UNIVERSIDAD NACIONAL DEL CALLAO

FACULTAD DE INDUSTRIAL Y DE SISTEMAS

UNIDAD DE INVESTIGACIÓN DE LA FACULTAD DE INDUSTRIAL Y DE SISTEMAS

INFORME FINAL DE TEXTO

"ESTADÍSTICA DESCRIPTIVA CON EXCEL Y SPSS"

AUTOR: Dr. MAS AZAHUANCHE, GUILLERMO ANTONIO

Período de ejecución: Del 01 de marzo del 2017 al 28 de Febrero del 2019

Resolución Rectoral N° 274-2017-R del 24 de Marzo del 2017

Resolución N° 042-2017-UIFIIS-UNAC

Callao marzo de 2019

I Contenido

DEDICATORIA

Dedico este humilde trabajo a Dios que da sabiduría y es mi apoyo en todo momento.

Dedico a mi esposa María Ysabel y a mis hijos Jorge, Luis y Carlos que dan aliento para avanzar en mis proyectos

AGRADECIMIENTOS

Agradezco al Dr. Félix Romero Revilla por darme los consejos y apoyo académicos y darme la oportunidad de dictar el curso de Estadística y de asesorar a los egresados para validar sus tesis para su titulación.

Agradezco a la Universidad Nacional del Callao por darme la oportunidad de escribir este texto que ha sido una experiencia muy grata en el quehacer académico.

II ÍNDICE

TABLAS

II ÍNDICE

FIGURAS

II PRÓLOGO

El presente texto de Estadística Descriptiva con Excel y SPSS, busca brindar los conocimientos en los temas y conceptos matemáticos y estadísticos que se usan en ciencias e ingeniería Industrial y de Sistemas, que servirán de base en la investigación inferencial de sus trabajos de tesis y/o trabajos de investigación de campo su especialidad de nuestros estudiantes.

El texto de Estadística Descriptiva con Excel y SPSS, ayuda a comprender y aplicar los conceptos básicos de matemática y estadística usando Excel y SPSS de forma práctica y presenta una colección de problemas y ejercicios que pueden ser utilizados como un manual de ejercicios para la asignatura de Estadística destinado fundamentalmente para los estudiantes de los primeros ciclos de las universidades. También puede ser útil para estudiantes de cursos superiores y los que estén elaborando su tesis. Los profesores pueden emplear este libro al preparar las clases prácticas.

El texto de Estadística Descriptiva con Excel y SPSS es un complemento para reforzar, facilitar y ampliar los conocimientos de Estadística de los estudiantes de toda carrera universitaria en especial a las ingenierías y de esta manera les permitirán elaborar sus tesis para sus títulos universitarios que la ley universitaria exige.

Agradecemos a nuestros lectores que nos hagan llegar sus aportes críticos y comentarios y algunas sugerencias que nos permitirá continuar mejorando el contenido del texto "Estadística Descriptiva con Excel y SPSS".

III INTRODUCCIÓN

Nuestro principal objetivo es la de comprender los fundamentos de teóricos y prácticos de la estadística con software educativo Excel y SPSS que es de gran ayuda al ejecutar los ejercicios y/o problemas de aplicación.

El texto tiene un alcance para los profesionales y estudiantes de las diferentes especialidades, especialmente para los alumnos que están haciendo sus tesis. Que requieren un análisis detallado sobre la estadística descriptiva para entender el fundamento de las aplicaciones que tienen desarrollar en sus investigaciones.

Este libro estar orientado en el terreno de la transcripción al estudiante de cada una de las esferas del amplio espectro de la función estadística, con casos y ejemplos Este texto ayuda a los profesionales y estudiantes a ver la conexión directa entre la teoría y la practica en el que campean los métodos y procedimientos, las fórmulas matemáticas y las gráficas, que están aplicando en el acervo estadístico.

Estadística es una asignatura que todo estudiante a nivel superior debe manejar con soltura para poder realizar sus investigación en donde se tiene que recoger información procesarla y saber que está pasando en su investigación para poder dar soluciones, este camino de la precisión estadística, capaz de captar la realidad circundante, con supresión de la habitual diferencia de ritmo entre el dato estadístico y el nivel de desarrollo demográfico, económico, social y tecnológico de la cambiante realidad de nuestro país y de su problemática.

IV CUERPO DEL TEXTO O CONTENIDO

CAPITULO I:

Nociones Básicas, Recolección y representación de los Datos

1.1. Revisión de conceptos matemáticos.

A continuación presentamos los diferente conceptos matemáticos que usamos en nuestras investigaciones al aplicar la estadística como herramienta.

Axioma o postulado: Proposición que se acepta como verdadera sin ninguna demostración.

Concepto: Acción de representar intelectualmente un objeto en forma abstracta. Acto o producto de la concepción intelectual o intelección.

Conclusión: Proposición final de un razonamiento obtenido por inducción o por deducción de las premisas o antecedentes.

Conocimiento: Conocimiento es la habilidad para transformar información en acción efectiva y exitosa. La información son datos estructurados y procesados.

Conocimiento clave: El conocimiento clave está compuesto por la unión del conocimiento colectivo de la organización y el conocimiento individual de sus miembros en el cual se basa la individualidad y la competitividad de la empresa. El conocimiento clave tecnológico es la base para los productos principales de la empresa y su potencial de desarrollo. El conocimiento clave también incluye competencias de la organización como el conocimiento específico del mercado, las experiencias de los clientes, etc.

Conocimiento tácito y Conocimiento explícito: Conocimiento explícito es aquél que está codificado, registrado, tal y como son los manuales, normas, procedimientos; como está en un soporte físico - o electrónico- es fácil de transmitir. Conocimiento tácito son aquellos saberes o habilidades residentes en las personas, el conocimiento tácito es más difícil de transmitir. El conocimiento tácito reside en las personas, el conocimiento tácito se hace explícito durante la interacción social.

Corolario: Es aquel teorema que se deduce fácilmente de otro precedente.

Criterio: Regla o norma para saber lo que es verdadero o puede tomarse como cierto. La teoría del conocimiento o epistemología busca un criterio para fundamentar nuestras certezas espontáneas y apoyar en él la objetividad de nuestro conocimiento. Existen también criterios para la rectitud del obrar moral.

Criterios de desempeño: Parte constitutiva de una norma de competencia laboral que hace referencia a aquellos aspectos que definen el resultado del desempeño competente, es decir, definen las condiciones con las que el elemento de competencia debe ser desempeñado. Los criterios de desempeño se asocian a los elementos de competencia.

Cuadros medios: Personal que trabaja en los niveles inferiores e intermedios de la dirección, en los sectores comerciales, industriales o administrativos.

Cuadros superiores: Personal que trabaja en los niveles superiores de la dirección, en los sectores industriales, comerciales o administrativos.

Deducción: Nexo lógico por el que una conclusión resulta de la comparación de dos o más premisas. La deducción procede de lo universal a lo particular o menos general. Su expresión se llama silogismo.

Definición: Es aquella proposición relativa a una descripción o convención. Manifestar lo que una cosa es. Deslindar o señalar los límites conceptuales de algo. Existen definiciones etimológicas, descriptivas, genéticas; la más perfecta es, sin embargo, la esencial, en la que se expresa el género próximo y la diferencia específica. No todos los objetos son, sin embargo, susceptibles de definición esencial.

Demostración: Conjunto de deducciones obtenidas a través un razonamiento lógico

Digitalización: Es la forma de representación de datos numéricos o alfabéticos al convertirlos a números binarios para transmitirlos en bits, lo que permite un mayor control y precisión en el manejo de datos. Mediante esta tecnología, la información puede adoptar múltiples formas como: multimedia, hipertexto, bases relacionales de imágenes o sistemas cliente servidor.

Data Warehouse: El principio consiste en instalar un sistema de base de datos que sea independiente de los sistemas de bases de datos operativos, en los que los datos de los niveles de toma de decisiones de la empresa se mantienen a parte. En una data warehouse, los datos necesarios no sólo son almacenados, sino que ya han sido parcialmente evaluados.

Empirismo: Escuela filosófica que no admite otro criterio de verdad que la experiencia sensible. Rechaza la suposición de ideas innatas, y también la intelección como penetración en las cosas sensibles hasta obtener de ellas su concepto o esencia. Reduce así los conceptos, de forma parecida al nominalismo, a meros nombres o términos designativos de colectividades agrupadas mentalmente.

Enseñanza con la ayuda del computador: Método de enseñanza aprendizaje que consiste en utilizar un computador cuyos terminales son puestos a disposición de los estudiantes para suministrarles los elementos y las secuencias de un curso previamente programado. El empleo del computador puede constituir uno de los elementos de un sistema multimedios.

Escolio: Es aquella advertencia o anotación que se formula con la finalidad de aclarar, ampliar o restringir proposiciones anteriores.

Estrategias: Es un elemento fundamental de la planificación que se apoya sobre dos aspectos básicos: la definición de una imagen prospectiva de la estructura y el funcionamiento del sistema económico social y la determinación de la trayectoria, o sea, de las acciones o proyectos estratégicos en un encadenamiento temporal de secuencia, considerando la viabilidad técnica, económica, sociopolítica de cada etapa del proceso de desarrollo e incluyendo las medidas básicas que permitirán realizar efectivamente dicha trayectoria.

Experiencia: Vivencia personal de una situación repetida. Posee experiencia quien ha conocido una realidad existencial, no sólo teóricamente. Experiencia sensible: captación de lo real a través de las facultades sensitivas de conocimiento. La escuela empirista hace de la experiencia sensible la única fuente válida de conocimiento.

Fáctico: Condición de existente o de ser "de hecho" (de facto), a diferencia de los entes de razón, los meramente posibles o los en potencia.

Fenotipo: Conjunto de caracteres que se manifiestan visiblemente en un individuo y que expresan la interacción de un genotipo con su medio.

Habilidad: Destreza y precisión necesaria para ejecutar las tareas propias de una ocupación, de acuerdo con el grado de exactitud requerido.

Hipermedia: Combinación de texto y multimedia. Actualmente es un recurso ampliamente explotado en el World Wide Web.

Hipertexto: Documentos que contienen vínculos con otros documentos, al seleccionar un vínculo automáticamente se despliega el segundo documento. Vea: HTML, HTTP, Página Web, World Wide Web.

Homepage: (Página inicial). Es la página web de entrada a un lugar del World Wide Web. Es considerada la página principal.

Host: (Anfitrión) Computadora a la que tenemos acceso de diversas formas (telnet, FTP, World Wide Web, etc). Es el servidor que nos provee de la información que requerimos para realizar algún procedimiento desde una aplicación cliente.

HTML: Lenguaje de marcado de hipertexto, (Hiper-Text Markup Languaje) es el lenguaje con que se escriben los documentos en el World Wide Web. A la fecha existen tras versiones de HTML. HTML 1, donde se sientan las bases para la disposición del texto y las gráficas, HTML 2 donde se agregan formas y HTML 3 (llamado también extensiones Netscape) donde se añaden tablas, mapas, etc.

HTTP: Protocolo de Transferencia de Hipertextos (Hiper-Text Transfer Protocol). Es el protocolo usado por el Word Wide Web para transmitir páginas HTML.

Indización: Proceso de descripción e identificación de documentos en función del contenido temático. Durante el indizado, se extraen documentos mediante un proceso de análisis, luego, se transcriben en los elementos de los instrumentos de indizado, tales como tesauros, planes de clasificación, etc.

Ingeniero: Especialista que posee conocimientos y competencias en uno o varios campos científicos, tecnológicos o técnicos, destinados generalmente a funciones de alta responsabilidad. La formación de ingeniero se adquiere en facultades superiores universitarias o en establecimientos especializados de enseñanza superior. Se distinguen, generalmente, según el tipo de formación recibida, dos categorías principales: los ingenieros de concepción o de organización, y los ingenieros de producción o de ejecución.

Innovación: Selección creadora, organización y utilización de recursos humanos y materiales bajo una forma nueva y original conducente a una mejor consecución de los fines y objetivos definidos. El esfuerzo de innovación debe ser continuado, con visitas a una utilización óptima de las facilidades y a la consecución del modelo ideal propuesto. Gracias a la innovación un sistema educativo puede acelerar su evolución.

Lema: Es el teorema preliminar que sirve de base para demostrar un teorema principal.

Link: Enlace entre páginas en la Web. Son sectores de la página (texto o imágenes) que están vinculados a otras páginas, de manera que basta con indicar en ellos para trasladarse a la nueva página, que puede estar ubicada en cualquier servidor de la red.

