenga osteoporosis?

- P(Osteoporosis)=64/1000=0,064=6,4%

¿Cuál es la probabilidad de que una mujer no tenga osteoporosis?

- P(No Osteoporosis)=1-P(Osteoporosis)=1-64/1000=0,936=93,6%

En ambas respuestas se aplica el concepto clásico de probabilidad o llamado también frecuentista.

¿Probabilidad de tener osteopenia u osteoporosis?

- P(OsteopeniaUOsteoporosis)=P(Osteopenia)+P(Osteoporosis)-P(Osteopenia∩Osteoporosis)=467/1000+64/1000=0,531 = 53,1%
  - Son eventos ó sucesos disjuntos
  - Osteopenia ∩ Osteoporosis = Ø

¿Probabilidad de tener osteoporosis o menopausia?

P(OsteoporosisUMenopausia)=P(Osteoporosis)+P(Menopausia) -P(Osteoporosis ∩ Menopausia)=64/1000+697/1000-58/1000=0,703 = 70,3%

- No son sucesos disjuntos

¿Probabilidad de una mujer normal?

P(Normal) = 469/1000 = 0,469 = 46,9%

P(Normal) = 1 - P(Normal') = 1-P(OsteopeniaUOsteoporosis) = 1-0,531 = 0,469 = 46,9%

Si es menopáusica… ¿probabilidad de osteoporosis?

- P(Osteoporosis|Menopausia)=58/697=0,098 = 9.8%

¿Probabilidad de menopausia y osteoporosis?

- P(Menopausia ∩ Osteoporosis) = 58/1000=0,058 = 5.8%

Si tiene osteoporosis… ¿probabilidad de menopausia?

- P(Menopausia|Osteoporosis)=58/64=0,906 = 90,6%

¿Probabilidad de menopausia y no osteoporosis?

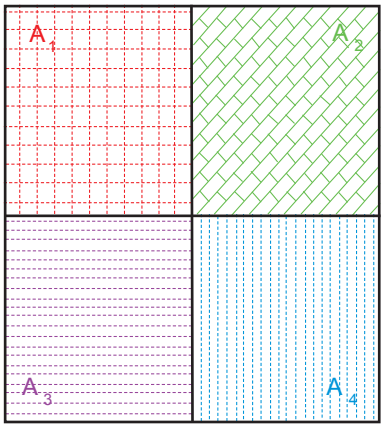
- P(Menopausia ∩ No Osteoporosis) = 639/1000=0,639 = 63,9%

Si tiene no tiene osteoporosis… ¿probabilidad de no menopausia?

- P(No Menopausia|NoOsteoporosis)=297/936=0,317 = 31,7%

Si generalizamos la propiedad de la probabilidad total para cuatro eventos exhaustivos y mutuamente excluyentes como lo indica la figura.

Sistema exhaustivo y excluyente de eventos ó sucesos

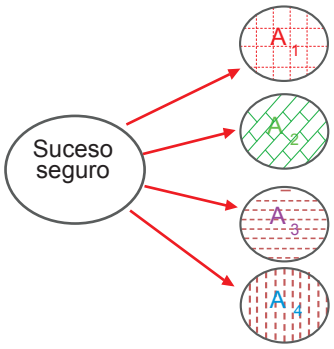


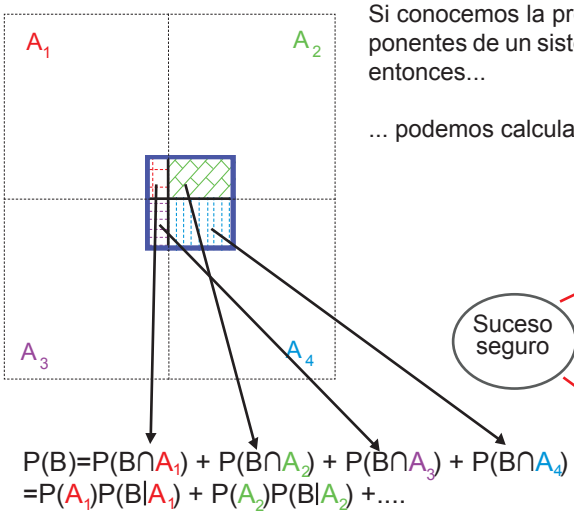
Son una colección de eventos ó sucesos

A_1, A_2, A_3, A_4 ...

Tales que la unión de todos ellos forman el espacio muestral, y sus intersecciones son disjuntas.

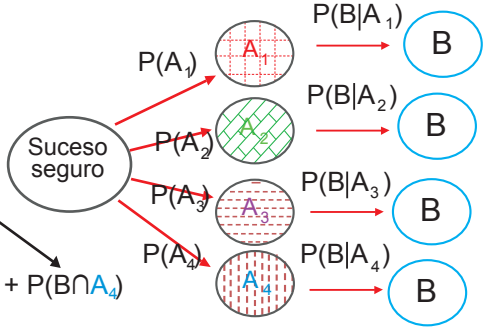
¿Recordemos cómo formar intervalos en tablas de frecuencias?





Si conocemos la probabilidad de B en cada uno de los componentes de un sistema exhaustivo y excluyente de sucesos, entonces...

... podemos calcular la probalidad de B.



$P(B)=P(B\cap A_1) + P(B\cap A_2) + P(B\cap A_3) + P(B\cap A_4)$
$=P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + ...$

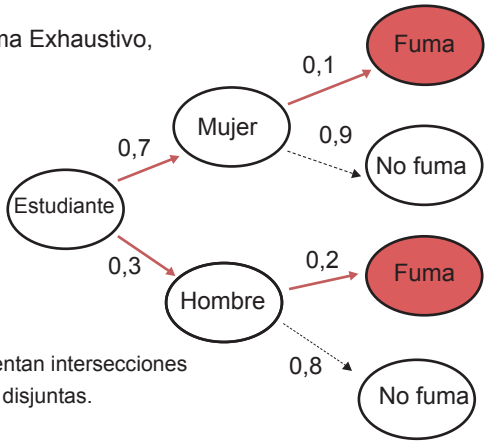
ACTIVIDAD DE APRENDIZAJE DESARROLLADA 02

En este curso 70% de los alumnos son mujeres.
De ellas el 10% son fumadoras. De los hombres, son fumadores el 20%.

¿Qué porcentaje de fumadores/as hay en el curso?

Propiedad de Probabilidad Total
Hombres y mujeres forman un sistema Exhaustivo, excluyente de eventos ó sucesos

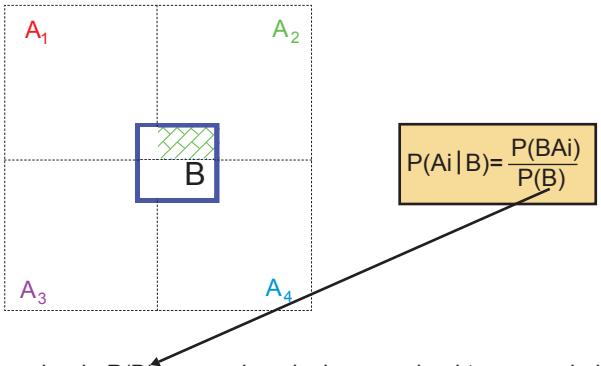
-P(F)=P(M∩F) + P(H∩F)
=P(M)P(F|M) + P(H)P(F|H)
=0,7 x 0,1 +0,3 x 0,2
=0,13 = 13%



- Los caminos a través de nodos representan intersecciones
- Las bifurcaciones representan uniones disjuntas.

Teorema de Bayes

Si conocemos la probabilidad de B en cada uno de los componentes de un sistema exhaustivo y excluyente de eventos, entonces, si ocurre B, podemos calcular la probabilidad (a posteriori) de ocurrencia de cada A_i.



donde P(B) se puede calcular usando el teorema de la probalidad total:

$P(B)=P(B\cap A_1) + P(B\cap A_2) + P(B\cap A_3) + P(B\cap A_4)$
$=P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \ldots.$

ACTIVIDAD DE APRENDIZAJE DESARROLLADA 03

En esta aula el 70% de los estudiantes son mujeres. De ellas el 10% son fumadoras. De los hombres, son fumadores el 20%.

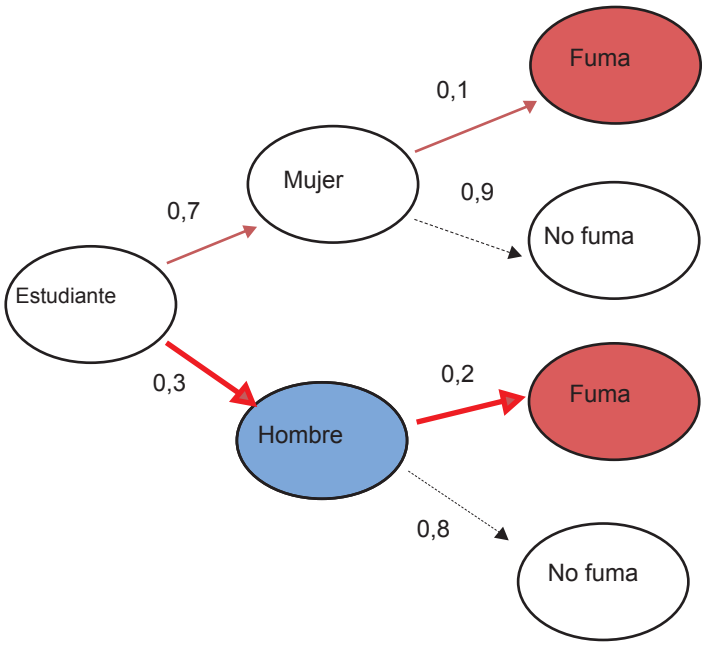
¿Qué porcentaje de fumadores hay?

– P(F) = 0,7 x 0,1 + 0,3 x 0,2 = 0,13

Se elige a un estudiante al azar y es… fumador ¿Cuál es la probabilidad de que sea un hombre?

$$P(H|F) = \frac{P(H \cap F)}{P(F)} = \frac{P(H) \cdot P(F|H)}{P(F)}$$

$$= \frac{0,3 \times 0,2}{0,13} = 0,46$$



Obervación.- Es interesante tomar en cuenta actividades de lo cotidiano, hoy en el Ecuador ha aumentado la enfermedad de la diabetes en niños y jovenes según el Instituto Nacional de Estadistica y Censos del Ecuador, INEC. Entonces, consideramos la siguiente actividad de aprendizaje de prueba diagnòstica: Diabetes.

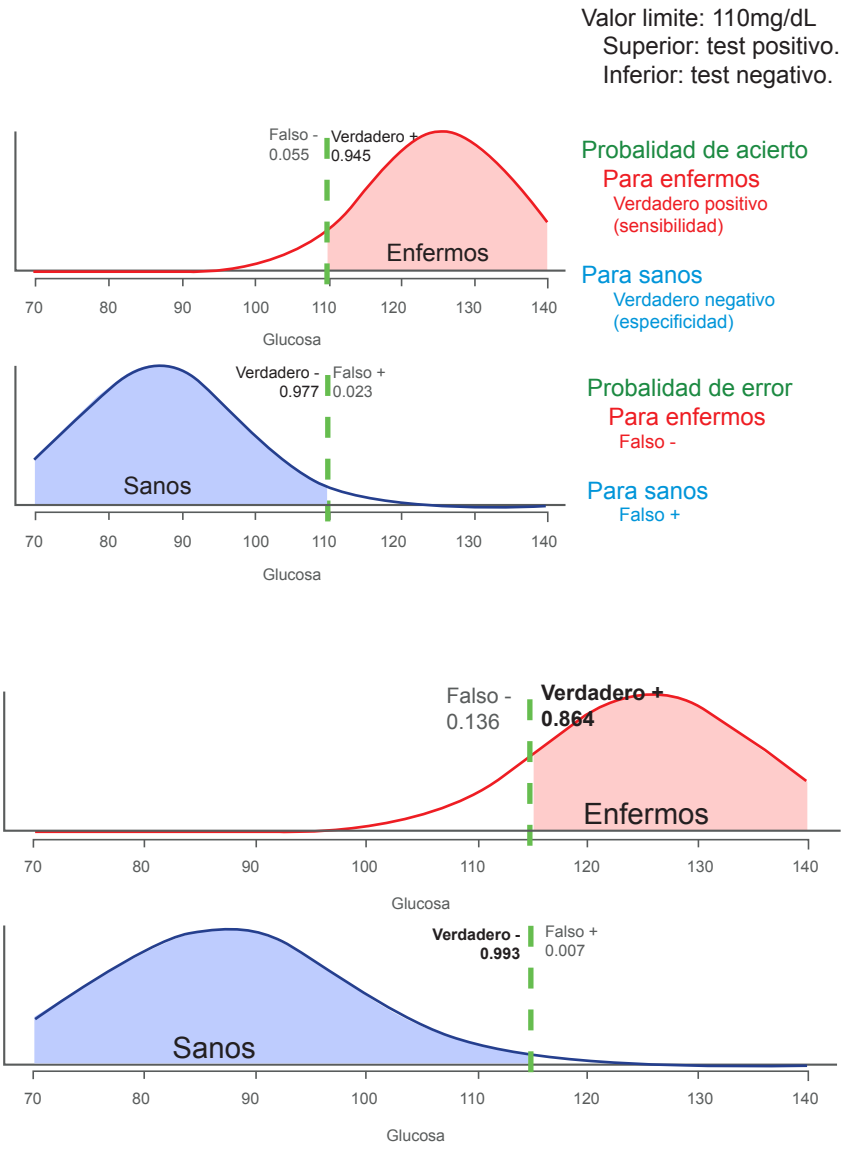
Los carbohidratos ingeridos terminan como glucosa en la sangre. El exceso se transforma en glucógeno y se almacena en hígado y músculos. Este se transforma entre comidas de nuevo en glucosa según necesidades.