Listas de interés (mailing lists): Servicio de Internet, que permite intercambiar mensajes de correo electrónico sobre un tema determinado, con un grupo definido de personas.

Login: Identificación o nombre electrónico de un usuario de correo electrónico. Equivale al nombre de la casilla (cuenta) que ese usuario tiene en el servidor de correo electrónico.

Multimedia: Sistema que integra texto, imágenes fijas o en movimiento y sonido en un único soporte. También, combinación de varios medios como sonido, gráficos, animación y video.

Objetividad: Es la relación lógica entre el contenido de una evaluación con la administración, calificación e interpretación de sus resultados. Todo lo cual debe ser independiente del juicio subjetivo de quien las elabora, las aplica y las evalúa; por tanto, estas consideraciones conducen a resultados más objetivos por cuanto hay igualdad en el tratamiento de todas las personas e identidad en los criterios de evaluación de empleados.

Ordinalidad. Utilización de los números para ordenar.

Password (palabra o clave de acceso): Código conocido sólo por el usuario de correo electrónico, y que se utiliza para proteger la privacidad de los mensajes.

Período de práctica: Designa el período, generalmente obligatorio, durante el cual las competencias adquiridas durante la formación práctica impartida por la educación técnica y profesional pueden ser experimentadas y puestas a punto en las empresas.

Proposiciones matemáticas: Son el conjunto de palabras que afirman o niegan propiedades y relaciones de entes (seres) matemáticos. Las principales proposiciones son: definición, axioma o postulado, teorema, corolario, lema.

Proporcionalidad. Correspondencia que existe entre magnitudes proporcionales.

Protocolo: Conjunto de normas que definen cómo se realiza el intercambio de datos entre computadores o programas computacionales, organizando el desplazamiento de la información a través de la red e indicando cuál es el origen de los datos, el camino que deben recorrer y el destino final.

Publicar en la Web: Instalar en un servidor Web, páginas con información para que sean accesibles a cualquier persona conectada a Internet.

Razón. Cociente entre dos cantidades. Se llama razón de una magnitud A respecto a una magnitud B al número que expresa la medida de A cuando se toma B por unidad.

Razonabilidad de un resultado. Coherencia del resultado de una operación o de la solución de un problema con el contexto en el cual está planteado. Ejemplo: El valor de una suma no debe ser mayor que un cierto número, la solución de un problema debe estar comprendida entre determinados valores.

Responsable de la formación: Persona empleada por una empresa o un grupo de empresas y cuyas principales funciones son planificar, organizar y controlar las actividades de formación que se desarrollan en la empresa, así como aquellas que se efectúan en el exterior.

Sistemas de gestión de documentos: Los sistemas de gestión de documentos comprenden la utilización de documentos electrónicos, su archivo conforme a estructuras definidas, la realización de un índice con palabras clave, la búsqueda de documentos así como hacer que los documentos estén disponibles para consultas o procesamientos posteriores.

Semiótica: Ciencia de los modos de producción, de funcionamiento y recepción de los diferentes signos de comunicación en los individuos o colectividades.

Servicios en Internet: Distintas actividades que se pueden realizar en Internet, como enviar y recibir correo electrónico, conversar con otra persona, publicar información, entre otros.

Servidor de correo electrónico: Computador que contiene y administra las casillas de correo electrónico de varios usuarios, y recibe y envía los mensajes a través de la red.

Servicios científicos y tecnológicos: Son todas aquellas actividades relacionadas con I+D que contribuyen a la generación, la difusión y la aplicación de los conocimientos de CyT.

Síntesis: Método de demostración que procede de lo simple a lo compuesto, de los elementos al todo; de las causas a los efectos, del principio a las consecuencias. Es de sentido inverso al análisis, y lo complementa para el estudio de un objeto.

Sistema de información: Conjunto organizado de acciones destinadas a lograr la transferencia de información.

Sitio Web: Conjunto de páginas Web que conforman una unidad entre sí.

Software: Conjunto de programas, documentos, procedimientos y rutinas asociados con la operación de un computador. Es la parte intangible del computador. Dentro del software destaca el sistema operativo, las aplicaciones y los documentos.

Tanteo. Estrategia que permite determinar la solución de una ecuación por medio del método de ensayo y error.

Teorema: Es aquella proposición que por no ser evidente necesita demostración. Consta de tres partes: hipótesis, tesis, demostración.

Tesis: Es la proposición que se quiere demostrar.

Transferencia de información: Conjunto de operaciones sucesivas mediante las cuales se pone el conocimiento a disposición de las diferentes categorías de usuarios después que haya sido producido. Incluye la producción, esto es, el registro de información primaria, secundaria y terciaria, la producción de los correspondientes documentos, su distribución, almacenamiento, tratamiento, difusión, búsqueda, acceso y explotación, a través de todos los canales posibles. Esta noción abarca la totalidad del ciclo, incluidos la producción y el empleo de información, y es, por consiguiente, más amplia que el manejo de la información.

Transferibilidad: Propiedad de las capacidades y de los conocimientos que le permiten al individuo la posibilidad de aplicarlos a ocupaciones o entornos técnicos diferentes.

Unidades de entrada: Leen datos e instrucciones procedentes de tarjetas, cinta magnética, cinta de papel, documentos codificados en caracteres magnéticos

(MICR), discos, terminales de teleproceso, etc. La unidad de entrada convierte datos e instrucciones leídas en impulsos eléctricos que magnetizan, en un sentido u otro, los núcleos del almacenamiento principal.

Unidades de producción: Talleres anexos a los establecimientos de formación y perfeccionamiento de los profesores de educación técnica y profesional, en los cuales los profesores y alumnos pueden aplicar y mejorar sus competencias prácticas, fabricando ellos mismos equipos y material pedagógico.

Unidades de salida: Gracias a las unidades de salida, los resultados pueden registrarse en tarjetas, cintas de papel, cinta magnética, copias impresoras, discos, terminales de teleprocesos, etc.

Fuente: file:///C:/2010/Temas%20Para%20CD/TIC/glosario%25202.pdf

1.2. Definición y Funciones de la Estadística

Estadística: Es una parte de la Ciencia de los Datos. Proporciona los procedimientos/técnicas para separar la "ciencia" de la "opinión" utilizando un modelo teórico del experimento. Este modelo es sometido a las leyes de la probabilidad para observar las posibilidades de los resultados que se originan únicamente por el azar y no hay otro factor que determina el resultado del experimento.

El término estadística, proviene de la palabra Estado. Surge cuando de hace necesario para el Estado, cuantificar datos como: características de las personas, de infraestructura, de recursos disponibles, etc. En general el estado realiza estos cálculos con fines de poder efectuar estudios económicos o militares.

La estadística es la herramienta que se utiliza en la mayoría de las investigaciones científicas en donde se hace un tratamiento de los datos e interpretamos los resultados y si es posible elaboramos una predicción de los mismos.

La importancia de la Estadística radica en proporcionar alternativas cuantitativas al mero juicio personal, de tal forma que realicemos conclusiones valederas y objetivas. Las pruebas con un nivel de significación se consideran como una herramienta muy importante en la toma de decisiones subjetivas, mediante sus resultados podemos tomar decisiones, que, en un principio, no dependen del investigador que las realiza.

Sabemos que el p-valor que arroja una prueba estadística es una medida directa de verdad que resulta de evaluar una muestra actual y ver si es cierta la hipótesis nula, o sea la hipótesis que se desea contrastar. Además, sabemos que el p-valor de emplea para indicar cuanto contradice la muestra actual la hipótesis alternativa.

Las implicancias de un mal uso del p-valor, no solo afectan la calidad de los trabajos científicos publicados, sino también influyen en tener una concepción equivocada de lo que se investiga.

División de la Estadística

Estadística clásica. Enfoque objetivo.

- Estadística descriptiva: Cuyo objetivo principal es la de recolectar un conjunto de datos, organizarlos mediante tablas, y representarlos de acuerdo a las variables en gráficos de barras, gráficos de círculos, ver sus asimetrías, valores extremos, datos inconsistentes, limpieza de datos, es decir describe y analiza una población sin pretender generalizar los resultados.

Estos métodos pueden ser gráficos o pueden incluir análisis mediante cálculos.

Puede reordenar y agrupar los datos de varias formas, de modo que sea posible tener de un golpe de vista una imagen global de los datos (Distribución de frecuencias).

Puede construir tablas, gráficas y figuras que permitan visualizar los resultados (Técnicas de representación gráfica).

Puede convertir los datos originales en formas que sean más útiles para fines específicos (Porcentajes, percentiles, distribución normal).

Puede calcular los promedios (Medidas de tendencia central).
Tomando el promedio como referencia, puede describir la dispersión de los datos respecto a ese punto central (Medidas de variabilidad o de dispersión).
Se puede obtener una relación entre dos variables (Coeficiente de correlación y regresión).

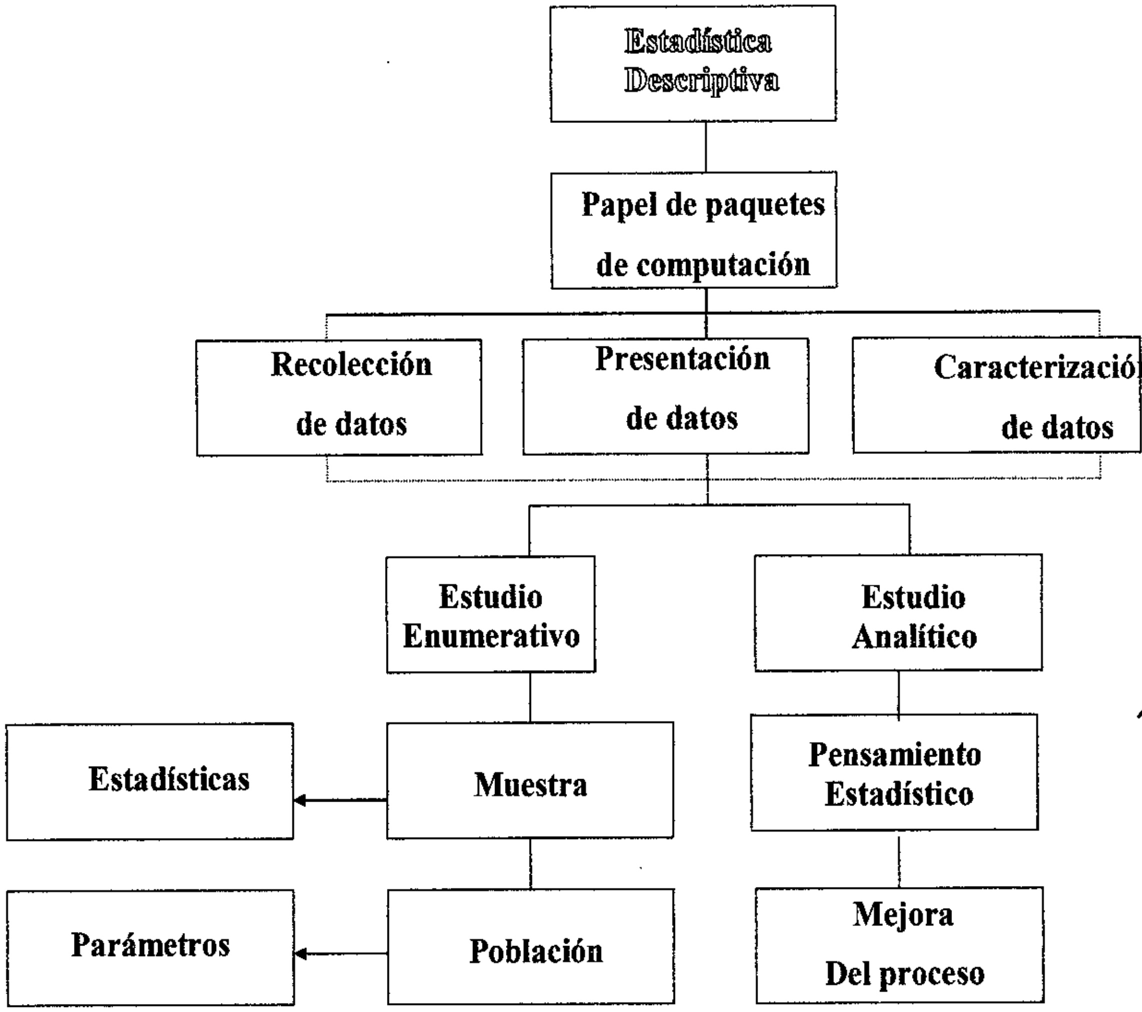
- Inferencia Estadística: Es el estudio de una población mediante una muestra representativa extraída de esa población a partir de una información parcial (muestra) y las condiciones que rigen su validez, basadas en un juicio de los administradores. Cuando la muestra es representativa significa que en ella se dan todas las características de esa población para ello esta muestra debe ser aleatoria con un tamaño apropiado para ello debemos tener un tamaño de muestra, los intervalos de confianza, hacer la prueba de hipótesis con un nivel de confianza, y un nivel de significación.

  Debido a que esas decisiones se toman en condiciones de incertidumbre, se requiere el uso de concepto de probabilidad. La inferencia puede contener conclusiones que pueden no ser ciertas en forma absoluta, por lo que es necesario que estas sean dadas con una medida de confiabilidad que es la probabilidad.

  El investigador que experimenta debe decidir si los resultados son producto de los tratamientos aplicados, o si es resultado del azar.

> La inferencia estadística es el conjunto de métodos y técnicas que permiten inducir, a partir de la información empírica proporcionada por una muestra, cual es el comportamiento de una determinada población con un riesgo de error medible en términos de probabilidad.

Fuente: http://www.ub.edu/aplica_infor/spss/cap4-1.htm



Extractado del libro Estadística Básica en Administración – Berenson, M. Levine, D.

1.3. Términos utilizados en Estadística.

Confiabilidad: Cuando se aplica un instrumento en una investigación las veces que se aplique debe dar el mismo resultado y/o parecido (es decir es la ausencia de error el instrumento). Cuando un instrumento está elaborado en escala de ordinal se usa el estadístico Alfa de Cronbach y cundo está en escala nominal se usa KR-20.

Datos: Son las observaciones recolectadas (como mediciones, géneros, respuestas de encuestas). Un número por sí mismo no es "de datos". Pero cuando un número se utiliza para representar algo real, se considera "datos". Hay dos tipos de datos que vamos a estar buscando en:

i. Datos categóricos: son datos que no necesitan se pueden clasificar en varios grupos o categorías. Por ejemplo: los colores preferidos, cargos, nombres de personas en la clase. Los datos dicotómicos son aquellos que provienen de dos categorías ejemplo la variable género.
ii. Datos cuantitativos: son valores numéricos. Por ejemplo: Altura, el salario, la edad.

Datos cuantitativos: son valores numéricos. Por ejemplo: Altura, el salario, la edad.

Desarrollo experimental: Consiste en trabajos sistemáticos basados en los conocimientos existentes, derivados de la investigación y/o la experiencia práctica, dirigido a la producción de nuevos materiales, productos y dispositivos; al establecimiento de nuevos procesos, sistemas y servicios o a la mejora substancial de los ya existentes. Es decir producir una tecnología.

Diseño: No experimental (no se manipula las variables) y Transversal (se aplica el cuestionario en un solo momento), su propósito es describir variables y analizar su incidencia e interrelación en un momento dado.

Estadística: Es la ciencia que estudia una parte de la Ciencia de los Datos en donde se usa los métodos y procedimientos de recopilación, tabulación, análisis e interpretación de datos para tomar decisiones y predecir situaciones que puedan expresarse de manera cuantitativa. Se divide en dos partes importantes: Estadística Descriptiva y Estadística Inferencial

Estadística No Paramétrica: Es la que tiene las siguientes características (la variable no tiene distribución conocida y su distribución es distinta a la normal).
Estadística Bayesiana: Es cuando su enfoque es probabilístico y subjetivo. El investigador usa el conocimiento del cálculo de las probabilidades.

Experimentación: Utilización de medios técnicos para analizar la producción de fenómenos y comprobar las hipótesis científicas.

Evaluar: Evaluar es medir y comparar la medida con patrón

Frecuencia o frecuencia absoluta: Es el número de veces que se presenta un dato o el valor de una variable.

Ejemplo: Dado el siguiente conjunto de datos que representan puntajes de una prueba de rendimiento matemático calificado a 80 puntos de 120 estudiantes

Tabla N°1 *Tabla de frecuencias del rendimiento matemático de 120 estudiantes*

IC	fi	hi
[11-20)	13	0.10833333
[20-29)	21	0.175
[29-38)	21	0.175
[38-47)	15	0.125
[47-56)	14	0.11666667
[56-65)	14	0.11666667
[65-74)	14	0.11666667
[74-83)	8	0.06666667
Total	120	1

Frecuencia relativa: Cociente entre la frecuencia absoluta del valor de una variable y el número total de datos. Ejemplo en la tabla 1, hi donde $h_i = \frac{f_i}{n}$ es la frecuencia relativa y n es el tamaño de la muestra.

Frecuencia de los acontecimientos: Es el porcentaje de veces que un evento específico se repite en un experimento aleatorio, cuando éste se realiza un número finito de veces. En el límite, cuando el número de veces que se repite el experimento aleatorio es tan grande como se quiera, frecuencia y probabilidad coinciden. Por ejemplo, al arrojar un dado N veces, cada una de las caras aparece con la frecuencia 1/6, a no ser que el dado esté cargado. La probabilidad de que aparezca una determinada cara está dada por (N/6)/N = 1/6, cuando N es muy grande.

Fuente de información: Origen material de la información o lugar donde se puede encontrar. Puede ser una persona, una institución o un documento. Las fuentes pueden ser primarias, secundarias o terciarias conforme al carácter de la información transmitida.

Gráfico Pie: Es un gráfico circular (es un gráfico círculo dividido en partes proporcionales). Ejemplo hi de la tabla 1.

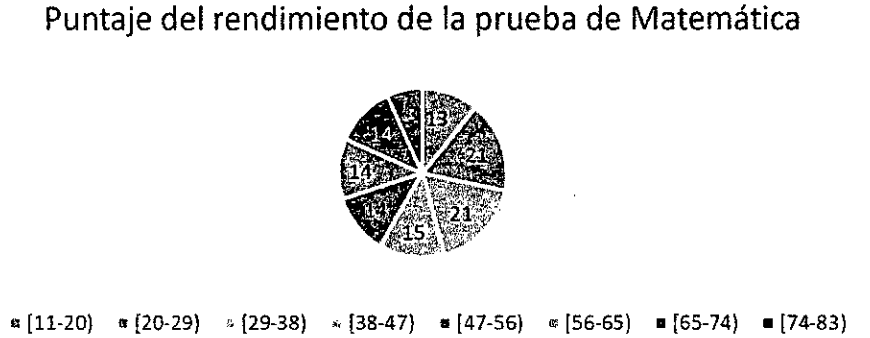


Figura N°1. Gráfico circular de la distribución del número de estudiantes según su puntaje de la prueba de matemática.

Hipótesis: Son los datos o proposición inicial que se toma por verdadera para iniciar el razonamiento.

Histograma: Gráfico en donde se asocia cada valor de la variable con un rectángulo cuya altura es igual a la frecuencia absoluta del valor de la variable. Se trabaja sobre valores de un intervalo (variable continua). Ejemplo: Histograma la tabla 1.

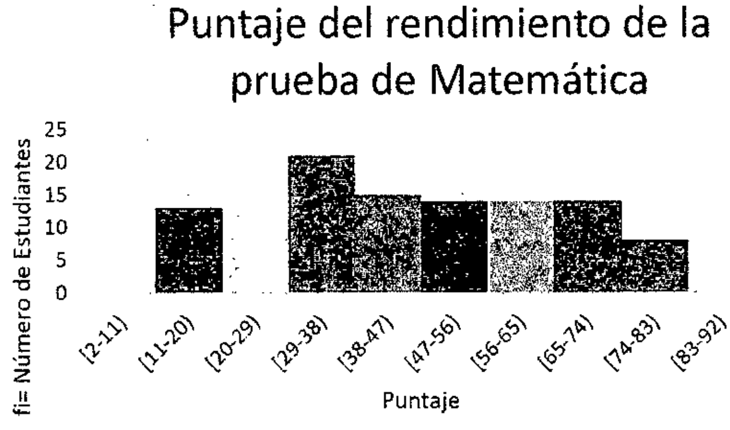


Figura N°2. Histograma de la distribución del número de estudiantes según su puntaje de la prueba de matemática.

Igual de probable: Resultado que puede ocurrir con igual frecuencia que otros.

Indicador: Es el elemento característico que describe una situación permitiendo su análisis. Por ejemplo, el coeficiente de promoción es un buen indicador del éxito de las acciones educativas.

Índices: Son números relativos que expresan el valor de una cierta cantidad al compararla con otra análoga (o con otra de época distinta) que se ha tomado como base igual a 100.

Investigación aplicada: Consiste también en trabajos originales realizados para adquirir nuevos conocimientos, pero fundamentalmente dirigidos hacia un fin u objetivo práctico específico.

Investigación básica: Consiste en trabajos experimentales o teóricos que se emprenden principalmente para obtener nuevos conocimientos acerca de los fundamentos de fenómenos y hechos observables, sin prever en darles ninguna aplicación o utilización determinada o específica.

Investigación y desarrollo (I+D): Se entiende por I+D cualquier trabajo creativo llevado a cabo en forma sistemática para incrementar el volumen de conocimientos, incluido el conocimiento del hombre, la cultura y la sociedad y el uso de éstos para derivar nuevas aplicaciones. Comprende: Investigación Básica, Investigación Aplicada y Desarrollo Experimental.

Investigador: Es la persona que trabaja en la concepción o creación de nuevos conocimientos, productos, procesos, métodos y sistemas y en la gestión de los respectivos proyectos. Incluye al personal superior que desarrolla actividades de planificación y gestión de los aspectos científicos y técnicos del trabajo de los investigadores.

Jerarquía: Clasificación vertical según la cual cada nivel o categoría está subordinado al nivel inmediatamente superior. El criterio de clasificación debe ser único. A veces, es posible articular varias jerarquías en una taxonomía.

Laboratorio: Espacio académico útil en la experimentación y trabajo de investigación científica. Aula de práctica en la cual se hacen demostraciones y el alumno puede realizar experimentos e investigaciones.

Media aritmética o promedio $\bar{x}$: Valor que se obtiene, en una distribución de datos no agrupados, al dividir la suma de los valores de los datos entre el número total de estos. Ejemplo: Si las calificaciones finales de un estudiante son:

12	13	12	11	8	9	10	17

entonces: la suma es 92 por lo que $\bar{x} = \frac{92}{8} = 11.5$

luego su promedio es 11.5 puntos. En algunos casos, la media aritmética puede calcularse también por compensación. Ejemplo: La media entre 12, 14 y 16 es 14 ya que 16 excede a 14 en 2, que es lo que le falta a 12 para llegar a 14.

Mediana: Valor de la variable cuantitativa que divide un conjunto de datos, una vez que estos han sido ordenados, en dos partes con igual número de datos. Ejemplo: Si las edades de un grupo de alumnos son 12, 15, 14, 11, 12, 14, 13, 12 y 15 años, al ordenar los datos se tiene: 11, 12, 12, 12, 13, 14, 14, 15, 15. La mediana de estas edades es 13. Si el número total de datos es par, por ejemplo: 8, 35, 13, 18, 14, 38, 6, 33, 10, 25, para determinar la mediana se ordenan los datos: 6, 8, 10, 13, 14, 18, 25, 33, 35, 38, y se calcula la semisuma de los dos términos centrales: $(14+18)/2 = 32/2 = 16$

Medida: Expresión numérica de la relación que existe entre dos valores de una misma magnitud, uno de los cuales se ha adoptado convencionalmente como unidad. Es la expresión numérica del resultado de medir una magnitud, dimensión o cantidad.

Medidas de tendencia central: Datos que se ubican hacia el lugar central de una lista de datos. Hay tres medidas de tendencia central: media aritmética, mediana y moda.

Medir: Determinar la cantidad de una magnitud por comparación con otra que se toma como unidad.

Menos probable: Resultado que ocurrirá con menos frecuencia que otro

Método: Hipotético Deductivo (parte de la observación, se plantean hipótesis, se realiza la contrastación de las hipótesis, se obtienen las conclusiones y se generalizan los resultados).

Moda: Dato que tiene mayor frecuencia en una distribución de datos no agrupados. Un conjunto de datos puede tener varias modas. Ejemplo: Las estaturas, en centímetros, de un grupo de alumnos de 5° grado son: 156, 158, 160, 170, 158, 170, 165, 168, 168, 158. La moda es 158.