La principal hormona que regula su concentración es la insulina. La diabetes provoca su deficiencia o bien la insensibilidad del organismo a su presencia. Es una enfermedad muy común que afecta al 2% de la población (prevalencia)

Una prueba común para diagnosticar la diabetes, consiste en medir el nivel de glucosa. En individuos sanos suele variar entre 64 y 110mg/dL.

El cambio de color de un indicador al contacto con la orina suele usarse como indicador (resultado del test positivo). Valores por encima de 110 mg/dL se asocian con un posible estado pre-diabético. Pero no es seguro. Otras causas podrían ser: hipertiroidismo, cáncer de páncreas, pancreatitis, atracón reciente de comida… Supongamos que los enfermos de diabetes, tienen un valor medio de 126mg/dL.

Funcionamiento de la prueba diagnóstica de glucemia



No es simple. ***No es posible aumentar sensibilidad y especificidad al mismo tiempo.*** Hay que elegir una solución de compromiso: Aceptable sensibilidad y especificidad.

Una prueba diagnóstica ayuda a mejorar una estimación de la probabilidad de que un individuo presente una enfermedad.

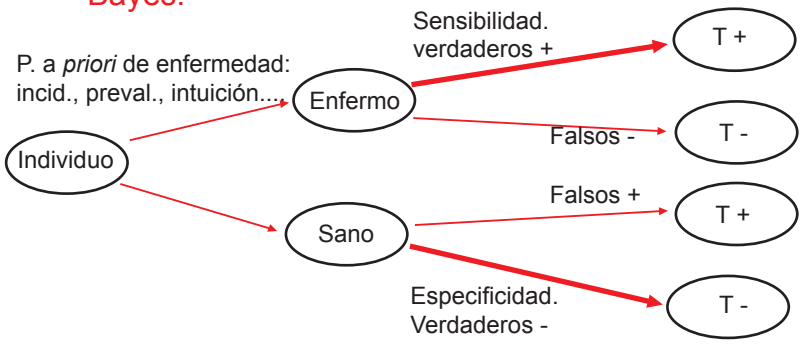
En principio tenemos una idea subjetiva de P(Enfermo). Nos ayudamos de…

- Incidencia: Porcentaje de nuevos casos de la enfermedad en la población.
- Prevalencia: Porcentaje de la población que presenta una enfermedad.

Para confirmar la sospecha, usamos una prueba diagnóstica. Ha sido evaluada con anterioridad sobre dos grupos de individuos: sanos y enfermos. Así de modo frecuentista se ha estimado:

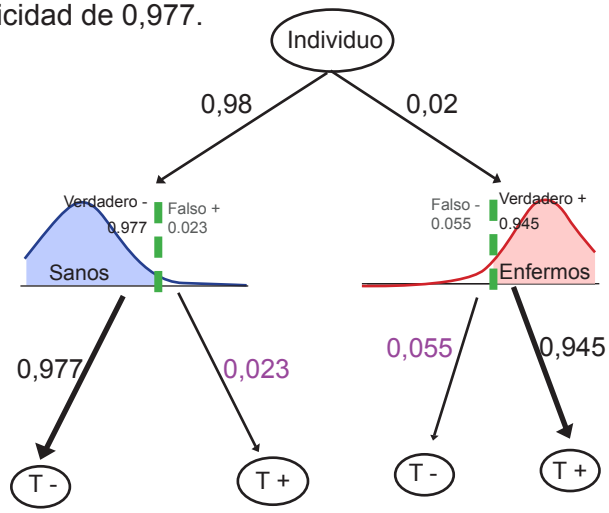
- P(+ | Enfermo)= Sensibilidad (verdaderos +)= Tasa de acierto sobre enfermos.
- P(- | Sano) = Especificidad (verdaderos -)= Tasa de acierto sobre sanos.
- A partir de lo anterior y usando el teorema de Bayes, podemos calcular las probabilidades a posteriori (en función de los resultados del test): Índices predictivos
- P(Enfermo | +) = Índice predictivo positivo
- P(Sano | -) = Índice predictivo negativo

Pruebas diagnósticas: aplicación Teorema de Bayes.



Índices predictivos

- La diabetes afecta al 2% de los individuos.
- La presencia de glucosa se usa como indicador de diabetes.
- Su sensibilidad es de 0,945.
- La especificidad de 0,977.



Calcular los índices predictivos. En efecto,

$$P(Sano\,|\,T-) = \frac{P(Sano \cap T-)}{P(T-)} = \frac{P(Sano)\,P(T-\,|Sano)}{P(Sano)\,P(T-\,|Sano) + P(Enf)\,P(T-\,|Enf)} = \frac{0{,}98 \cdot 0.977}{0{,}98 \cdot 0{,}977 + 0{,}02 \cdot 0{,}055} = 0{,}999$$

$$P(Enfermo\,|\,T+) = \frac{P(Enf \cap T+)}{P(T+)} = \frac{P(Enfermo)\,P(T+|\,Enfermo)}{P(Sano)\,P(T+|\,Sano) + P(Enfermo)\,P(T+|\,Enfermo)} = \frac{0{,}02 \cdot 0.945}{0{,}02 \cdot 0{,}945 + 0{,}98 \cdot 0{,}023} = 0{,}456$$

Observaciones

- En la actividad de aprendizaje de prueba diagnóstica: Diabetes, al llegar un individuo a la consulta tenemos una idea *a priori* sobre la probabilidad de que tenga una enfermedad.

- ¿ Qué probabilidad tengo de estar enfermo ?

- En principio un 2%. Le haremos unas pruebas.

- A continuación se le pasa una prueba diagnóstica que nos aportará nueva informaciòn: Presenta glucosa o no.

- En función del resultado tenemos una nueva idea(*a posteriori*) sobre la probabilidad de que esté enfermo.

  - Nuestra opinión a priori ha sido modificada por el resultado de un exprerimento.

- Presenta glucosa. La probabilidad ahora es del 45,6%

3.4 VARIABLES ALETORIAS Y DISTRIBUCIONES DE PROBABILIDAD

3.4.1 DEFINICIÓN Y CLASIFICACIÓN DE LAS VARIABLES ALETORIAS

Una función

$$X: \Omega \rightarrow R$$

tal que para todo subconjunto B de R. $X^{-1}(B)$ es un evento, se denomina ***variable aleatoria*** (v.a.). O sea una v.a. es una función de valores reales definida en un espacio muestral Ω.

Se dice que X es *"aleatoria"* porque involucra la probabilidad. Las v.a. se clasifican en dos grupos: ***discretas y continuas.***

Discretas

Una v.a. X se dice discreta si el recorrido X(Ω) es un conjunto numerable de valores (finito o infinito).

Continuas

Una v.a. X sobre un espacio (Ω, F, P) se dice continua si el recorrido X(Ω) consiste en uno o más intervalos de la recta de los reales.

ACTIVIDAD DE APRENDIZAJE PROPUESTA

Clasifique las siguientes v.a. como discretas ó continuas.

X: " Número de accidentes automovilísticos por año en Riobamba"_____

Y: " Tiempo de duración de una lámpara"_____

M: " Cantidad de leche producida anualmente por una vaca particular"_____

N: " Número de huevos puestos cada mes por una gallina"_____

P: " Peso de un cierto grano producido en una hectárea de terreno"_____

Q: "Número de matrículas en el semestre oct. 2014 - feb. 2015 de la ESPOCH"_____

R: " Estatura de los estudiantes de la Facultad de Ciencias de la ESPOCH"_____

S: "Calificación o Puntaje con enteros entre 0-10 en la ESPOCH"_____

3.4.2.1 Distribución de probabilidad de variables aleatorias discretas.

Se han utilizado mayúsculas, como X, para denotar variables aleatorias; se utilizará minúsculas como x, para denotar valores particulares que puede tomar una v.a.

La expresión (X=x) se puede leer como "***el conjunto de todos los puntos de Ω a los que la v.a. X les asignó el valor x".***

Ahora tiene sentido hablar de la probabilidad de que X tome el valor x denotado por P(X=x) o p(x). Esta probabilidad se define como la suma de probabilidades de ciertos puntos muestrales. También se puede indicar p(x) = P(X=x).

Definición.- Sea X una v.a. discreta, se llama a p(x) = P(X=x) ***función de probabilidad de X(o distribución de probabilidad),*** si satisface las siguientes propiedades:

$$1)\quad p(x) \geq 0 \quad \forall x \in X$$

$$2)\quad \sum_{x} p(x) = 1$$

Observación. Se observa que al hacer referencia a la distribución de probabilidad de X no solo implica la existencia de la función de probabilidad sino también la existencia de la función de distribución acumulativa.

Definición.- La ***función de distribución acumulativa de X discreta*** es la probabilidad de que X sea menor o igual a un valor específico de X, x_0 y está dada por:

$$F(x_0) = P(X \leq x_0) = \sum_{x_i \leq x_0} p(x_i)$$

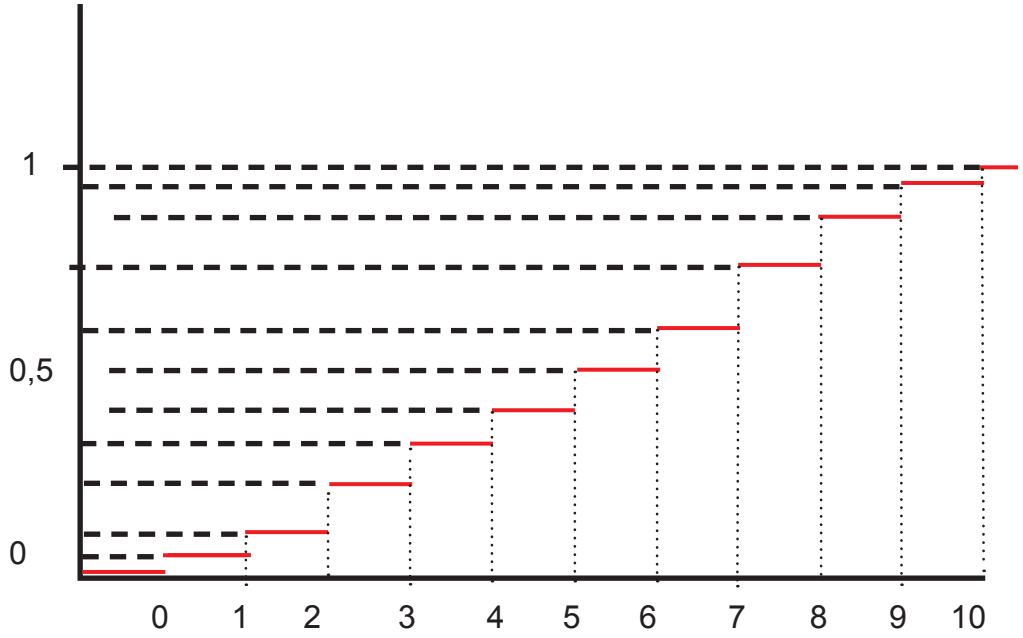
En general, la función de distribución acumulativa F(x) de una v.a. discreta es una función no decreciente de los valores de X, de tal manera que:

1. $0 \leq F(x) \leq 1 \quad \forall x \in X$.
2. $F(x_i) \leq F(x_j)$ si $x_i < x_j$.
3. P(X>x) = 1 - F(x).

Además, puede establecerse que para una v.a. de valor entero se tiene:

$$P(X=x) = F(x) - F(x-1)$$
$$P(x_i \leq X \leq x_j) = F(x_j) - F(x_i - 1)$$

Observación.- La función de distribución acumulativa de una v.a discreta es una función escalonada, como se indica en la siguiente gráfica.



3.4.2.2 Distribución de probabilidad de variables aleatorias continuas

La distribución de probabilidad de una v.a. continua X está caracterizada por una función f(x) que recibe el nombre de función de ***densidad de probabilidad.***

La función de densidad de probabilidad no representa la probabilidad de que X = x. más bien, ésta proporciona un medio para determinar la probabilidad de un intervalo [a, b] por ejemplo.

Si existe una función f(x) tal que satisface

1) $\int_{-\infty}^{+\infty} f(x)dx = 1$

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

para cualesquiera $a,b \in R$ y $f(x)$ es ***la función de densidad de probabilidad de v.a. continua X.***

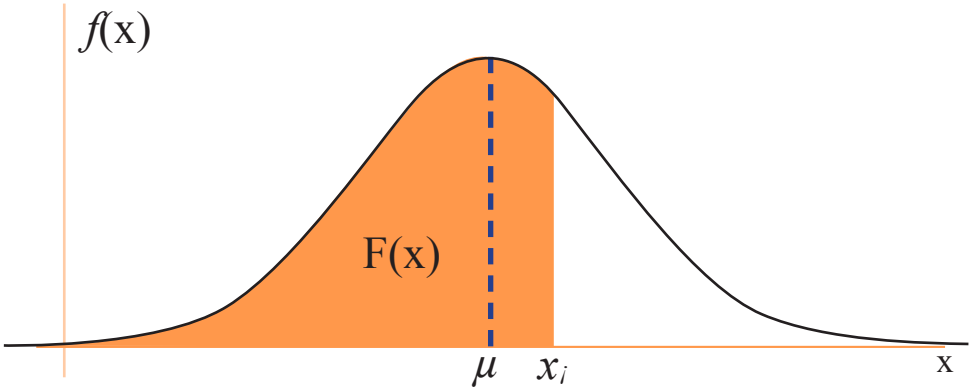
La función de distribución acumulativa de una v.a. continua X o distribución acumulada de X es la probabilidad de que X tome un valor menor o igual a algún x_i específico. Esto es:

$$F(x_i) = P(X \leq x_i) = \int_{-\infty}^{x_i} f(t)dt$$

Nota. *En el caso discreto, se asignan probabilidades positivas a todos los valores puntuales de la v.a. pero la suma de todos ellos es 1, aún a pesar de que el conjunto de valores sea infinito numerable.*

Para el caso de una v.a. continua, lo anterior, no es posible. Se verá que la probabilidad de que una v.a. continua X tome un valor específico x es cero.

Geométricamente $F(x_i)$ es el área acotada por la función de densidad *f(x)* y la recta X = x_i. Dado que $P(X=x_i) = 0$ entonces $P(X \leq x_i) = P(X < x_i) = F(x_i)$. En general $F(x) = P(X \leq x)$



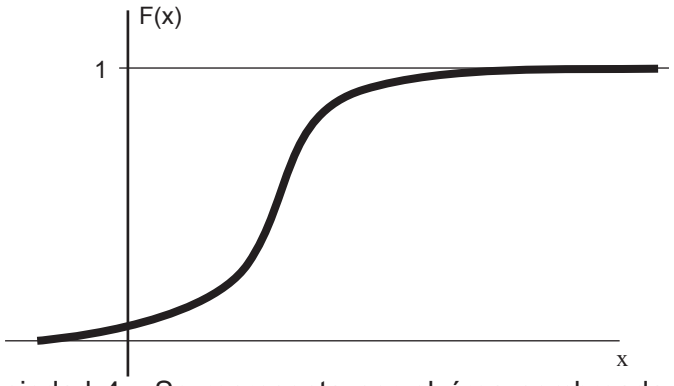
Propiedades de la distribución acumulada

La distribución acumulada F(x) es una función suave no decreciente en $-\infty < x < +\infty$ y satisface las siguientes propiedades:

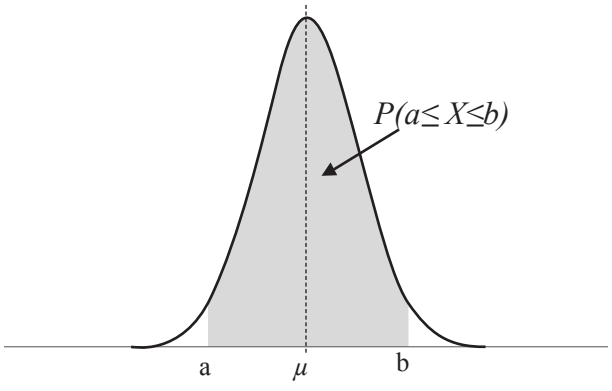
1. $\lim_{x \to \infty} F(x) = F(+\infty) = 1$
2. $\lim_{x \to -\infty} F(x) = F(-\infty) = 0$
3. $F(x_i) \leq F(x_j)$ si $x_i < x_j$
4. $P(a < X < b) = F(b) - F(a)$
5. $dF(x)/dx = f(x)$, o sea, f(x) es la primitiva de F(x).

Con estas propiedades se puede definir una v.a. continua, en efecto, sea X una v.a, con una función de distribución acumulada F(x), se dice que X es continua si F(x) es continua, para $-\infty < x < +\infty$. Exponemos a continuación gráficas de algunas propiedades de la distribución acumulada.

Para las propiedades 1, 2 y 3



Para la propiedad 4. Se representa por el área sombreada (área bajo la función de densidad f(x) y limitada por las rectas x = a y x = b).



La esperanza matemática o el valor esperado de una v.a. X, el cual denotamos por E(X), está dada por

$$E(X)=\sum_i x_i p(x_i) \quad \text{si X es discreta;} \quad \text{ó} \quad E(X)=\int_{-\infty}^{+\infty} xf(x)dx \quad \text{si X es continua}$$

En donde $p(x)\ y\ f(x)$ son las funciones de probabilidad y de densidad de probabilidad respectivamente, dependiendo de que la sumatoria o la integral, converjan absolutamente, es decir.

$$\sum_i x_i p(x_i) < +\infty \quad \text{ó} \quad \int_{-\infty}^{+\infty} xf(x)dx < +\infty$$

Si la suma o la integral no convergen absolutamente entonces el valor esperado no existe(o no tiene esperanza finita).

Observación.- La esperanza matemática de una v.a. X es una propiedad de la distribución de probabilidad de X.

Además no es una función de xϵX, sino un número fijo que indica el promedio estadístico o media de un número grande de observaciones xϵX, por otro lado E(X) no necesariamente está definido en el recorrido de X, es decir; en X(Ω) y se denota por μ a la media de la población, esto es μ = E(X).

Sea X:"Calificaciones de 4 asignaturas sobre 10 de un estudiante del sexto nivel de la carrera de Ingeniería Estadística Informática de la ESPOCH", supongamos que estas cuatro calificaciones son: 8, 7, 7 y 9 entonces el rango de X es X(Ω) = {8, 7, 7, 9 } y su valor (esperado) obtenemos en el Excel, además del rango y mínimo.

CALIFICACIONES	
Media	7,75
Rango	2
Mínimo	7
Máximo	9

Nota. *Note que la esperanza matemática o valor esperado de X es 7,75, es decir E(X)∉X(Ω).*

Si el estudiante referido tuvo las mismas calificaciones en estas cuatro asignaturas supongamos 8 entonces X(Ω) = {8, 8, 8, 8 } por tanto E(Ω) = 8 que es la media o promedio de las calificaciones y diremos que la v.a. X si tiene los mismos valores, es constante.

Propiedades de la esperanza matemática

Demuestre que si X es una v.a. cualquiera con valor esperado μ = E(X), a y b constantes reales, entonces E(a) = a o E(b) = b y además:

si Y = aX + b, entonces E(Y) = aE(X) + b = a*μ + b

Definición.- La varianza de toda v.a. X se denota por Var (X) o σ^2 (X) y está dada por

$$Var(X) = E[(X-\mu)^2]$$

Donde μ = E(X) y desarrollando ésta expresión se obtiene:

$$Var(X) = E(X^2) - E^2(X) = E(X^2) - \mu^2$$

Nota.- *Recuerde que la esperanza matemática o media y la varianza son medidas de centralización y de dispersión de la distribución de probabilidad respectivamente, entonces ¿cómo están asociadas? La respuesta la damos con la siguiente medida adimensional ¿Por qué es adimensional?...*

Propiedades de la varianza.

Demuestre que:

Propiedad 1.- La varianza de toda v.a. X es positiva o cero, esto es $Var(X) \geq 0$
Propiedad 2.- Si X es una v.a. y a, b son dos constantes reales entonces

$$Var(aX+b) = \sigma^2(aX+b) = a^2\sigma^2(X).$$

Además por la propiedad 1, se puede determinar la raíz cuadrada positiva de la varianza y este valor recibe el nombre de **desviación estándar** y se denota por σ. Se puede emplear también la notación d.e.(X) o sd(X).

Definición.- Una medida que compara la dispersión relativa de dos o más distribuciones de probabilidad es el ***coeficiente de variación*** que está definido por:

$$cv = \sigma/\mu \quad \text{(desviación estándar/ esperanza matemática)}$$

Si estuviéramos manejando datos sobre una muestra entonces

$$CV = S/\bar{X}$$

donde S es la desviación estándar muestral y $\bar{X}$ es la media muestral.

3.5.1 COEFICIENTES DE ASIMETRÍA Y CURTOSIS

3.5.1.1 Coeficiente de asimetría

Una medida que le indica la forma o el sesgo de la curva de una distribución de datos se llama ***sesgo o asimetría*** y es apropiada tomar e indicar los ***momentos centrados de orden n*** de una v.a. X como $\mu^n = E(X-\mu)^n$ con $n \epsilon N$, entonces la asimetría de una v.a. X, está definido por:

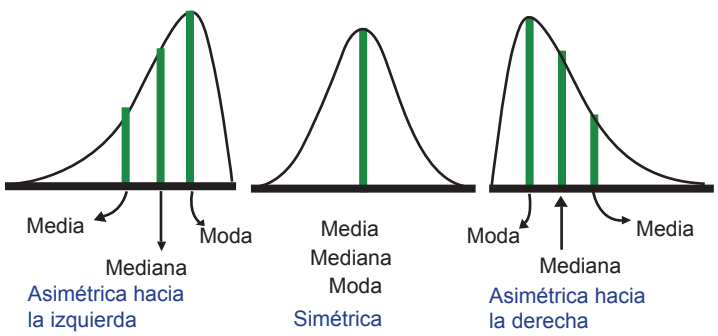
$$\alpha_3 = \mu^3 / \sigma^3 \; ; \; donde \; \mu^3 = E(X-\mu)^3$$

α_3 , recibe el nombre de ***coeficiente de asimetría.***

Donde μ^3 es el momento centrado de orden 3. Dependiendo de los valores que tome se definirá:

- $\alpha_3 < 0$ la distribución es asimétrica hacia la izquierda o presenta sesgo negativo.
- $\alpha_3 = 0$ la distribución es simétrica o presenta sesgo cero.
- $\alpha_3 > 0$ la distribución es asimétrica hacia la derecha o presenta sesgo positivo.