Muestra: Subconjunto de una población, seleccionado para la investigación y análisis. Para analizar la muestra debemos hacernos las siguientes preguntas: ¿Es la variable cualitativa o cuantitativa?, ¿Es la población finita o infinita?, ¿Qué técnicas estadísticas o algoritmos computacionales existen para analizar los datos, ¿cuántos datos se necesitan en cada una de las técnicas o algoritmos computacionales para su correcta aplicación?, ¿Qué método de muestreo debe utilizarse?, ¿Qué tamaño de muestra se debe elegir?, ¿Cuánto de error debe utilizarse? Excel

Ejemplo: Si se quiere conocer la edad promedio de los alumnos de la UNAC, una muestra podría estar conformada por los estudiantes de cuarto ciclo de la FIIS-UNAC.

Muestreo: Método o sistema usado para obtener una muestra ante un caso particular, las circunstancias determinarán el método de muestreo o la combinación de métodos a ser empleados.

Observaciones: Son los datos que se recolectan para su estudio.

Parámetro: Es una medida resumen que describe una característica de la población. Es la expresión que designa un valor numérico que caracteriza a una población.

Patrón de Medida. Valor tipo que o que sirve para definir una unidad de medida.

Placebo: Una sustancia que no tiene efecto en el cuerpo (a menudo llamada píldora de azúcar), y es dada a un grupo en un estudio de placebo controlado. En estudios de medicamentos de placebo controlado, el placebo es dado a un grupo de participantes mientras el medicamento que se está estudiando se le da a otro grupo. Los resultados obtenidos en cada grupo son posteriormente comparados.

Pictograma: Los pictogramas son gráficos similares a los gráficos de barras, pero empleando un dibujo en una determinada escala para expresar la unidad de medida de los datos. Generalmente este dibujo debe cortarse para representar los datos. Se usan para lograr el interés masivo del público. Ejemplo:

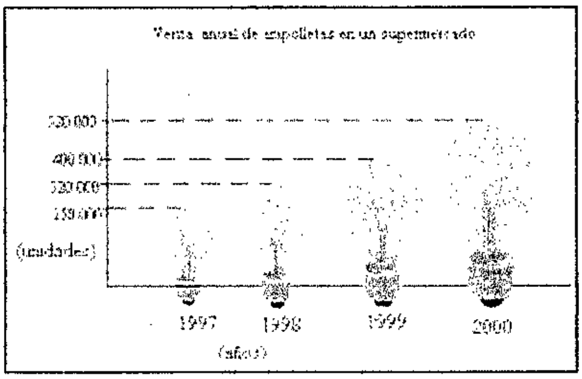


Figura N°3. Venta anual de ampolletas en un supermercado

Población: Conjunto de personas u objetos con características comunes, observables y medibles. Ejemplo: Si se quiere investigar el precio promedio de los automóviles en Perú, entonces la población está constituida por todos los automóviles de Perú.

Por ciento. Palabras que significan "centésimos" o "por cada cien" 42 por ciento (42%) es igual a 0,42 ó 42/100.

Porcentaje: Tanto por ciento, proporción de una cantidad respecto a otra evaluada sobre la centena. Relación existente entre un número y el número cien.

Procesamiento de datos: Es el tratamiento de datos originales según reglas preestablecidas para clasificarlos, ordenarlos, sumarizarlos, someterlos a cálculos y registrarlos. Puede ser manual (lápiz, formulario, tarjetón, archivador); mecánico (máquina de escribir, de sumar, de contabilidad, etc.) o electrónico (computador u ordenador).

Proceso: Conjunto de actividades relativas a la producción, obtención, elaboración, fabricación, preparación, conservación, mezclado, acondicionamiento, envasado, manipulación, ensamblado, transporte, distribución, almacenamiento y expendio o suministro al público de productos y servicios.

Probabilidad: Mayor o menor posibilidad de que ocurra algo en un suceso de azar. Relación entre el número de casos favorables y el número total de casos posibles para un suceso cualquiera, suponiendo que todos los casos son igualmente probables. (Un número, de 0 a 1, indica la probabilidad de que un suceso determinado ocurra. Cuanto más cerca de 0, menor es la probabilidad de que ocurra). Ejemplo: Si en una bolsa hay 4 bolsas amarillas y 6 bolsas azules, la probabilidad de sacar al azar una bolsa azul es 6/10.

Problema: Situación que comprende una pregunta que no puede ser respondida de manera inmediata, lo cual requiera hácer uso de conceptos previamente aprendidos y de destrezas que se han desarrollado para encontrar alguna solución de carácter cualitativo o cuantitativo.

Proporción: Relación en cuánto a magnitud, cantidad o grado, de una cosa con otra o de una parte con el todo. Igualdad entre dos razones.

Proyecto de investigación y desarrollo: Es un conjunto coordinado de tareas científicas y tecnológicas específicas que comprende total o parcialmente actividades de I+D, y que a partir de conocimientos preexistentes permite llegar a un objetivo cuyas características han sido previamente determinadas y/o aumentan su conocimiento.

Recolección de datos. Obtención de datos relevantes sobre objetos, situaciones, etc.

Redondeo. Aproximación de un número a la decena, a la centena, a la unidad de mil, etc., más próxima Ejemplo: 582 redondeado a la decena más próxima es 580, redondeado a la centena más próxima es 600.

Seguimiento: El seguimiento consiste en la evaluación continua de un proyecto o del desarrollo de algún acontecimiento. Es más importante la observación del proceso interno que la evaluación de los resultados. Esto se hace para garantizar la pronta intervención con el fin de prevenir los desarrollos equivocados en lugar de identificar y eliminar los resultados erróneos únicamente al final del proceso. El seguimiento presupone por lo tanto unos parámetros de éxito, en fases o subproyectos que hayan sido previamente definidos como resultados provisionales, hitos, etc.

Sistema de reclutamiento y selección: Es un conjunto de procedimientos e instrumentos tendentes a captar y evaluar los méritos y capacidades de las personas que aspiran a desempeñar un determinado cargo administrativo en una institución.

Suceso imposible: Suceso o evento que nunca puede ocurrir. Suceso en el cual la probabilidad de que ocurra es cero. Ejemplo: Sacar un número mayor que 7 al lanzar un dado.

Suceso posible: Suceso o evento que puede ocurrir. Suceso en el cual la probabilidad de que ocurra es mayor que cero y menor que uno. Ejemplo: Sacar un tres al lanzar un dado.

Suceso seguro: Suceso cierto, que ocurrirá sin lugar a dudas. Suceso en el cual la probabilidad de que ocurra es uno. Ejemplo: Sacar 2 números que sumen menos que 19 de una bolsa que contiene bolas numeradas del 1 al 9.

Tabla de frecuencias: Disposición en la cual se presenta un conjunto de valores, junto con sus frecuencias. Ejemplo tabla 1.

Tabla de proporcionalidad: Tabla que permite observar la relación que hay entre dos magnitudes. Se le llama también tabla de variación.

Tabla o cuadro estadístico: Disposición de los datos numéricos en filas y columnas de manera que se puede apreciar de la mejor forma las características y la cuantía del fenómeno estudiado, y establecer comparaciones entre aquellas. (Si el cuadro estadístico se refiere a una sola variable se llama de simple entrada, si se trata de dos variables correlacionadas, como pesos y estaturas de un grupo de personas, se construye un cuadro o tabla de doble entrada). Ejemplo:

Tabla N°2: *Tabla del tipo de ansiedad de 59 hombres y 63 mujeres*

	Ansiedad Alta	Ansiedad Media	Ansiedad Baja
Hombres	17	25	17
Mujeres	25	19	19

Tesis: Es un trabajo de investigación que debe presentar el estudiante después de concluir el plan de estudio de la carrera. Dicho trabajo de investigación debe versar sobre problemas nacionales o relacionados con la carrera elegida, y constituye un requisito básico para la obtención del título.

Tests de aptitudes: Pruebas concebidas para verificar las aptitudes de los individuos con la finalidad de definir las orientaciones a darles, más que a medir sus capacidades reales.

Unidad de competencia: Función integrada por una serie de elementos de competencia y criterios de desempeño asociados, los cuales forman una actividad que puede ser aprendida, evaluada y certificada.

Unidad de competencia básica: La referida a las habilidades consideradas como mínimo para la realización de cualquier trabajo.

Unidad de competencia específica: Se refiere a conocimientos, habilidades y destrezas propios de una función que se identifica generalmente con una ocupación.

Unidad de competencia genérica: La que se refiere a funciones o actividades comunes a un número significativo de áreas de competencia.

Unidad de observación (o unidad de análisis): Son los elementos o unidades observados. Es decir, es la unidad de la cual se obtiene el dato estadístico.

Usuarios: Designa toda la comunidad de aquellos que pudieran necesitar información en uno u otro momento. Incluye los formularios de políticas, los decisores y, en general, los especialistas de la investigación y el desarrollo (este término se emplea corrientemente para designar a personas que utilizan efectivamente un determinado servicio o producto de información, aquellos que es probable que hagan lo mismo se llaman usuarios potenciales).

Validez: Es el instrumento que mide realmente lo que se pretende medir (Es decir ausencia de sesgo). ¿Qué tipo de validez? Validez de contenido deberá ser medido por juicio de expertos, que evalúan la pertinencia, la relevancia y la claridad. Si indican que existen suficiencia (se aplicará la prueba Aiken) luego se podrá aplicar dichos instrumentos.

Variable: Según Hernández, Fernández y Batista (2010). Señalan que una variable es una propiedad que puede variar y cuya variación es susceptible de medirse u observarse. Una variable es una característica de las unidades de la muestra, propiedad que se presenta un fenómeno que varía, en efecto puede ser medidos o evaluado. Objeto matemático que puede tomar diferentes valores.

Variable continua. Variable estadística que teóricamente puede tomar cualquier valor entre dos valores dados. Ejemplo: La talla de una persona es una variable continua.

Variable cualitativa. Variable estadística que se expresa mediante un nombre. Ejemplo: *Ejemplos:*

- *Sexo (hombre, mujer)*
- *Salud (buena, regular, mala)*

Variable cuantitativa. Variable estadística que se expresa mediante un número. Ejemplo: La edad, la temperatura, la estatura, etc.

Variable cuantitativa discreta: Variable estadística que teóricamente no puede tomar cualquier valor entre dos valores dados, son números enteros positivos. Ejemplo: El número de hijos que puede tener una familia es una variable discreta, pues es un número entero.

Variable cuantitativa continua: Variable estadística que teóricamente puede tomar cualquier valor entre dos valores dados, pueden ser números racionales positivos. *Edad (12,5; 24,3; 35;...). Continua.*

Variable estadística: Es el conjunto de valores que puede tomar cierta característica de la población sobre la que se realiza la investigación y sobre la cual es posible su medición, o sea es cualquier característica de una persona, un medio ambiente o una situación experimental, que puede variar de persona a persona, de un medio ambiente a otro, o de una situación experimental a otra. Ejemplo: El peso, la edad, el sexo, etc. El Indicador, es una sub-variable que se desprende con el propósito de medirla ejemplo la variable inteligencia puede ser subdividida en los indicadores: Rendimiento obtenido en un test, respuesta novedosa en la solución de un problema, rapidez en la emisión de la respuesta intelectual, etc.

Verificación: Es la constatación o comprobación del cumplimiento de las normas.

Fuente: file:///C:/2010/Temas%20Para%20CD/TIC/glosario%25203.pdf

1.4. Niveles de Medición.

En estadística se definen como variables a los atributos, rasgos o propiedades de un grupo de elementos que toman diferentes valores, magnitudes o intensidades.

En el proceso de medición de ellas se les asignan números o códigos de observación. La manera más aceptada para ordenar y cuantificar una variable, es dividirlas en cualitativas (según su calidad o atributo) o cuantitativas (de acuerdo a la magnitud de su medición). Cuando la variable cualitativa no tiene punto de comparación como el color de los ojos (café, azul, verde, negro) se le denomina variable cualitativa nominal; cuando hay un determinado orden como clase social (alta, media, baja), duración de una enfermedad (aguda, subaguda, crónica), orden en la familia (primero, segundo, tercero, etc.) se le llama variable cualitativa ordinal. Cuando la variable cuantitativa sólo se puede medir en valores enteros: como el número de alumnos, el número de partos, el número de empleados, se le denomina variable cuantitativa discreta, discontinua o de intervalo, mientras que, si la variable se puede expresar en fracciones, como peso al nacimiento (3,565 g) o estatura (50.9 cm), se les denominan variables cuantitativas continuas o de razón y puede ser que los datos tengan una distribución normal (sesgo de -0.5 a +0.5 y curtosis de 2 a 4)

MEDICIÓN

¿Qué es medir?