La siguiente gráfica presenta las tres formas de una distribución de datos y se observan cómo se presenta las tres medidas de agrupación: Media, Mediana y Moda



Resultado importante

- Si $\alpha_3 < 0$ entonces Media < Mediana < Moda
- Si $\alpha_3 = 0$ entonces Media = Mediana = Moda
- Si $\alpha_3 > 0$ entonces Media > Mediana > Moda

3.5.1.2 Coeficiente de curtosis

Es una medida que indica qué tan apuntada (alargada) es la distribución de probabilidad y recibe el nombre de curtosis. Al igual que para el coeficiente de asimetría, es preferible emplear el cuarto momento centrado y se llama ***coeficiente de curtosis*** a:

$$\alpha_4 = \mu^4 / \sigma^4 \text{ ; donde } \mu^4 = E(X-\mu)^4$$

Las distribuciones de probabilidad que presentan un pico si:

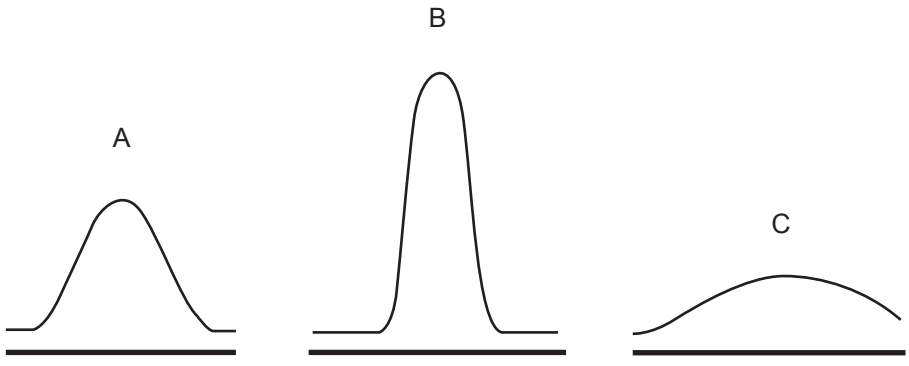
- $\alpha_4 = 3$ la distribución no presenta un pico muy alto ni muy bajo y se llama mesocúrtica.
- $\alpha_4 > 3$ el pico que presenta es relativamente alto y se llama **leptocúrtica.**
- $\alpha_4 < 3$ la distribución es relativamente plana y se llama **platicúrtica.**

Nota. La curtosis también se define y denota por $K = \alpha_4 - 3$

Resultado importante

A. Si K = 0 entonces la distribución es mesocúrtica (normal)
B. Si K > 0 entonces la distribución es leptocúrtica(más apuntada)
C. Si K < 0 entonces la distribución es platicúrtica (aplanada)

La siguiente gráfica presenta las tres formas de una distribución de probabilidad



Mesocúrtica (Normal) k=0

Leptocúrtica K > 0

Platicúrtica K< 0

Nota. Los coeficientes α_3 *y* α_4 *también se denominan* ***factores de forma****, debido a que en gran medida, determinan la forma de la distribución de probabilidad, estos valores pueden calcularse mediante software estadístico como Minitab y están determinados por skewness (asimetria) y kurtosis (curtosis).*
Los valores $\alpha_3=0$ *y* $\alpha_4=3$ *(K = 0) de asimetría y curtosis respectivamente se toman de la distribución normal que definiremos más adelante.*

ACTIVIDAD DE APRENDIZAJE PROPUESTA

Nota. La estandarización de una v.a. afecta a la media y a la varianza, pero no afecta a los factores de forma como lo indica el numeral 2 de ésta actividad de aprendizaje propuesta.

Además si x es un valor de X el valor $z = (x-\mu)/\sigma$ es la estandarización, la desviación del valor x del valor esperado μ en términos de las unidades de la desviación estándar.

Sea X una v.a. cualquiera no constante con media o esperanza matemática $\mu = E(X)$ y desviación estándar σ. Llamamos ***variable aleatoria estandarizada*** o simplemente variable estándar de X a la v.a.

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

Se demuestra que la v.a. estandarizada Z:

1. Z es centrada, es decir, $E(Z) = 0$ y $\sigma(Z) = 1$.
2. $\alpha_3(Z) = \alpha_3(X)$ y $\alpha_4(Z) = \alpha_4(X)$

3.6 DISTRIBUCIONES: BINOMINAL, POISSON Y NORMAL

En esta parte se pone en juego los conocimientos vistos principalmente del párrafo anterior. De la clasificación de variables aleatorias cuantitativas: discretas y continuas tenemos que existen distribuciones discretas de probabilidad: ***binomial*** (Bernoulli), ***Poisson, geométrica, hipergeometrica, binomial negativa*** entre otras y distribuciones continuas de probabilidad: ***normal, t-student, Chi-cuadrada, F, uniforme continua, gamma, beta, Weibull, exponencial*** y otras.

Se verán principalmente en esta parte las dos primeras familias de distribución de probabilidad discreta y una distribución continua de probabilidad, la normal.

3.6.1 DISTRIBUCIÓN BINOMIAL

Un experimento binomial es aquel que tiene las siguientes características:

a) El experimento consta de n pruebas idénticas.
b) Cada prueba tiene dos resultados posibles. Se llamará a uno el éxito E y al otro fracaso F.
c) La probabilidad de tener éxito en una sola prueba es igual a p y permanece constante de prueba en prueba, la probabilidad de un fracaso es igual a q = 1 - p.
d) Las pruebas son independientes.
e) La v.a. bajo estudio es X: "número de éxitos observados en las n pruebas".

La pregunta que debemos contestar en problemas que presentan estas características es: ***¿cuál es la probabilidad de que este experimento tenga x éxitos?***

Cada punto muestral de Ω se puede denotar mediante una n-ada, de elementos E y F. Por ejemplo el punto muestral EEEFEFFE...EF en donde la letra en la i-ésima posición indica el resultado de la i-ésima prueba.

Considérese ahora un típico punto muestral con x éxitos en el evento numérico X=x.

$$\underline{EEE...EE}\ \underline{FFF...FF}$$

$$x \qquad n\text{-}x$$

Es la intersección de n pruebas independientes, por lo tanto la probabilidad de este punto muestral es el producto de x éxitos y (n-x) fracasos por lo que su probabilidad está dada por:

$$p.p.p...pp\ qqq...qq = p^x q^{n-x}$$

Cualquier otro punto muestral del evento numérico X = x aparecerá como un arreglo de las letras E y F que contendrá x letras E y (n-x) letras F y que se determinarán con la misma probabilidad, siendo que el número de arreglos distintos de x letras E y (n-x) letras F es

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

maneras distintas de escoger x elementos de entre n. Es decir, son las combinaciones de n objetos tomados x. La pregunta queda entonces contestada por la siguiente definición.

Definición.- Sea X una v.a. que representa el número de éxitos en n pruebas y p la probabilidad de éxito. Se dice entonces que la v.a. X tiene una distribución binomial con función de probabilidad p(x; n, p) si.

$$p(x;\ n,\ p) = \begin{cases} \binom{n}{x} p^x\,(1\text{-}p)^{n-x}\ con...x = 0,1,2,...,n \quad 0 \le p \le 1 \\ 0\ para\ cualquier\ otro\ valor \end{cases}$$

Los parámetros de ésta distribución son n y p; éstos parámetros definen una familia de distribuciones binomiales.

El término “binomial” proviene de las probabilidades p(x; n, p), que son términos del desarrollo del binomio.

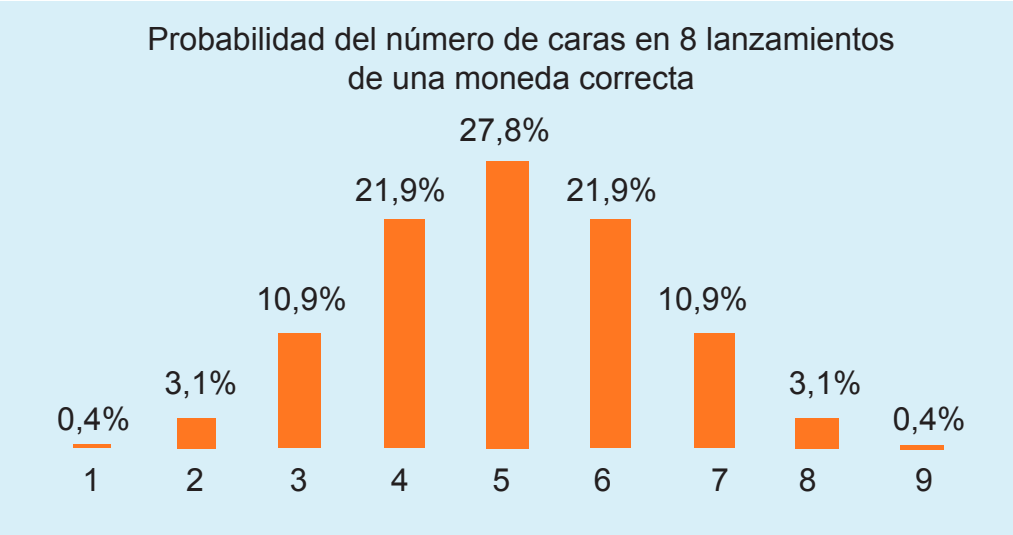
$$(q+p)^n = \binom{n}{0} q^n + \binom{n}{1} pq^{n-1} + \binom{n}{2} p^2 q^{n-2} + ... + \binom{n}{n} p^n = \sum_{x=0}^{n} p(x,n.p)$$

Probemos que p(x; n, p) es efectivamente una f.p. en efecto,

1) $\forall\, x;\ x$=0, 1, 2, ..., n $\quad p(x;n,p) \ge 0.$

2) $$\sum_{x=0}^{n} p(x;n.p) = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} = (q+p)^n = 1$$

Si lanzamos una moneda correcta, p = 0.5; 8 veces.
¿Qué representa gráficamente los resultados cuando cae cara?



Otros experimentos binomiales:

1) El lanzamiento de una moneda, es un experimento binomial cuando lanzamos n veces la moneda y la probabilidad de obtener cara es 1/2.

2) También es un experimento binomial si un estudiante que no se ha preparado en absoluto para un examen, observa que éste contiene 15 items de verdadero y falso. Decide lanzar al aire una moneda para responder: anota V si la moneda muestra cara y F si muestra sello.

3) La ley binomial es aplicada en el control de calidad, por ejemplo si se tiene un lote de artículos entre defectuosos y no defectuosos y se quiere ver si aceptamos o rechazamos el lote.

Definición.- La distribución acumulativa de una v.a. X con ley binomial se determina

$$P(X \leq x) = F(x; n, p) = \sum_{i=0}^{x} \binom{n}{i} p^i (1 - p)^{n-i}$$

Las probabilidades individuales se calculan por:

$$p(x;n, p) = F(x;n, p) - F(x-1;n, p)$$

Se pueden probar los valores de la siguiente tabla para una distribución binomial:

		FACTORES DE FORMA	
MEDIA	VARIANZA	COEFICIENTE DE ASIMETRIA	COEFICIENTE DE CURTOSIS
np	npq	$\alpha 3 = \frac{q - p}{(npq)^{1/2}}$	$\alpha 4 = 3 + \frac{1 - 6pq}{npq}$

Con q=1 - p.

ACTIVIDAD DE APRENDIZAJE PROPUESTA

1. Pruebe de la tabla anterior para una distribución binomial que:

   Si p<1/2 entonces la ley o distribución binomial presenta un sesgo positivo.
   Si p=1/2 entonces la ley binomial es simétrica.
   Si p>1/2 entonces la ley binomial tiene sesgo negativo.