Medir es asignar un valor a una característica, el cual puede ser un numero o una categoría.

Por ser muy importante la medición en una investigación le daremos un especial desarrollo para clarificar lo que se quiere medir en una investigación.

Instrumentos de Investigación:

Los instrumentos que se requiere en una investigación pueden ser de diferentes tipos: medición, de constatación, de acopio de información, de verificación de situaciones, etc. Los instrumentos más conocidos y los que proporcionan información más valiosa al investigador son los instrumentos de medición. La medición es una actividad muy importante cuando se trata de conocer la naturaleza de los fenómenos que proporcionan la información precisa acerca de sus características.

Para realizar un proceso de medición es necesario reconocer que el fenómeno a medir tiene su propia magnitud y que el problema radica en que el investigador, con los instrumentos que dispone, no la puede conocer plenamente. La medición es, en estricto sentido, conocer la verdadera magnitud del fenómeno, de ahí que resulta muy importante que toda auténtica medición sea isomórfica con la realidad que se está midiendo, es decir, que los datos que se obtengan como resultado de la medición debe ser parecidos, equivalentes, o correspondientes a los que realmente posee el fenómeno que se mide; aunque en realidad no se mide el fenómeno directamente, sino los indicadores de sus características. Es muy importante tener en cuenta esto cuando se realizan procesos de medición de fenómenos o variables del comportamiento que, por naturaleza, son muy elusivas, impredecibles y difíciles de identificar, como la creatividad, el talento, el coeficiente intelectual, la personalidad, la agresividad, el rendimiento académico, la angustia, etc. Como estos fenómenos ofrecen dificultades a los esfuerzos por medirlos, solo se puede inferir sus características a partir de la observación o el análisis de sus indicadores.

Medición: Proceso que vincula conceptos abstractos con indicadores empíricos.

Medir: Significa "asignar números, símbolos o valores a las propiedades de objetivos o eventos de acuerdo con reglas". (Stevens, 1951)

Para ciencias sociales: la medición es "el proceso de vincular conceptos abstractos con indicadores empíricos", el cual se realiza mediante un plan explícito y organizado para clasificar (y con frecuencia a cuantificar) los datos disponibles (los indicadores), en términos del concepto que el investigador tiene en mente (Carmines y Zeller, 1991). En este proceso, el instrumento de medición o de recolección de datos tiene un papel central. Sin el, no hay observaciones clasificadas.

La Medición

La medición es un proceso que consiste en asignar numerales a determinados fenómenos o eventos, siguiendo reglas previamente establecidas. Esta definición dada por Stanley Stevens (1946) en su publicación Teoría Escalas de Medición nos dice que es intencionalmente muy genérica, pero es muy útil porque hace posible abarcar todos los aspectos que se derivan del proceso de medición y, además, porque permite sostener que es posible, teóricamente, medir cualquier fenómeno siempre y cuando las reglas tengan un fundamento lógico.

Niveles de medición

Existen cuatro niveles (o escalas) de medición según Mohammad (2005) y Hernández, Fernández y Batista (2010), que son: Cuando la variable es nominal. Ordinal, de intervalo y cundo es de razón:

1. Nivel de medición nominal. Es la escala de medición más baja. Consiste en establecer relaciones de igualdad y/o desigualdad entre objetos que se están midiendo, no admitiéndose puntuaciones ordenadas. Esta escala establece los grupos de acuerdo con la presencia o ausencia de un atributo característica. Se usa para medir variables cualitativas cuyos valores no pueden ser ordenados de acuerdo a intensidad, es decir, de menor a mayor,

la escala nominal solo permite hacer muy pocas operaciones estadísticas, tales como contar frecuencias. La única relación que puede establecerse entre observaciones serán iguales si están en la misma clase.
En este nivel hay dos categorías del ítem o la variable. Las categorías no tienen orden ni jerarquía. Lo que se mide (objetivo, persona, etc.) se coloca en una u otra categoría, lo cual indica tan solo diferencias respecto de una o más características. La clasificación de los objetos debe ser: Exhaustiva y Excluyente.
Esta escala puede ser de dos tipos:

Dicotómicas, que admite dos valores. Ejemplo: genero.

	Categoría	Números	Etiquetas o símbolos
Sexo	Masculino	1	*
	Femenino	2	&

Polinómicas, admite más de dos posibles valores. Ejemplo: Estado civil, raza, etc.

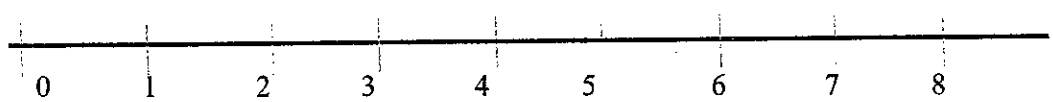
2. Nivel de medición ordinal. Se utiliza cuando se requiere colocar en orden (1°, 2°, 3°, etc.) en relación al atributo. Las observaciones no solo difieren de categoría a categoría, si no que pueden clasificarse de acuerdo con algún criterio. Las etiquetas o símbolos de las categorías si indican jerarquía. Por ejemplo. Se miden las actitudes haciendo varias preguntas con múltiples respuestas, las cuales pueden ordenarse en forma ascendente o descendente, es el caso del grado de satisfacción de los trabajadores de cierta empresa puede ser: ¡Esta usted a gusto con su jefe inmediato? Si se pide al entrevistado que marque el número que representa su grado de satisfacción.

Muy contento	Contento	Descontento	Muy descontento
1	2	3	4

3. Nivel de medición por intervalos. Además de haber orden o jerarquía entre categorías, se establecen intervalos iguales en la medición. Las distancias entre categorías son las mismas a lo largo de toda la escala. Hay intervalo

constante, una unidad de medida. Permite identificar las diferencias en monto, tamaño, cantidad o distancia utilizando puntuaciones para ello.

Para comparar cualitativamente se deben tener unidades de medición, algunas variables que se pueden medir con escala de intervalo son: altura, temperatura, ingreso y otras.



4. Nivel de medición de razón. Este nivel tiene todas las características del nivel de intervalos (periodos iguales entre categorías y aplicación de operadores aritméticas básicas y sus derivaciones) y agrega el cero es real y es arbitrario. Ejemplo. Exposición a las redes sociales (en minutos), el número de hijos, las ventas de un producto, los metros cuadrados de construcción, ingresos (en moneda), presión arterial, etc.

Ejemplo: Reconozca el nivel de medición de las siguientes variables

1. Género (Hombre – Mujer)
2. Notas en la Asignatura "Estadística"
3. Tiempo de Espera para ser atendido en un hospital
4. Cantidad de horas que veo TV al día
5. Número de teléfono
6. Compañía de teléfono a la cual puedo estar asociado
7. Nivel de ansiedad frente al examen de la asignatura

Respuestas

1. Nominal 2. Intervalo 3. Razón 4. Razón 5. Nominal 6. Nominal 7. Intervalo

Instrumento de medición:

Recurso que utiliza el investigador para registrar información o datos sobre las variables que tiene en mente.

Un instrumento de medición adecuado es aquel que registra datos observables que representan verdaderamente los conceptos o las variables que el investigador tiene en mente.

En una investigación cuantitativa, los instrumentos se aplican para medir las variables contenidas en las hipótesis.

En una investigación cualitativa, cuando no hay hipótesis los instrumentos se utilizan para medir las variables de interés.

En una investigación cuantitativa, los instrumentos se aplican para medir las variables contenidas en las hipótesis.

Requisitos que debe cubrir un instrumento de medición: Confiabilidad, Validez, Objetividad.

Confiabilidad: Se refiere al grado en que su aplicación repetida al mismo sujeto y objeto produce resultados iguales.

Ejemplo. Un termómetro que marca en este momento una temperatura de 22° centígrados y un minuto más tarde se consulta marca 42° centígrados y media hora más tarde, señala 7° centígrados, el termómetro no sería confiable.

Pregunta: ¿el instrumento mide adecuadamente las principales variables en cuestión?,

Validez: Se refiere al grado en que un instrumento realmente mide la variable que pretende medir.

Ejemplo: Un instrumento válido para medir la inteligencia debe medir la inteligencia y no la memoria. Un instrumento sobre conocimientos de Aritmética tiene que medir esto y no conocimientos de Algebra.

Tipos:

Validez de contenido: se refiere al grado en que un instrumento refleja un dominio especifico de contenido de lo que mide. Esta validez se obtiene mediante las opiniones de expertos y asegurarse que las dimensiones medidas por instrumento sean representativas del universo o domino de las dimensiones de las variables de interés.

Validez de Criterio: establece la validez de un instrumento de medición al comparar sus resultados con los de algún criterio estreno que pretende medir lo mismo. Si el criterio se ajusta al futuro se habla de validez predictiva. Por ejemplo, una prueba de admisión en las universidades puede comparar los resultados con el rendimiento futuro de los estudiantes en la carrera. Si el criterio se fija en el presente se habla de validez concurrente; es cuando los resultados del instrumento correlacionan con el criterio en el mismo momento o en tiempo real. Ejemplo. Una encuesta administrada un día antes de las votaciones para detectar preferencias del electorado, correlaciona con los resultados de la elección.

Validez de constructo: es la más importante sobre todo desde la perspectiva científica, y se refiere a qué tan exitosamente un instrumento representa y mide un concepto teórico.

Incluye tres etapas:

1. Se establece y especifica la relación teórica entre los conceptos (sobre la base de la revisión de la literatura)
2. Se correlacionan los conceptos y se analiza cuidadosamente la correlación.
3. Se interpreta la evidencia empírica de acuerdo con el nivel en que clarifica la validez de constructo de una medición particular.

Esta validez se obtiene mediante el análisis de factores. Tal método nos indica cuantas dimensiones integran a una variable y que ítems conforman cada dimensión. Los reactivos que no pertenezcan a una dimensión y no midan lo mismo que los demás ítems, se pueden eliminar.

Validez Total = Validez de contenido +Validez de criterio + validez de constructo.

La objetividad: Se refiere en qué grado el instrumento es permeable a la influencia de los sesgos y tendencias de los investigadores que lo administran, califican e interpretan.

Instrumentos para la recolección de datos (documento)

A. Cuestionarios: Debemos tomar en cuenta lo siguiente:
    a) Escribir las preguntas en positivo.
    b) Organizar las preguntas por temas.
    c) Como regla general, por cada pregunta que se coloque en el cuestionario se necesitan de 5 a 10 cuestionarios. Por ejemplo, si la técnica estadística a utilizarse es análisis factorial y el cuestionario tiene 20 preguntas, entonces se necesitan al menos 100 cuestionarios.
    d) Tratar en lo posible tener preguntas cerradas. Si se decide utilizar preguntas abiertas, entonces usar text mining.
    e) Google forms
B. Cuestionarios: Los instrumentos para recolectar datos deben ser confiables, para ello debemos usar el:
    a) Alfa de Cronbach: coeficiente que sirve para medir la fiabilidad de una escala de medida,
    b) Índice de Kayser-Meyer-Olkin (KMO): Mide la asociación entre dos variables una vez que se han descontado los efectos de otras variables. Los valores aceptables de KMO deben ser mayores a 0.7.
    c) Prueba de Bartlett: comprobar que la matriz de correlaciones no es una matriz identidad. El indicador de la prueba (p-valor) debe ser menor a 0.05.

1.5. Clasificación y presentación de la información en tablas

Para analizar una muestra y organizarlos en tablas se debe hacerse las siguientes preguntas: ¿Es la variable cualitativa o cuantitativa?, ¿Es la población finita o infinita?, ¿Qué técnicas estadísticas o algoritmos computacionales existen para analizar los datos, ¿cuántos datos se necesitan en cada una de las técnicas o algoritmos computacionales para su correcta aplicación?, ¿Qué método de muestreo debe utilizarse?, ¿Qué tamaño de muestra se debe elegir?

Organización de los datos: distribución de frecuencias.

Después de obtener los datos es necesario resumirlos y presentarlos en forma tal, que faciliten su comprensión y su posterior análisis y utilización. Para ello se ordenan en cuadros numéricos y luego se representan en gráficos.

Todo cuadro numérico debe tener:

- Un *título* adecuado para evitar confusiones y para expresar brevemente su contenido.
- La *fuente* de los datos, si no son datos propios.
- Las *unidades* en que se expresan los datos.

Los cuadros numéricos de una sola variable estadística se denominan *distribución de frecuencias.*

En el procedimiento para construir distribuciones de frecuencias *nos referiremos a muestras,* mientras no se diga lo contrario.

Distribución de frecuencias: variable cualitativa

Supongamos que en una muestra de n unidades estadísticas se observan k categorías o modalidades diferentes de alguna variable cualitativa X.

La tabulación de estos n datos, es la DISTRIBUCIÓN DE FRECUENCIAS POR CATEGORÍAS del cuadro 1.

La **frecuencia absoluta** f_i, es el número de datos observados en cada categoría o modalidad.

La **frecuencia absoluta** F_i, es la frecuencia absoluta acumulada y F^*_i su inversa

La suma de todas las frecuencias absolutas es igual al total n de datos observados.

La **frecuencia relativa** h_i se define en cada categoría por $h_i = f_i/n$. La suma de todas las frecuencias relativas es igual a uno.

La **frecuencia relativa** H_i es la frecuencia relativa acumulada y H^*_i su inversa

La frecuencia porcentaje pi, se define en cada fila por $p_i = h_i$ x 100%. El total de las frecuencias porcentajes es igual a cien.

Ejemplo:

La siguiente tabla pertenece a los planes de estudios superiores de un grupo de 548 estudiantes del último año del bachillerato:

Tabla N°3. *Frecuencia absoluta simple de los planes de estudios superiores de un grupo de estudiantes del último año de bachillerato*

	f	h	%hi
Planean ir a la universidad	240	0,4379	43,79
Quizá vayan a la universidad	146	0,2703	27,03
Planea ir o quizá vayan a una escuela vocacional	57	0,1055	10,55
No irán a ninguna universidad	105	0,1944	19,44

Para la tabulación de datos cualitativos también se pueden usar tablas de contingencia o

Super tablas, el valor de una tabulación cruzada consiste en que proporciona una idea de la relación entre las variables (ya sean ambas cualitativas, ambas cuantitativas o combinación de ambas).

a) ***Amplitud del intervalo.***

- Identificar el valor mayor X_{max} y el valor menor X_{min} del conjunto de datos.
- Determinar el número de intervalos m. Se recomienda que $5 \leq m \leq 20$.

Cuando no se tiene referencia para el valor de m, se puede calcular utilizando alguna fórmula. Una de las más utilizadas es la de Sturges: k=1+3.3 log *n;* donde k es el número aproximado de intervalos y n es el número de datos. En general, cuando el número de intervalos es demasiado pequeño, podría quedar ocultas algunas características importantes de los datos al agruparlos. Si el número de intervalos es grande, es posible tener muchos intervalos vacíos y la distribución dará una deficiente descripción de los datos.

Variable cuantitativa continua (cc) regla de Sturges:

Rango o Recorrido (R) : R = Dato Máximo - Dato Mínimo

Número de Intervalos a formar (K) :

- K= 1+3.322*log(n) ... (Redondear a enteros)

Tamaño del Intervalo de Clase (TIC) o Amplitud del Intervalo (Ci) :

- TIC= Ci = R/K (Redondear con = # de decimales de los datos)

Diferencia o Exceso (D) : D = (K*TIC) - R

- Calcular la amplitud del intervalo C, utilizando la fórmula $C = {(x_{\max} - x_{\min})}/{m}$

  Si se desea que la amplitud del intervalo sea de valor entero, como generalmente ocurre, el valor de C debe redondearse al entero mayor (por exceso), si el valor calculado no es entero. Si el valor calculado es entero, debe incrementarse en 1.
- Es deseable que todos los intervalos tengan la misma amplitud para poder hacer comparaciones uniformes de las frecuencias de los intervalos.
- Establecer los límites de los intervalos, definiendo primero los límites de los intervalos inferiores a partir de $x_{\min}$ y luego los límites superiores teniendo en cuenta que los intervalos no sean traslapes.
- Realizar el conteo para obtener las frecuencias para cada intervalo.

Marca de clase.

- Es el punto central de cada intervalo. Esta se suele representar por m_i. Para el intervalo *i-ésimo* viene será: $m_i = (L_i + L_{i-1})/2$

Ejemplo1: se tiene 300 postulantes a un puesto de trabajo dan una prueba de conocimientos cuyos puntajes están entre 180 puntos y 270 puntos elaborar la clasificación del número de filas y columnas y la distribución de frecuencias y el grafico del histograma frecuencia absoluta simple y el polígono de frecuencias dentro del histograma.

250	210	207	262	228	241	190	246	254	193
256	265	240	185	254	187	191	201	256	219
184	256	238	208	232	213	243	189	269	211
199	253	270	265	209	244	181	209	226	267
270	227	211	214	198	231	204	265	241	227
199	190	191	189	255	253	230	258	225	185
183	250	257	245	269	237	206	181	182	212
228	255	222	209	264	183	224	259	266	225
270	256	260	246	218	180	205	250	189	191
252	245	193	205	224	210	218	202	231	201
235	227	180	209	240	251	261	200	270	199
259	244	255	215	255	187	240	220	233	245
205	247	222	259	232	263	212	259	195	219
181	255	182	222	187	222	252	240	248	255
185	228	245	247	241	210	205	213	250	180
242	241	235	185	246	196	216	212	245	225
240	205	234	221	208	256	268	192	258	226
234	254	184	261	247	204	193	261	226	206
268	219	250	183	215	219	210	215	198	190
239	182	261	261	196	237	263	251	185	249
202	228	224	268	209	246	211	251	221	197
257	248	250	208	199	191	224	228	221	237
215	196	221	254	210	183	248	180	241	253
204	246	221	192	205	222	246	222	257	262
260	226	180	200	206	229	260	234	191	268
207	246	261	223	184	212	203	216	222	252
227	227	263	180	196	218	232	270	243	209
252	224	243	231	267	222	256	261	189	222
233	207	202	190	205	211	245	180	228	235
204	233	230	256	249	225	193	237	202	259

Aplicación de la regla de Sturges

Vmax=	270			
Vmin=	180			
Rango=	90			
k=1+3.3*log(n)=	9.17450014	10	11	Número de filas de la tabla
TIC=R/k=	9			Tamaño del intervalo de clase
D=TIC*k-R=	0		9	Diferencia, restar 4 a valor mínimo sumar 5 al valor máximo

Tabla N°4. *Tabla de frecuencias estudiadas*

IC	fi	Fi	hi	Hi	F_i^*	H_i^*
[176-185)	20	20	0.06666667	0.0667	300	1
[185-194)	27	47	0.09	0.1567	280	0.93333333
[194-203)	20	67	0.06666667	0.2233	253	0.84333333
[203-212)	36	103	0.12	0.3433	233	0.77666667
[212-221)	21	124	0.07	0.4133	197	0.65666667
[221-230)	40	164	0.13333333	0.5467	176	0.58666667
[230-239)	22	186	0.07333333	0.62	136	0.45333333
[239-248)	33	219	0.11	0.73	114	0.38
[248-257)	38	257	0.12666667	0.8567	81	0.27
[257-266)	29	286	0.09666667	0.9533	43	0.14333333
[266-275)	14	300	0.04666667	1	14	0.04666667
	300		1			

Grafico del ejemplo 1: Histograma de la frecuencia absoluta simple y un polígono de la frecuencia absoluta simple.

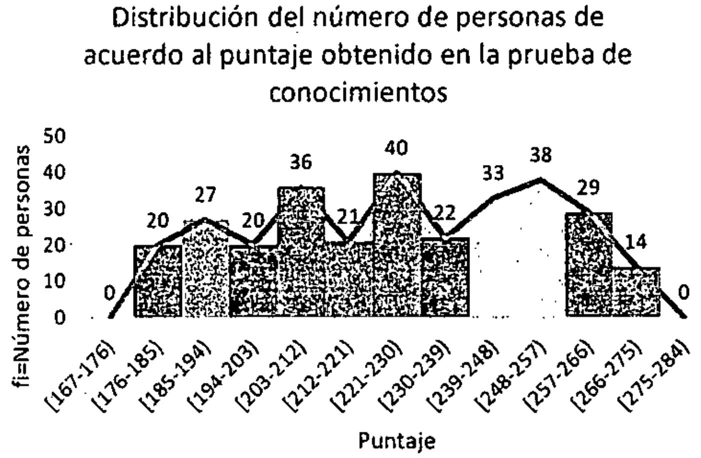


Figura N°4. Distribución del número de personas de acuerdo al puntaje en la prueba de conocimientos.

4. Para los siguientes datos discretos elaborados de la siguiente manera: Primero haga el sorteo aleatorio (en Excel) hasta la tercera vez entre 17 y 85 (edades de las personas que van al cine a ver la película de estreno) entre las columnas A y J y las filas del 1 y 10 de tal manera que tenga 100 datos aleatorios.

A) Determine: el Valor máx., Valor mín., el rango, el número de filas k, *el tamaño de intervalo de clase* TIC, el valor del exceso D.

Llene la tabla con datos agrupados usando la regla de Esturges.

B) Elabore un histograma de Frecuencia abs. Simple y un polígono de frecuencias incrustado en el polígono de frecuencias.

77	49	79	46	79	59	44	41	27
70	65	42	85	25	76	80	25	51
71	50	72	54	25	83	30	35	37
37	70	37	27	70	31	34	19	47
81	20	25	22	25	51	34	43	82
42	74	63	38	17	67	28	19	83
35	74	41	39	44	80	52	20	46
27	37	58	84	63	61	69	55	79
27	54	21	58	49	22	24	73	21
27	74	77	48	23	61	76	35	67

Solución

Usando Excel se tiene:

Vmax	85			
Vmin	17			
Rango=	68			
K=	1+3.3*log(n)	7.6	8	
TIC=R/K	8.5	9		
D=K*TIC-R	4			

Tabla N°5. *Tabla de frecuencias de personas de 17 años a 85 que van al cine.*

IC	fi	hi	Fi	Hi	Fi*	Hi*	hi*%	Hi*%	mi	mi*fi
[15-24)	10	0.1	10	0.1	100	1	10	10	19.5	195
[24-33)	14	0.14	24	0.24	90	0.9	14	24	28.5	399
[33-42)	15	0.15	39	0.39	76	0.76	15	39	37.5	562.5
[42-51)	13	0.13	52	0.52	61	0.61	13	52	46.5	604.5
[51-60)	12	0.12	64	0.64	48	0.48	12	64	55.5	666
[60-69)	8	0.08	72	0.72	36	0.36	8	72	64.5	516
[69-78)	16	0.16	88	0.88	28	0.28	16	88	73.5	1176
[78-87)	12	0.12	100	1	12	0.12	12	100	82.5	990
	100									5109

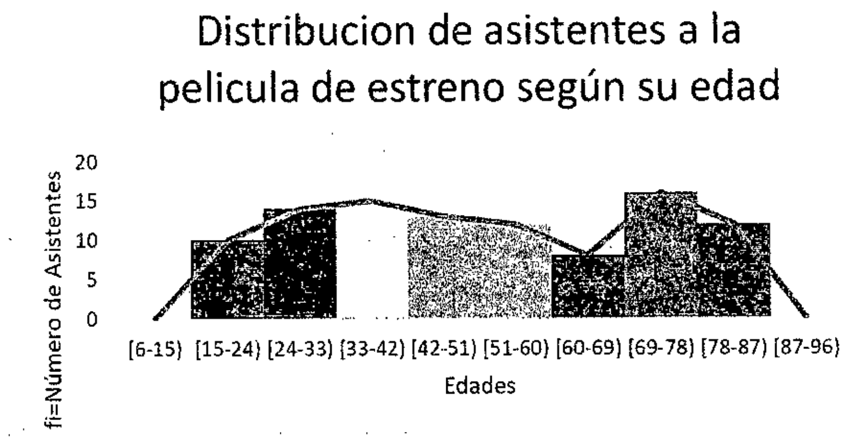


Figura N°5. Distribución de personas de 17 años a 85 que van al cine.

Gráficos circulares, de sectores o de pastel

Es un gráfico formado por un círculo dividido en sectores circulares cuyas amplitudes son proporcionales a las frecuencias de los datos representados son bastante útiles para visualizar diferencias de porcentajes, para representar datos cualitativos, etc. Pasos para su construcción:

1. Buscamos los porcentajes que representan a cada elemento.
2. Cada porcentaje se multiplica por 3,6 y daría el valor de los ángulos centrales.
3. Utilizar un transportador para ubicar cada ángulo.

También es posible hallar los ángulos centrales estableciendo una regla de tres entre la totalidad

del fenómeno (al cual le corresponden 360°) y la frecuencia de cada parte del fenómeno.