2. Demuestre los valores de la tabla anterior. (Consulte las fórmulas de los factores de forma)

3. (Intente resolver). ¿Qué valor o valores de p se consideraría para que la ley binomial sea leptocúrtica, mesocúrtica y platicúrtica? Aplique los valores de la tabla anterior

ACTIVIDAD DE APRENDIZAJE DESARROLLADA 04

Supongamos que el 30% de los estudiantes de la unidad educativa Pensionado Olivo de Riobamba no paga puntualmente las pensiones mensuales, hallar la probabilidad de que en una muestra aleatoria de 15 estudiantes el número de estudiantes que no pagan la pensión sea:

(a) exactamente 5
(b) mayor que 5
(c) cinco o menos
(d) un número comprendido entre 6 y 10

Solución:

Sea X: "Número de estudiantes que no pagan pensiones" con probabilidad p = 0.30 y n = 15 entonces q = 1 - p = 0.70 y

a) P(X=5) = p(5;15,0.3) = 0.2061 (o véase tabla de la distribución binomial p(5;15,0.30) = 0.2061)

b) P(X>5) = 1 - P(X≤ 5) = 1 - F(5;15,0.30) = 1 - 0.7216 = 0.2784.

c) P(X≤5) = F(5;15,0.3) = 0.7216

d) P(6≤X≤10) = F(10;15,0.3) - F(5;15,0.3)= 0.9993 - 0.7216 = 0.2777

ACTIVIDAD DE APRENDIZAJE DESARROLLADA 05

Supóngase que un lote de 300 fusibles eléctricos contiene 5% de defectuosos. Determine la probabilidad que se puede encontrar al menos un fusible defectuoso en una muestra de cinco fusibles.

Solución:

Sea X: "número de fusibles defectuosos observados" con probabilidad p = 0.05, n = 5.

$$P(X\geq 1) = 1 - P(X=0) = 1 - p(0;5,0.05) = 1 - (0.95)^5 = 0.226.$$

Observando las tablas de probabilidades binomiales del apéndice tenemos:

$$P(X\geq 1) = 1 - P(X=0) = 1 - p(0;5,0.05) = 1 - 0.7738 = 0.2262$$

Observación. Obsérvese que existe una probabilidad bastante grande de obtener al menos un defectuoso, aunque la muestra sea relativamente pequeña.

ACTIVIDAD DE APRENDIZAJE DESARROLLADA 06

El lote grande de fusibles de la actividad de aprendizaje anterior supuestamente contiene solamente el 5% de defectuosos. Determine la probabilidad de que se encuentren al menos tres defectuosos en una muestra aleatoria de 20 fusibles.

Solución:

Sea X: "número de fusibles defectuosos en la muestra" con probabilidad p = 0.05 y

$$n = 20, \text{ luego } P(X\geq 3) = 1 - P(X\leq 2) = 1 - 0.9245 = 0.0755$$

Observando las tablas de probabilidades binomiales del apéndice tenemos:

$$P(X\geq 3) = 1 - P(X\leq 2) = 1 - [p(0;20,0.05) + p(1;20,0.05) + p(2;20,0.05)]$$
$$=1-(0.3585 + 0.3774 + 0.1887) = 1 - 0.9245 = 0.0755$$

Esta probabilidad es relativamente pequeña, lo que nos lleva a concluir que, si se observaran realmente más de tres defectuosos en los 20 fusibles, la proporción de defectuosos del 5% está equivocada.

3.6.1.1 Distribución de Bernoulli.

Una variable aleatoria X tiene distribución de Bernoulli de parámetro p con $0 \leq p \leq 1$, si $P(X=0) = 1- p = q$, $P(X = 1) = p$.

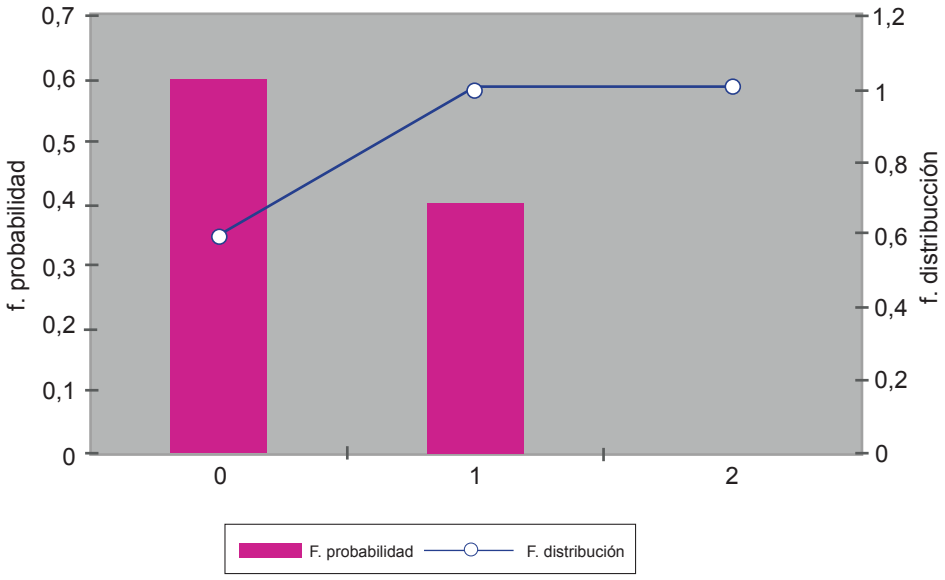
Es decir,

$$p(x,p)=\begin{cases} p^{x}\, q^{1-x} \;\; si\_ x=0,1 \\ 0...\text{para} \cdot cualquier \;\_otro\_valor \end{cases}$$

Con función de distribución acumulada.

$$F(x,p) = \begin{cases} 0,.....x < 0 \\ q,.....0 \leq x < 1 \\ 1,.....x \geq 1 \end{cases}$$

Distribución de Bernoullil b(0.4)



Nota. *Se note que la distribución de Bernoulli es un caso particular de la binomial cuando* $n = 1$, *luego* $\mu = p$ $Var(X) = pq$

Se puede probar los valores de la siguiente tabla **para una distribución de Bernoulli**

		FACTORES DE FORMA	
MEDIA	VARIANZA	COEFICIENTE DE ASIMETRIA	COEFICIENTE DE CURTOSIS
p	pq	$\alpha 3 = \frac{q - p}{(pq)^{1/2}}$	$\alpha_4 = 3 + \frac{1 - 6pq}{pq}$

3.6.2 LA DISTRIBUCIÓN DE POISSON

La distribución de Poisson es otra distribución discreta de probabilidad muy útil en la que la v.a. representa el número de eventos independientes que ocurren a una velocidad constante. En el campo educativo generalmente no se aplica, pero es bueno ver desde el punto de vista de la teoría de los límites que una distribución Binomial se aproxima a una distribución de Poisson.

Definición.- Sea X una v.a que representa el número de eventos aleatorios independientes que ocurren a una rapidez constante sobre el tiempo o el espacio. Se dice entonces que la v.a. X tiene una ***distribución de Poisson*** con función de probabilidad (f.p.)

$$(*)\quad p(x,\lambda)=\begin{cases}\dfrac{e^{-\lambda}\lambda^{x}}{x!},\ldots x=0,1,2,\ldots;\ \lambda>0\\ 0,\ldots para\cdot cualquier\cdot otro\cdot valor\end{cases}$$

Donde λ es el número promedio de ocurrencias del evento aleatorio por unidad de tiempo, λ define una familia de distribuciones con una f.p. determinada.

En efecto (*) es una f.p. pues;

1. $p(x, \lambda) > 0$ para $x = 0,1,2,3,\ldots$
2. $\sum_{x=0}^{\infty} p(x;\lambda)=\sum_{x=0}^{\infty}\dfrac{e^{-\lambda}\lambda^{x}}{x!}=e^{-\lambda}\sum_{x=0}^{\infty}\dfrac{\lambda^{x}}{x!}=e^{-\lambda}e^{\lambda}=1$

ACTIVIDAD DE APRENDIZAJE DESARROLLADA

Las siguientes actividades de aprendizaje representan v.a. que tiene distribución aproximada de Poisson.

a)Número de accidentes automovilísticos, accidentes industriales u otro tipo de accidentes en una unidad de tiempo dada.
b)Número de llamadas telefónicas manejadas por un computador en un intervalo de tiempo.
c)Número de solicitudes de seguro procesadas por una compañía en un periodo específico, etc.
d)Se utiliza para analizar problemas de líneas de espera, "confiabilidad o control de calidad.

Definición.- La probabilidad de que una v.a. de Poisson X sea menor o igual a un valor específico de x, se define la función de distribución acumulada (f.d.a):

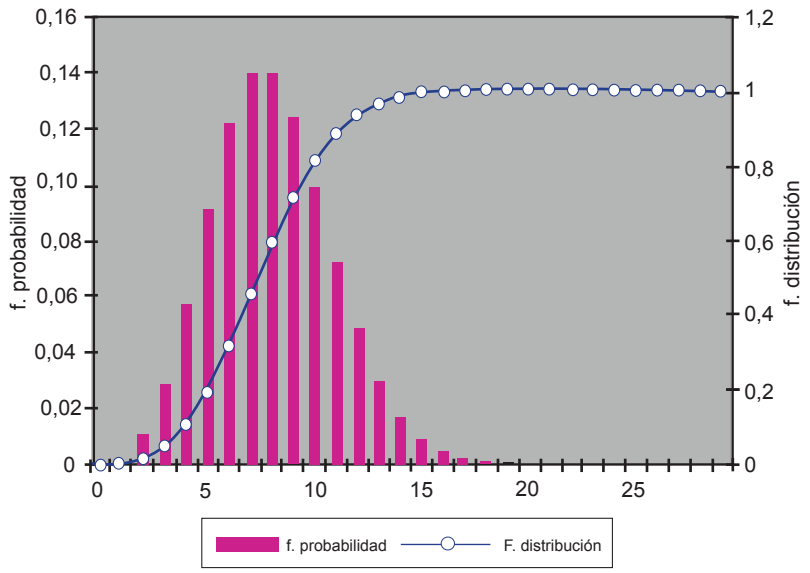
$$F(x,\lambda)=P(X\le x)=\sum_{i=0}^{x}\frac{e^{-\lambda}\lambda^{i}}{i!}$$

Cuyos valores se determinan en las tablas, las mismas que se pueden utilizar para calcular probabilidades individuales en lugar de la ecuación anterior, por:

$$p(x,\lambda) = F(x,\ \lambda) - F(x-1,\ \lambda)$$

La siguiente gráfica representa la función de probabilidad y distribución acumulada de una v.a. de Poisson con $\lambda = 8$.

Distribución de Poisson P(8)



Se pueden probar los valores de la siguiente tabla para una distribución de Poisson

		FACTORES DE FORMA	
MEDIA	VARIANZA	COEFICIENTE DE ASIMETRIA	COEFICIENTE DE CURTOSIS
λ	λ	$1/\lambda^{1/2}$	$3+1/\lambda$

En conclusión, la distribución de Poisson es leptocúrtica con sesgo positivo y se emplea para modelar el número de eventos aleatorios independientes que ocurren a una rapidez constante ya sea sobre el tiempo o el espacio.

Nota. *El siguiente teorema hace referencia a la relación de aproximación de la distribución binomial a la distribución de Poisson, para valores grandes de n y pequeños de p.*

Teorema.- Sea X una v.a. con ley binomial de parametros n y p.

$$p(x,\ n,\ p) = \binom{n}{x} p^x\, q^{n-x} \qquad x=0,1,2,...,n$$

Si para n=1, 2,... la relación $p= \lambda/n$ es cierta para alguna constante $\lambda > 0$ entonces:

$$\lim_{n\to\infty} p(x,n,p) = \frac{e^{-\lambda}\,\lambda^x}{x!} \qquad x= 0,1,2,... \qquad (p\to 0)$$

Demostración

$$\lim_{n\to\infty} p(x,n,p) = \lim_{n\to\infty}\binom{n}{x} p^x(1-p)^{n-x} = \lim_{n\to\infty} \frac{n!}{x!(n-x)!}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{\lambda^x}{x!}\lim_{n\to\infty}\frac{n(n-1)...(n-x+1)}{n^x}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-x}$$

$$= \frac{\lambda^x}{x!}\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n \lim_{n\to\infty}\left[1\left(1-\frac{1}{n}\right)...\left(1-\frac{x-1}{n}\right)\right]$$

$$= \frac{e^{-\lambda}\,\lambda^x}{x!}$$

Pues $\lim_{z\to 0}(1+z)^{1/z} = e \quad con \quad z = -\lambda/n$

ACTIVIDAD DE APRENDIZAJE DESARROLLADA 07

Aplicación de Bioestadística. Para un volumen fijo, el número de células sanguíneas rojas es una v.a. que se presenta con una frecuencia constante. Si el número promedio para un volumen dado es de nueve células para personas normales, determinar la probabilidad de que el número de células rojas para una persona se encuentra dentro de una desviación estándar del valor promedio y a dos desviaciones estándar del promedio.