Diagramas de sectores (también llamados *tortas*). Se divide un círculo en tantas porciones como clases existan, de modo que a cada clase le corresponde un arco de círculo proporcional a su frecuencia absoluta o relativa

El arco de cada porción se calcula usando la *regla de tres*:

$$n \rightarrow 306^\circ \quad , \qquad n_i \rightarrow x_i = \frac{360 \times n_i}{n}.$$

1. **Ejemplo:** Un informe de un Ingeniero de un estado occidental en EUA, indico que 56% de los impuestos recaudados se destinaron a la educación, 23% al fondo general, 10% a los condados, 9% a los programas de la tercera edad, y el

remanente otros programas sociales. Trace una gráfica circular para mostrar la distribución del presupuesto.

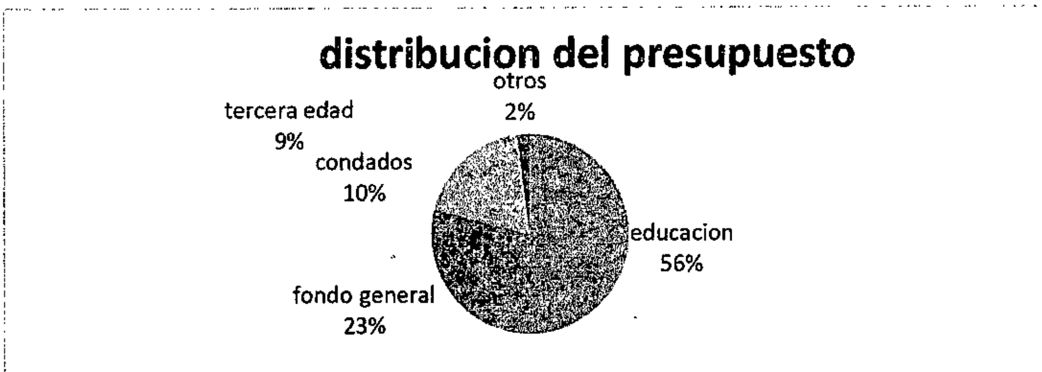
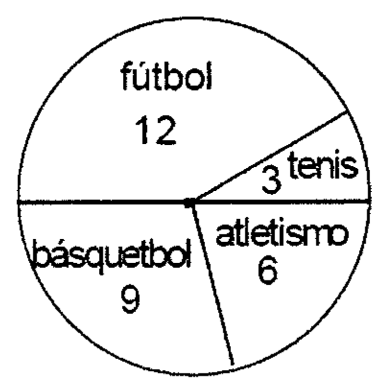


Figura N°6. Distribución de presupuesto de un estado de EUA

Ejemplo: Se da la siguiente tabla con diferentes disciplinas deportivas. Elabore el diagrama de sectores:

Tabla 6: *Tabla con diferentes disciplinas deportivas*

Deporte	Frecuencia Absoluta	Frecuencia Relativa	Frecuencia Relativa %
Fútbol	12	0,40	40
Tenis	3	0,10	10
Atletismo	6	0,20	20
Básquetbol	9	0,30	30
Total	30	1	100



I) La frecuencia relativa del grupo de fútbol es de 40%. (**Correcta**)

II) La frecuencia relativa del grupo de básquetbol es de 30%. (**Correcta**)

III) La mitad del grupo no prefirió fútbol ni tenis. (**Correcta**)

Gráficos de tallo y hoja

Es un diseño ideado por John Tukey que proporciona una impresión visual rápida del número de observaciones o de datos de una clase. Cada observación del conjunto de datos se divide en dos partes: el(los) dígito(s) principal(es) se convierte(n) en el tallo, y el (los) dígito(s) posterior(es) se convierte(n) en la(s) hoja(s). Los tallos se escriben a lo largo del eje principal, y por cada porción de datos se escribe una hoja para mostrar la distribución de los datos.

Entre las ventajas que tiene podemos mencionar que es más fácil de construir que un histograma y dentro de un intervalo de clase, este gráfico da más información que un histograma porque muestra los valores reales.

Ejemplo 1: Construir una representación tallo-hoja para el siguiente conjunto de 20 calificaciones en un test de habilidad gráfica:

82, 74, 88, 66, 58, 74, 78, 84, 96, 76, 62, 68, 72, 92, 86, 76, 52, 76, 82, 78

Elaborar el diagrama tallo-hoja agrupando 5 valores para cada tallo

Solución: Ordenando se tiene

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Ordenados	52	58	62	66	68	72	74	74	76	76	76	78	78	82	82	84	86	88	92	96

Tabla N°7: *Tabla de agrupación de Tallo y hojas*

Número de Agrupaciones	Agrupación de 5 valores	Tallo	Hoja	
1	(50-54)	5	2	
2	(55-59)	5	8	Menor frecuencia
3	(60-64)	6	2	
4	(65-69)	6	6, 8	
5	(70-74)	7	2, 4, 4	
6	(75-79)	7	6, 6, 6, 8, 8	Mayor frecuencia
7	(80-84)	8	2, 2, 4	
8	(85-89)	8	6, 8	
9	(90-94)	9	2	
10	(95-99)	9	6	

El diagrama de tallo-hoja sería:

Ejemplo 2: En un estudio del crecimiento de varones, se obtuvieron estas observaciones sobre el perímetro cefálico (en centímetros) de un niño al nacer:

33.1 34.6 34.2 36.1 34.2 35.6 34.5 35.8 34.5 34.2 34.3 35.1

35.2 33.7 36.0 34.2 34.7 34.6 34.3 33.4 34.9 33.8 33.6 35.3

35.2 34.6 33.7 34.8 33.9 34.7 35.1 34.2 36.5 34.1 34.0

Solución: Ordenando los datos de manera ascendente por columna, tenemos:

33.1	33.7	*33.9*	34.2	34.2	34.3	*34.6*	34.7	*34.9*	35.2	*35.6*	36.1
33.4	33.7	*34.0*	34.2	34.2	34.5	*34.6*	34.7	*35.1*	35.2	*35.8*	36.5
33.6	33.8	*34.1*	34.2	*34.3*	34.5	*34.6*	34.8	*35.1*	35.3	*36.0*	

Tabla N°8: *Tabla de agrupación de Tallo y hojas*

Número de agrupaciones	**Agrupación de los valores**	**TALLO**	**HOJAS**								
2	*33,0 - 33,4*	*33*	1	4							
5	*33,5 – 33,9*	*33*	7	7	9	8	6				
9	34.0 – 34,4	34	2	2	2	3	2	2	1	3	0
9	34,5 – 34,9	34	6	6	8	7	6	7	5	5	9
5	*35,0 – 35,4*	*35*	2	2	1	1	3				
2	*35,5 – 35,9*	*35*	6	8							
2	36,0 – 36,4	36	0	1							
1	36,5 – 36,9	36	5								

Dato menor: 33,1. Dato mayor: 36,5. Tallos posibles: 33, 34, 35 y 36.

Sólo existen 4 tallos, debiendo ser el mínimo 5 tallos, entonces los tallos propuestos se desdoblan, de la siguiente manera:Luego clasificamos los datos en función al valor del decimal:

Las principales propiedades de los datos son: Son 35 casos, la menor observación es de 33,1 cm. y la mayor observación es de 36,5 cm. El 51,43% (18 casos) de los datos están alrededor de los 34 cm. Son muy poco frecuentes los casos mayores a 36 cm. (3 casos) y menores a 33,5 cm. (2 casos). La distribución de datos está sesgada a la derecha (sesgo positivo), es decir, hay una mayor concentración de datos a la izquierda.

Ejercicios de Aplicación

2. El tercer renglón de una representación de tallo y hoja aparece como: 21 | 0 1 3 5 7 9.

Considere números enteros.

a) ¿Cuál es la amplitud de variación posible (o intervalo total) de los valores en este renglón?
El intervalo en este renglón es de 210 hasta 219.

b) ¿Cuántos valores de datos hay en este renglón?
Hay 6 valores.

c) Enumere los valores
Los valores son: 210 , 211 , 213 , 215 , 217 , 219.

3. La siguiente representación de tallo y hoja informa el número de cajas de leche Gloria del Centro Comercial "Vega"

Tabla N°9: *Tabla de agrupación de Tallo y hojas del Centro Comercial "Vega"*

3	12	689
6	13	123
10	14	6889
13	15	589
15	16	35
20	17	24568
23	18	268
(5)	19	13456
22	20	34679
16	21	2239
12	22	789
9	23	179
4	24	3
3	25	13
1	26	
1	27	0

a) ¿Cuántos días se estudiaron?
Se estudiaron 50 días.

b) ¿Cuántas observaciones hay en la última clase?
Hay 1 observación.

c) ¿Cuál es el valor más grande y el más pequeño en todo el conjunto de datos?
El más pequeño es 126 y el más grande es 270.

d) Enumere los valores reales del 4to renglón
Los valores reales del 4to renglón son: 155, 158, 159.

e) Enuncie los valores reales del penúltimo renglón.
Sin valores.

f) ¿En cuántos días se rentaron menos de 160 películas?
En 13 días.

g) ¿En cuántos días se alquilaron 220 películas o más?
En 22 días.

h) ¿Cuál es el valor intermedio o mediano?
El valor intermedio es 194.

i) ¿Cuántos días se rentaron entre 170 y 210 películas?
19 días.

4. El banco BBVA Está analizando el número de veces que se utiliza cada día su cajero automático en el Supermercado TOTUS. El siguiente es el número de veces que se usó durante cada uno de los últimos 30 días. Desarrolle una representación de tallo hoja. Resuma los datos referentes al número de veces que se utilizó el cajero automático.
¿Cuántas veces se usó en un día típico? ¿Cuáles son el mayor y el menor número de veces que se empleó el cajero? ¿Alrededor de que valores tiende a concentrarse la cantidad de veces de uso del cajero?

83	64	84	76	84	54	75	59	70	61
63	80	84	73	68	52	65	90	52	77
95	36	78	61	59	84	95	47	87	60

Solución

Tabla N°10: *Representación de tallo y hoja del Cajero automático en Totus.*

3	6
4	7
5	2 2 4 9 9
6	0 1 1 3 4 5 8
7	0 3 5 6 7 8
8	0 3 4 4 4 4 7
9	0 5 5

¿Cuántas veces se usó en un día típico?

En un día típico se usó aproximadamente 63 veces, puesto que en el intervalo de 60 a 69 se muestra más observaciones de uso.

¿Cuáles son el mayor y el menor número de veces que se empleó el cajero?

El mayor número de veces es 36 y el menor número de veces que se usó el cajero fue 95.

¿Alrededor de que valores tiende a concentrarse la cantidad de veces del uso del cajero?

Alrededor de 60 a 89 veces el uso del cajero.

5. La siguiente representación de tallo y hoja informa el número de pedidos recibidos por día en una empresa de venta por correo.
Solución
Tabla N°8: *Representación de tallo y hoja de pedidos recibidos en un día*

1	9	1
2	10	12
5	11	235
7	12	69
8	13	2
11	14	135
15	15	1229
22	16	2266778
27	17	01599
(11)	18	00013346799
17	19	03346
12	20	4679
8	21	0177
4	22	45
2	23	17

a) ¿Cuántos días se estudiaron?
   Se estudiaron 55 días.
b) ¿Cuántas observaciones hay en l 4ta clase?
   Hay 2 observaciones.
c) ¿Cuál es el valor más pequeño y el más grande?
   El más pequeño es 91 y el más grande es 237.
d) Enuncie los valores reales en la sexta clase.
   En la sexta clase tenemos los siguientes valores: 141, 143, 145.
e) ¿Cuántos días recibió la empresa menos de 40 pedidos?

8 días.

f) ¿Cuántos días recibió 200 pedidos o más?
12 días.

g) ¿Alrededor de cuantos días recibió 180 pedidos?
Alrededor de 3 días.

h) ¿Cuál es el valor central?
El valor central es 180.

6. La siguiente gráfica muestra el número de pacientes que ingresaron diariamente a la sala de urgencias del hospital Memorial.