Solución:

Sea X: "Número de células sanguíneas en un volumen fijo", $\lambda = 9$ por tanto $\mu = 9$ y $\sigma = 3$, $P(\mu-\sigma \leq X \leq \mu+\sigma) = ?$ y $P(\mu-2\sigma \leq X \leq \mu+2\sigma) = ?$

Reemplazando los valores y determinando en Excel tenemos

$P(9-3 \leq X \leq 9+3) = P(6 \leq X \leq 12) = F(12) - F(5) = 0.8758 - 0.1157 = 0.7601$
$P(9-6 \leq X \leq 9+6) = p(3 \leq X \leq 15) = F(15) - F(2) = 0.9779 - 0.0062 = 0.9717$

ACTIVIDAD DE APRENDIZAJE DESARROLLADA 08

Aplicación de Control de Calidad. Una compañía compra cantidades muy grandes de componentes electrónicos, la decisión para aceptar o rechazar un lote de componentes se toma con base en una muestra aleatoria de 100 unidades, si el lote se rechaza al encontrar tres o más unidades defectuosas en la muestra,
¿Cuál es la probabilidad de rechazar un lote si éste contiene 1% de componentes defectuosos?,
¿Cuál es la probabilidad de rechazar un lote que contenga un 8% de unidades defectuosas?

Solución:

Sea X: "Número de unidades defectuosas en el lote, $n = 100$" la muestra es grande, $n>30$. El lote se rechaza si $X \geq 3$.

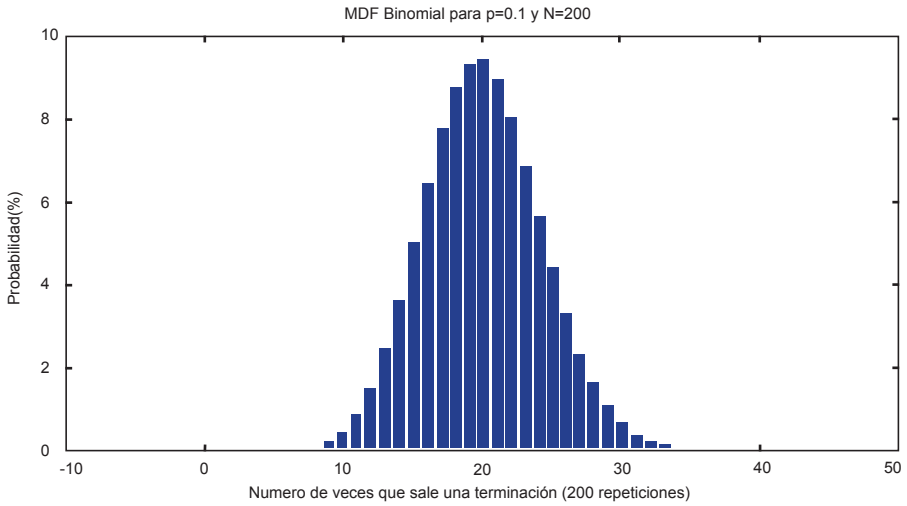
$P(\text{rechazar}) = P(X \geq 3) = ?$ si $p_1 = 1/100$
$P(\text{rechazar}) = P(X \geq 3) = ?$ si $p_2 = 8/100$.

Puesto que las probabilidades de unidades defectuosas son pequeñas aplicamos el teorema que aproxima la distribución binomial con la distribución de Poisson entonces

$\lambda_1 = np_1 = 100(1/100) = 1$ por tanto $P(\text{rechazar}) = 1 - P(X \leq 2) = 1 - 0.9197 = 0.0803$.

$\lambda_2 = np_2 = 100(8/100) = 8$ por tanto $P(\text{rechazar}) = 1 - P(X \leq 2) = 1 - 0.0138 = 0.9862$

Nota histórica.- La función de densidad de probabilidad que proporcionamos a continuación fue descubierta por de Moivre en 1733 como una forma límite de la función de probabilidad binomial; después la estudió Laplacé. También Gauss la cita en un artículo que publicó en 1809, de allí que la distribución normal es frecuentemente llamada distribución gaussiana, en honor de Karl Friedrich Gauss (1777-1855).



Observación: Se puede demostrar fácilmente mediante integrales dobles y utilizando una transformación a coordenadas polares, el siguiente resultado (α).

$$(\alpha)\quad \int_{-\infty}^{+\infty} \exp\left[-\frac{1}{2}\frac{(X-\mu)^2}{\sigma^2}\right] dx = \sigma\sqrt{2\pi}; \qquad -\infty < \mu < +\infty \qquad \sigma > 0$$

Definición.- Una variable aleatoria X se dice que esta normalmente distribuida si su función de densidad de probabilidad está dada por:

$$(1)\quad f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{1}{2}\left(\frac{(X-\mu)}{\sigma}\right)^2\right]$$

$$-\infty < x < +\infty \qquad -\infty < \mu < +\infty \qquad \sigma > 0$$

La ecuación (1) se llama función de densidad normal, o ley Laplace-de Moivre o función de densidad gaussiana.

Se note que tiene dos parámetros μ y σ. Se puede demostrar que la media es μ y la varianza es σ^2, por conveniencia escribiremos $X \sim N(\mu,\sigma)$ como abreviación que X tiene ***distribución normal con media μ y desviación estándar σ.***

Demostremos que efectivamente f(x) dada por (1) es una f.d.p. esto es:

1) $f(x) \geq 0 \quad -\infty < x < +\infty$

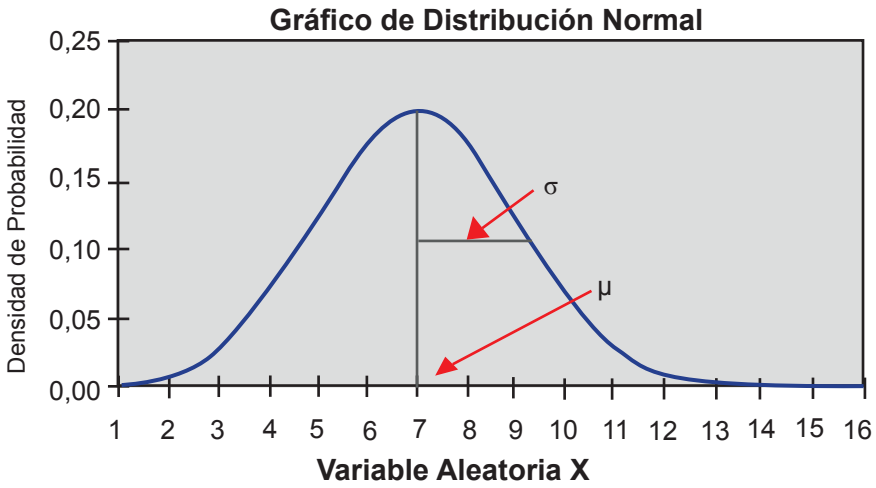
2) $\int_{-\infty}^{+\infty} f(x)dx = 1$

Demostración

1.- Por como está definida f(x) es siempre positiva por x $\epsilon(-\infty,+\infty)$

2.- $\int_{-\infty}^{+\infty} f(x)dx = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{+\infty} \exp\left[-\frac{1}{2}\left(\frac{(X-\mu)}{\sigma}\right)^2\right]dx = \frac{\sqrt{2\pi\sigma}}{\sqrt{2\pi\sigma}} = 1$ por el resultado(α)

Gráficamente la ecuación (1) representa una curva de forma de campana denominada indistintamente curva normal o campana de Gauss, siendo de gran utilidad en Estadística Inferencial. El área bajo la curva es igual a 1 o 100%. La media (μ) se encuentra localizada en el centro dividiendo la curva en dos partes iguales, correspondiéndole a cada una de ellas el 50%.



Curva normal o campana de gaussiana

Los valores de la siguiente tabla se demuestran aplicando la expresión (1) de la definición. Se debe tener en cuenta de la misma los valores de los factores de forma, que nos indican que la curva normal es simétrica ($\alpha_3 = 0$) y mesocúrtica (α_4=3, K = 0).

Tabla (Propiedades básicas de la distribución normal).

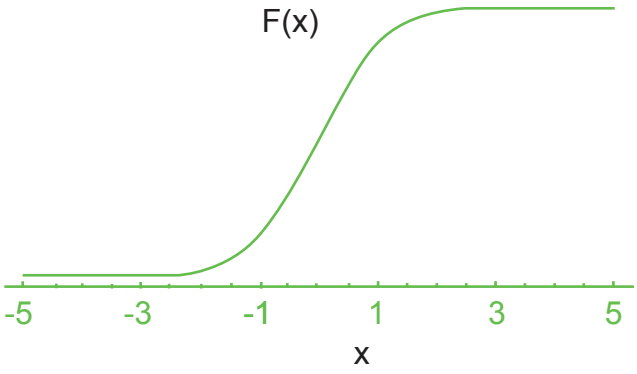
MEDIA	VARIANZA	RECORRIDO INTERCUARTIL	RECORRIDO INTERDECIL	COEFICIENTE DE ASIMETRÍA	COEFICIENTE DE CURTOSIS
μ	σ^2	1.36σ	2.56σ	0	3

Función de distribución acumulada.

La probabilidad de que una v.a. normalmente distribuida X sea menor o igual a un valor específico x está dada por la función de distribución acumulada.

(2) $P(X \leq x) = F(x;\, \mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi\sigma}} \int_{-\infty}^{x} \exp\left[-\frac{1}{2}\left(\frac{(t-\mu)}{\sigma}\right)^2\right]dt$

La curva está dada por una “ojiva” de la forma de una S echada de la siguiente forma, realizada en STATGRAPHICS.



Distribución acumulada normal

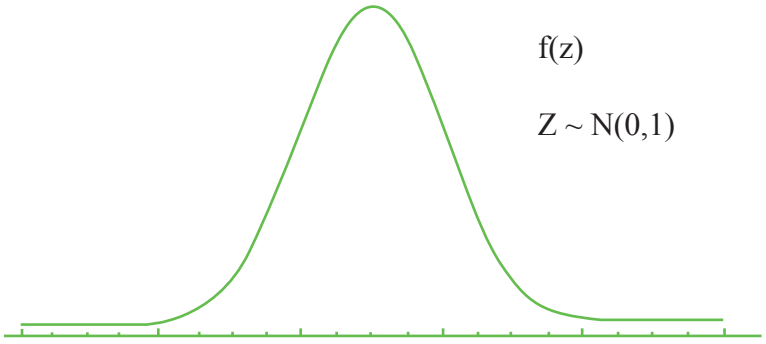
Observación.- La integral (2) no puede evaluarse de forma cerrada; sin embargo, se puede tabular $F(x;\mu,\sigma)$ como una función de μ y σ, lo que necesitaría una tabla para cada par de estos valores. Como existe un número infinito de valores de μ y σ, esta tarea es virtualmente imposible. Afortunadamente, lo anterior puede simplificarse estandarizando la variable aleatoria X, ($X \sim N(\mu,\sigma)$) esto es, sea Z una v.a. definida por la siguiente relación:

(*) $Z = (X-\mu)/\sigma$

Z es una v.a. estandarizada con $E(Z) = 0$ y $Var(Z) = 1$, con función de densidad f(z) definida por (sustituya 0 y 1 por μ y σ respectivamente en la ecuación (1) de la definición de función de densidad normal y cambie X por Z), tenemos

$$(3)\quad f(z) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2} Z^2\right]$$

$$-\infty < z < +\infty$$



Curva normal estándar

La curva es simétrica respecto a la media 0 y el área bajo la curva es nuevamente 1 o 100%.

Observación: Se observe que $X \sim N(\mu,\sigma) \Rightarrow Z \sim N(0,1)$
En palabras se puede decir que: si X se encuentra normalmente distribuida con media μ y desviación estándar σ, entonces $Z = (X-\mu)/\sigma$ también se encuentra normalmente distribuida con media cero y desviación estándar uno.

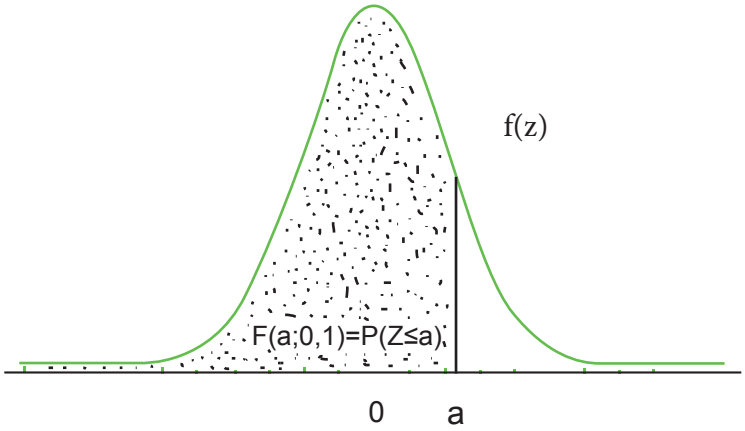
De esta observación se deduce que las distribuciones acumuladas de X y Z son exactamente iguales, es decir que: $F_x(x; \mu, \sigma) = F_z(z;0,1)$.

La expresión de $F_z(z;0,1)$ viene dada por:

$$(4)\quad P(Z \leq z) = F_Z(z;0,1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp(-t^2/2)dt$$

Para z con z ϵ R, los valores de la ecuación (4) vienen dados en tablas fácilmente de leerlos. Ver tabla 1 del apéndice

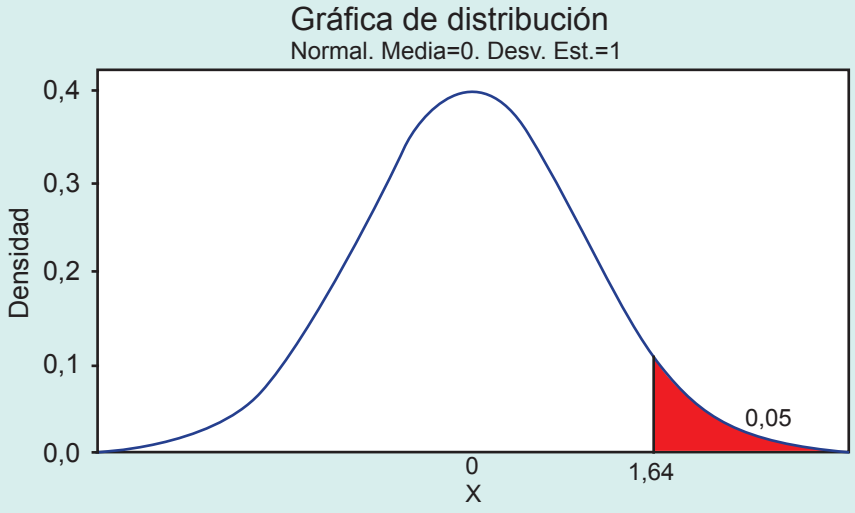
La interpretación geométrica de la ecuación (4) es, el área bajo la curva de f(z) desde $-\infty$ hasta z, considérese el valor z = a, entonces el área de éste valor $F_z(a;0,1)$, se indica en la siguiente figura.



Presentación geométrica de la distribución acumulada (Fa;0,1)

Observación. Para nuestra tabla 1. Distribución normal tenemos que $P(Z_\alpha > a) = \alpha$.

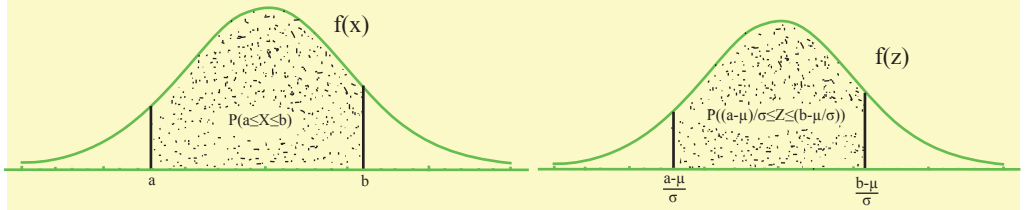
Por ejemplo $P(Z_\alpha > 1.64) = 0.05$, es decir la probabilidad para $Z_\alpha > 1.64$ es igual a 0.05 ó 5% en porcentaje representa el área sombreada de la gráfica de distribución abajo indicada.



Determine la probabilidad de que Z≤a, con la tabla 1 del apéndice de distribución normal estándar dados los siguientes valores de z = a.

1. z = -2.50 P(Z≤-2.50) = P(Z≥2.50) = 0.0062
2. z = -1.96, P(Z≤-1.96) = P(Z≥1.96) =F(1.96;0,1) = 0.0250
3. z = 0, P(Z≤0) = F(0;0,1) = 0.5000
4. z = 1.00, P(Z≤1.00) = 1 - P(Z≥1.00) =1 - F(1.00) = 1 - 0.1587 = 0.8413
5. z = 1.24, P(Z≤1.24) =1.0000 - F(1.24) = 1.0000 - 0.1075 = 0.8925.

Observación. La función de probabilidad (3) es simétrica, es decir, f(-z) = f(z) y se tiene F(-z;0,1) = 1 - F(z;0,1). Muchas tablas vienen dadas únicamente para valores z positivos, como la del apéndice. Nótese que la probabilidad de un valor de la variable aleatoria X se encuentre entre a y b (a < X < b) si X~N(μ,σ); por las propiedades de distribución acumulada y estandarizando la misma se tiene que



a) Si (a-μ)/σ>0 y (b-μ)/σ>0 respecto a Z, entonces
$P(a \leq X \leq b) = F_z[(a-\mu)/\sigma] - Fz[(b-\mu)/\sigma]$

Por ejemplo. Si la v.a. X representa las calificaciones sobre diez puntos de los estudiantes de la asignatura de Econometría de la carrera de Ingeniería en Estadística Informática de la ESPOCH y están distribuidas normalmente con media 7.0 y desviación estándar 1.2 es decir X˜N(7.0,1.2) calcular la probabilidad de que estas calificaciones se encuentren entre 7.1 y 8.5, o sea vamos a calcular $P(7.1 \leq X \leq 8.5)$.

Estandarizamos las calificaciones 7.1 y 8.5 tenemos respectivamente 0.08 y 1.25 por tanto

$P(7.1 \leq X \leq 8.5) = P(0.08 \leq Z \leq 1.25) = F(0.08)-F(1.25) = 0.4681 – 0.1056 = 0.3625$ o 36.25%

b) Si $(a-\mu)/\sigma < 0$ y $(b-\mu)/\sigma>0$ respecto a Z, entonces $P(a \leq X \leq b) = F_z[(b-\mu)/\sigma;0,1] +[1- F_z[-(a-\mu)/\sigma;0,1]]$

Con la tabla del apéndice calcular $P(-1.1\leq Z \leq 2.5) = 1- F(2.5) – F(1.1) = 1-0.0062-0.1357 =0.8581$ o 85.81% .

c) Si $(a-\mu)/\sigma < 0$ y $(b-\mu)/\sigma < 0$ respecto a Z, entonces $P(a \leq X \leq b) = 1- F_z[- (b-\mu)/\sigma;0,1] - [1- F_z[-(a-\mu)/\sigma;0,1]] = F_z[-(b-\mu)/\sigma;0,1] - F_z[-(a-\mu)/\sigma;0,1]$.

Con la tabla del apéndice calcular $P(-2.7\leq Z\leq -1.5) = F(1.5) – F(2.7) = 0.0668-0.0035= 0.0633$ o 6.33%.

Nota. *Se note que para tablas que tienen únicamente valores positivos como la del apéndice la probabilidad $P(a\leq X\leq b)$ se determinaría por:*

$P(a \leq X \leq b) = P[(a-\mu)/\sigma \leq Z \leq (b-\mu)/\sigma]$

ACTIVIDAD DE APRENDIZAJE DESARROLLADA 09

Si X˜N(μ, σ), ¿cuáles son las probabilidades de que el valor de X se encuentre a una, dos y tres veces la desviación estándar de la media?

Solución:

$P(\mu - \sigma \leq X \leq \mu + \sigma) = P[(\mu - \sigma - \mu)/\sigma \leq Z \leq (\mu + \sigma - \mu)/\sigma] = P(-1 \leq Z \leq 1)$
$= F_z(1;0,1) - F_z(-1;0,1) = 2\,F_z(1;0,1) - 1 = 0.6826.$

Con la tabla del apéndice también determinamos

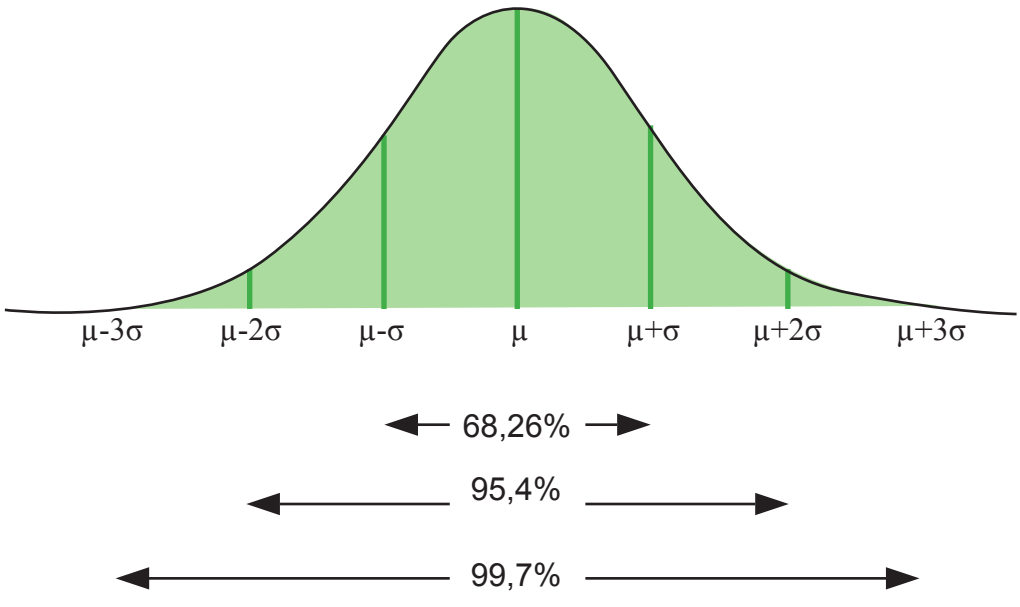
$P(-1 \leq Z \leq 1) = F(-1) - F(1) = 1 - F(1) - F(1) = 1 - 2F(1) = 1 - 2*0.1582 = 1 - 0.3174 = 0.6826.$

Análogamente

$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = P(-2 \leq Z \leq 2) = 1 - 2*F(2) = 1 - 2*0.0228 = 0.9544.$
$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = P(-3 \leq Z \leq 3) = 1 - 2*F(3) = 1 - 2*0.0026 = 0.9974.$

Así para cualquier v.a. normal las probabilidades “una sigma”, “dos sigma” y “tres sigma” son 0.6826, 0.9544 y 0.9974 respectivamente.

Estos resultados indican que en la distribución normal existe una concentración de valores alrededor de la media, como indica la siguiente gráfica.



ACTIVIDAD DE APRENDIZAJE DESARROLLADA 10

Una universidad ecuatoriana espera recibir, para el siguiente año escolar, 16000 solicitudes de ingreso al primer año de ingeniería. Se supone que las calificaciones obtenidas por los aspirantes en la prueba SAT se pueden calcular, de manera adecuada, por una distribución normal con media 950 y desviación estándar 100, si la universidad decide admitir al 25% de todos los aspirantes que obtengan las calificaciones más altas en la prueba SAT.

¿Cuál es la mínima calificación que es necesario obtener en esta prueba, para ser admitido por la universidad?

Solución:

Sean X: "calificaciones por los aspirantes en la prueba SAT", $P(X>x) = 0.25$, es decir, $P(X\leq x) = 1 - P(X>x) = 1 - 0.25 = 0.75 \Rightarrow P(Z\leq z_{0.75}) = 0.75$ donde $z_{0.75} = ?$

Por la tabla de la función de distribución normal del apéndice se calcula : $P(z_{0.2514}=0.67) = 0.2514$ y $P(z_{0.2485}=0.68) = 0.2485$ interpolando estos valores, es decir,

$$z_{0.75} = 0.6745.$$

Luego $0.6745 = (x-950)/100 \Rightarrow x \tilde{=} 1018$, es la calificación mínima para ser admitido por la universidad.

ACTIVIDAD DE APRENDIZAJE DESARROLLADA 11

Un fabricante de escapes para automóviles desea garantizar su producto durante un periodo igual al de la duración del vehículo.

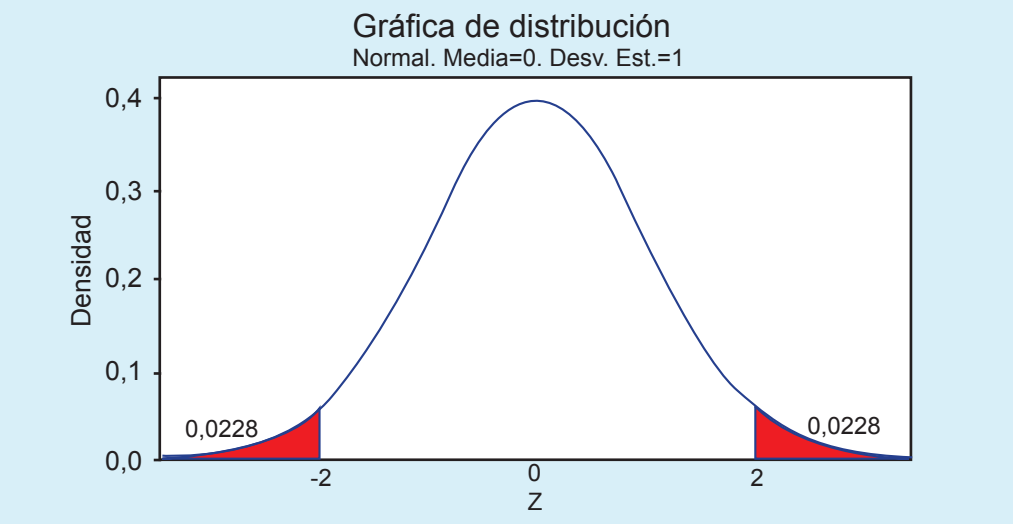
El fabricante supone que el tiempo de duración de su producto es una v.a. con una distribución normal, con una vida promedio de tres años y una desviación estándar de seis meses.

Si el costo de remplazo por unidad es de $10, ¿cuál puede ser el costo total de reemplazo para los primeros dos años, si se instalan 1'000.000 unidades?

Solución:

Sea X: "tiempo de duración del escape de un vehículo".

$$X\tilde{}N(3, 1/2), \quad P(X<2) = ?$$



Por tanto P(X<2) = P(Z<-2) = F(-2) = F(2) = 0.0228 que es la proporción del costo total de reemplazo. Entonces el costo total de reemplazo para los dos primeros años si se instalan 1'000.000 de unidades con el costo de $10 por unidad es (0.0228) *10^7 = $228000.

3.6.3.1 Aproximación de la distribución binomial por la distribución normal estándar.

Teorema de Moivre-Laplacé. Sea X una variable aleatoria binomial con media np y desviación estándar [np(1-p)]1/2. La distribución de la variable aleatoria.

$$Y = \frac{X - np}{\sqrt{np(1-p)}}$$

Tiende a la normal estándar, conforme el número de ensayos independientes tiende a infinito (n $\rightarrow\infty$).

Nota. Este teorema pone de manifiesto que: si X es una variable aleatoria binomial, para la que el número de ensayos independientes es suficientemente grande, se dice que X posee una distribución normal aproximada con media np y desviación estándar [np(1-p)] $^{1/2}$

La aproximación es adecuada tanto como:
np > 5 cuando p ≤ ½ o n(1-p) >5 cuando p >1/2, esto es

$$(*) \quad P(a \leq X_B \leq b) \approx P\left(\frac{a - np}{\sqrt{np(1-p)}} \leq Z_n \leq \frac{b - np}{\sqrt{np(1-p)}}\right)$$

Con Z_n ~ N(0,1)

Se sabe que $P(X_B=x) \neq 0$ pero $P\left(Z = \frac{x - np}{\sqrt{np(1-p)}}\right) = 0$ esto resulta inadecuado, por tanto en lugar de emplear ésta expresión se utilizará:

$$P(X_B = x) \approx P\left(\frac{x - np - 0.5}{\sqrt{np(1-p)}} \leq Z \leq \frac{x - np + 0.5}{\sqrt{np(1-p)}}\right)$$

Que determina la probabilidad de un intervalo de longitud uno (y no del punto) de manera que el punto medio del intervalo sea igual al valor x.

Luego la expresión (*) se escribirá:

$$(**) \quad P(a \leq X_B \leq b) \approx P\left(\frac{a - np - 0.5}{\sqrt{np(1-p)}} \leq Z_n \leq \frac{b - np + 0.5}{\sqrt{np(1-p)}}\right)$$

ACTIVIDAD DE APRENDIZAJE DESARROLLADA 12

Un periódico llevó a cabo una encuesta entre 400 bachilleres seleccionados aleatoriamente, en una provincia, sobre pruebas de ingreso a la universidad. De los 400 bachilleres, 220 se pronunciaron a favor de las pruebas.

a) ¿Qué tan probable resulta el hecho de tener 220 o más a favor de las pruebas de ingreso, si la población graduada de esta provincia se encuentra dividida en opinión de igual manera?

b) Supóngase que se encuesta a 2000 personas teniendo la misma proporción de éstas a favor de las pruebas, que la del inciso anterior. ¿Cómo cambiaría su respuesta respecto al inciso a)?

Solución:

a) Sean n = 400 p = ½ entonces np = 200, np(1-p) = 100 y
X: "Número de bachilleres a favor de pruebas de ingreso a la universidad" entonces lo que queremos calcular es P(X $\geq$ 220) = ? Por el teorema de Moivre Laplace tenemos

$$P(X \geq 220) \cong P(Z \geq \frac{220\text{-}200\text{-}1/2}{\sqrt{100}})$$
$$= P(X \geq 1.95) = 0.0256$$

b)Aquí n = 2000, p = ½, entonces np = 1000, np(1-p) = 500. Aplicando la regla de tres en a) tenemos

400 220
100 x = 55% se ha tomado sobre una población de 400.

Ahora

100 55
2000 x = 1100 personas.

Luego

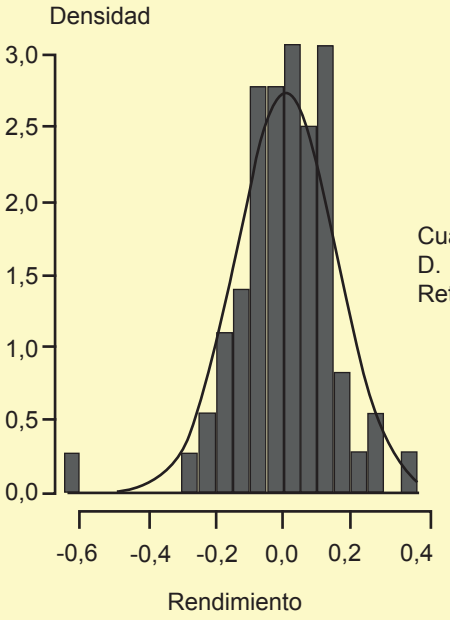
$$P(X \geq 1100) \cong P(Z \geq \frac{1100\text{-}1000\text{-}1/2}{\sqrt{500}})$$

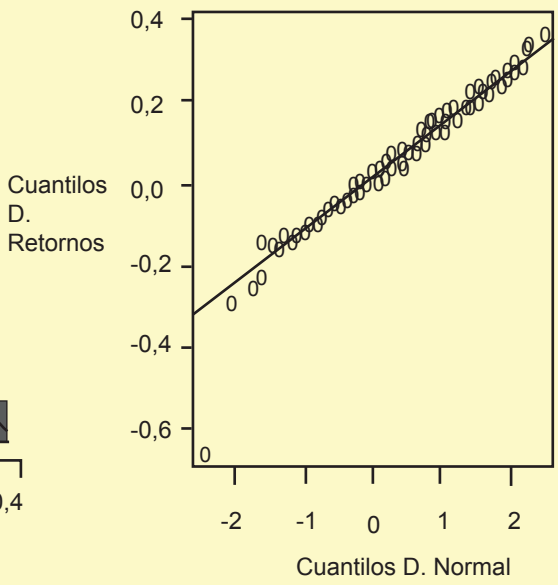
$$= P(Z \geq \frac{99.5}{\sqrt{500}} = P(Z \geq 4.4) = 0$$

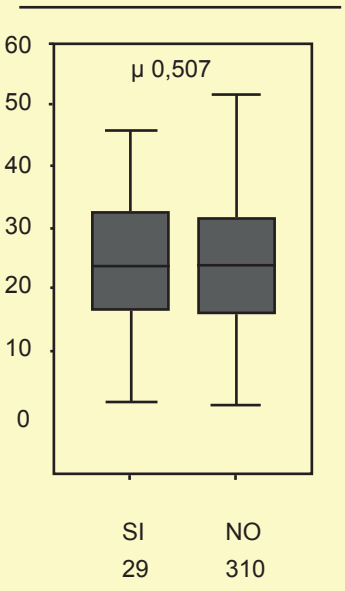
Al tener una probabilidad cero resulta ser un evento imposible.

Observación. La hipótesis de normalidad puede ser convalidada con presentaciones semejantes a un histograma diagrama de probabilidad normal (Q-Q plot). También, las presentaciones gráficas pueden dar al investigador (educativo) una impresión de la variabilidad de la distribución. El uso gráfico nos ayuda a extraer información acerca de propiedades de un conjunto de datos.

Por ejemplo el diagrama de caja y bigotes o box and wisher plot proporciona al investigador la simetría para una distribución normal y otras propiedades de los datos como localizar puntos atípicos, ver los cuartiles entre ellos podemos detectar la mediana (segundo cuartil).







ESTADÍSTICA APLICADA A LA EDUCACIÓN CON ACTIVIDADES DE APRENDIZAJE

En los programas ministeriales de matemática se establecen argumentos de Estadística y Probabilidades, sin embargo no se estudian adecuadamente o no se estudian tales temas. Se quiere estimular a una mejor enseñanza de los conceptos básicos de dichos fundamentos. No se quiere dar un recetario de formulas con el único propósito de acatar cumplimientos a un programa establecido, al contrario se pretende presentar este texto de manera atractiva, interesante y aplicativa. El texto Estadística aplicada a la educación con actividades de aprendizaje para su desarrollo se ha dividido en cuatro capítulos tomando en cuenta los aspectos de generalidades, descripción, herramienta y conclusiones bajo los nombres de Generalidades, Estadística Descriptiva, Teórica de las Probabilidades e inferencias Estadística. Al final de los capítulos 2 al 4, se puede realizar ejercicios aplicativos a la teoría vista, a lo que llamamos actividades de aprendizaje utilizando paquetes estadísticos como el MINITAB, SPSS entre otros y se pueden también realizar en la hoja de calculo EXCEL. En el tercero exponemos la parte teórica-practica de las Probabilidades requerida en el cuatro de inferencia Estadística.

Jorge Washington Congacha Aushay

Jorge W. Congacha A. Doctor en Matemática, graduado en la Escuela Superior Politécnica de Chimborazo, ESPOCH- ECUADOR. Estudió STATISTICA MATEMÁTICA y ANALISI SUPERIORE en Dipartimento di Matemática dell'Universita di Pavía - Italia. Especialista en Computación Aplicada al Ejercicio Docente. Estudió la maestría en Docencia Universitaria e Investigación Educativa en la Universidad Nacional de Loja, LOJA-ECUADOR. Docente de Econometría y Estadística Inferencial en la carrera de Ingeniería en Estadística -Informática de la Escuela de Física y Matemática de la FACULTAD DE CIENCIAS-ESPOCH, Director del Grupo de Investigación ESTADISMATICA en "MODELIZACION ESTADISTICA-INFORMATICA".

La esencia de la vida es la improbabilidad estadística a escala colosal.

Richard Dawkins

Jorge Washington Conchaga Aushay

Estadística aplicada a la educación