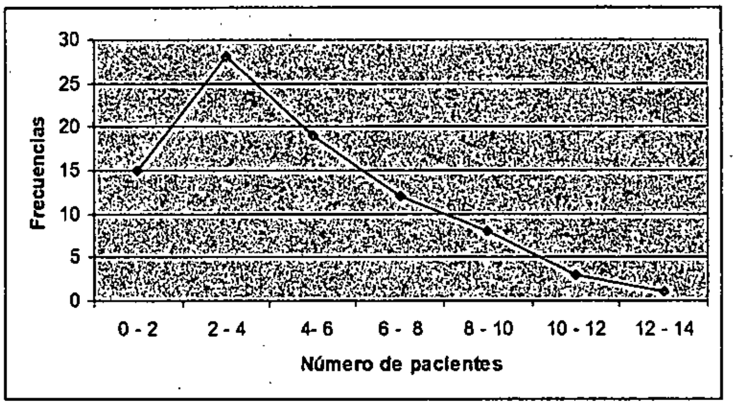


Figura N°7: El número de pacientes que ingresaron diariamente a la sala de urgencias del hospital Memorial

a) ¿Cuál es el punto medio de la clase 2 a 4? El punto medio o marca de clase es de 3.

b) ¿Cuántos días ingresaron de 2 a 4 pacientes?
Ingresaron aproximadamente de 25 a 30 pacientes

c) ¿Cuántos días se estudiaron? Se estudiaron aproximadamente 86 días.

d) ¿Cuál es el intervalo de la clase? El intervalo es de 2

e) ¿Cómo se llama esta grafica? Se llama 'Polígono de Frecuencia'

7. La siguiente grafica muestra el precio de venta (en miles de dólares) de casa vendidas en el área de Lima Metropolitana.

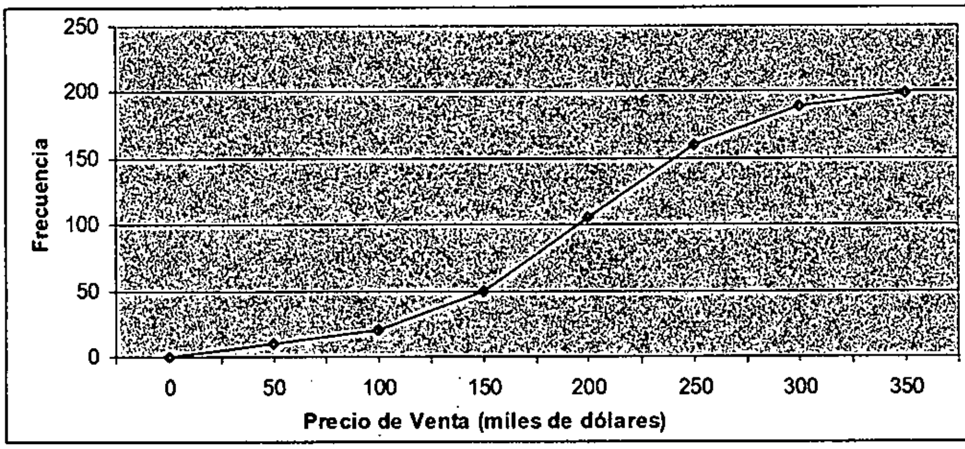


Figura N°8: Precio de venta de casa vendidas en el área de Lima Metropolitana.

a) ¿Cuántas casas se estudiaron?

Se estudiaron aproximadamente 200 casas

b) ¿Cuál es el intervalo de clase?

El intervalo de la clase o amplitud es de 50

c) ¿Cien casas se vendieron en una cantidad inferior a?

Se vendieron a una cantidad inferior de 200 mil dólares.

d) ¿Alrededor de 75% de las edificaciones se vendieron en menos de que cantidad?

Se vendieron en una cantidad mayor a 250 mil dólares.

Gráfico de barras

Consiste en una serie o conjunto de rectángulos que de acuerdo a su longitud y anchura representan un fenómeno, se puede utilizar para representar datos cualitativos y cuantitativos. Observaciones:

1. En el eje donde irá la base del rectángulo se especifican los indicadores o nombres que se usan para cada una de las bases.

2. La escala que se debe tomar para la base debe ser la misma para cada rectángulo.

3. La separación que exista entre las barras debe ser la misma, depende del

número de barras a construir y del espacio con que se cuenta.

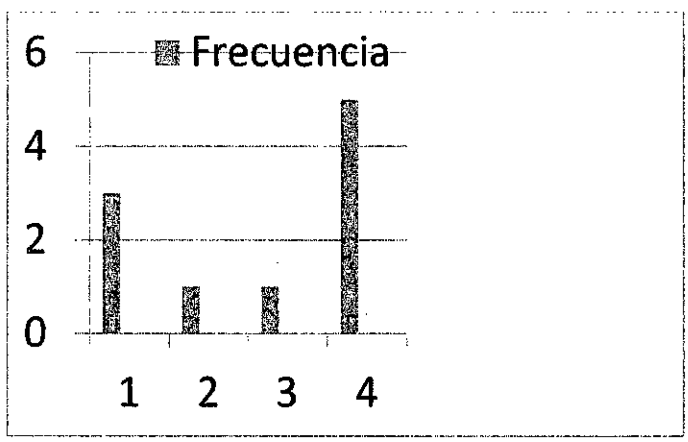
4. En el eje vertical se puede representar una escala de frecuencias, frecuencias relativas o de porcentajes.

Entre los tipos de gráficos de barra tenemos:

Gráficos de barras simples: son aquellos que representan una sola característica.

Ejemplo cuantitativo:

En una caja hay 10 bolitas marcadas con los números del 1 al 4. En la siguiente tabla se muestra la



Número	Frecuencia
1	3
2	1
3	1
4	5

Figura N°9: Bolitas marcadas con los números del 1 al 4

Histogramas

Se emplean cuando las variables son continuas, son gráficos de barras en los cuales no hay separación entre los rectángulos que se forman, se construye mediante la representación de las clases de una distribución de frecuencias en el eje horizontal y las frecuencias en el eje vertical. A través de él se pueden visualizar tres características de los datos: forma, acumulación o tendencia posicional y la dispersión o variabilidad

Pasos para la construcción:

1. Se trazan dos ejes de coordenadas sobre un plano.
2. Se llevan sobre el eje horizontal a los límites de clase.
3. En el eje vertical podemos representar no sólo el número de frecuencias,

también podemos colocar la proporción y el porcentaje de observaciones para

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

cada intervalo de clase, por eso tenemos varios tipos de nombres:

4. Se levantan perpendiculares por los límites de cada clase hasta la frecuencia de clase respectiva.

5. Se unen las dos perpendiculares que representan cada clase.

Observaciones:

1. Los histogramas no se pueden utilizar con respecto a distribuciones de frecuencias de clases abiertas (a menos que la persona cierre el intervalo de una manera conveniente).

2. El histograma representa las frecuencias de los intervalos mediante áreas y no mediante alturas; sin embargo, si los intervalos de clase tienen todos igual tamaño entonces el área de los rectángulos representa las frecuencias, por ello las alturas de los rectángulos son proporcionales a las frecuencias de clase y se acostumbra en tal caso a tomar las alturas numéricamente iguales a las frecuencias de clase.

Si los intervalos de clase no son de igual tamaño, las áreas no representan a las frecuencias, por lo tanto, es necesario ajustar la altura de los rectángulos (estas alturas deberán ser calculadas para que las superficies sean proporcionales a las frecuencias de clase).

Ejemplo 1: Los resultados de un test de inteligencia hecho a 25 personas se han registrado en la siguiente tabla de frecuencias por intervalos.

Tabla N°12: *Frecuencia absoluta simple de un test de inteligencia*

IC Puntaje	fi= N° de Personas
[64-73]	4
[74-83]	4
[84-93]	5
[94-103]	7
[104-113]	2
[114-123]	3

Construye un histograma de los datos propuestos.

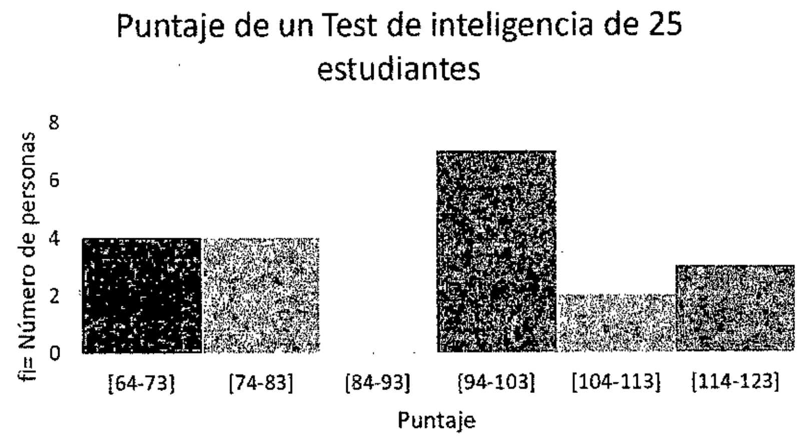


Figura N°10: Puntaje de un test de inteligencia de 25 estudiantes

Ejemplo 2: Se toma un examen de Expresión Gráfica y se les quiere clasificar en estudiantes deficientes en esta habilidad, Regulares, Buenos y Muy buenos, indique dicha clasificación. Elabore un cuadro de frecuencias estudiadas y un histograma incrustando un polígono de frecuencias en dicho histograma, cuyas notas están dadas en la siguiente tabla de las notas de un grupo de 120 estudiantes: Para agrupar estos datos recurrimos a la regla de Sturges cuya leyenda adjuntamos: Para estos datos observamos que el valor D es 6 estos valores estarán dentro del valor mínimo y máximo y elaboramos nuestro cuadro de frecuencias:

113	129	117	130	110	133	86	91	97	102
111	138	121	129	145	86	90	98	133	100
116	91	93	101	91	124	104	123	92	110
99	143	128	123	116	147	148	87	135	106
95	136	148	85	115	141	104	128	108	131
100	110	122	105	139	99	142	91	97	115
91	103	144	110	144	106	117	107	106	139
115	86	82	128	106	138	125	114	116	147
143	123	137	120	131	88	116	148	93	91
89	104	113	128	147	97	144	143	116	113
94	94	86	133	116	128	87	100	126	130
93	87	113	130	147	125	96	121	141	92

Vmax= 148 , Vmin= 82 ; R = 66 , k = 7.86129811 = 9 , TIC = 8.25 = 9

D = 6 , N= 120

Tabla N°13: *Tabla de frecuencias de notas del examen de Expresión Gráfica*

IC	fi	Fi	hi	Hi	Fi*
[79-88)	9	9	0.075	0.075	120
[88-97)	18	27	0.150	0.225	111
[97-106)	16	43	0.133	0.358	93
[106-115)	16	59	0.133	0.492	77
[115-124)	18	77	0.150	0.642	61
[124-133)	16	93	0.133	0.775	43
[133-142)	12	105	0.100	0.875	27
[142-151)	15	120	0.125	1	15

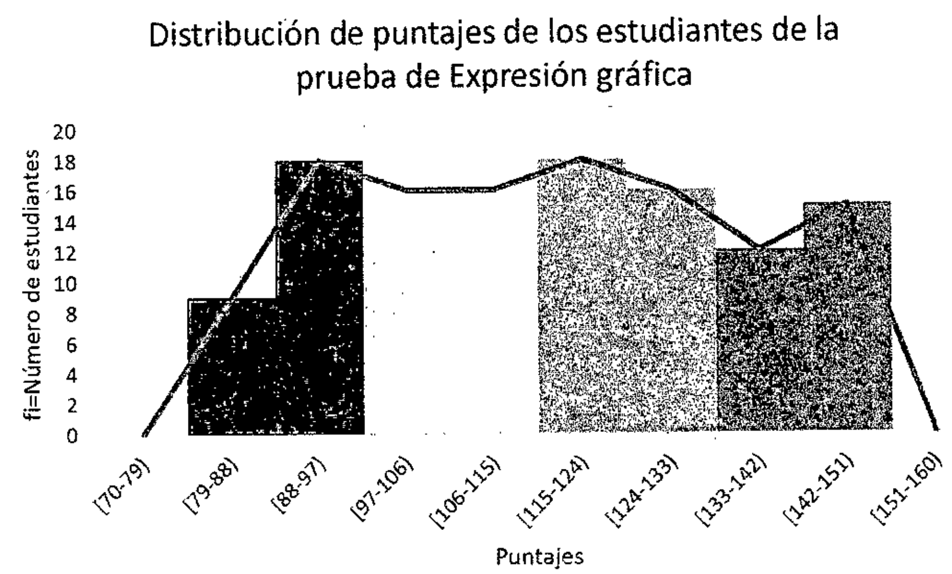


Figura N°11: Distribución de notas del examen de Expresión Gráfica

Otras Representaciones Gráficas:

1. Pictogramas: Se utilizan "dibujos" para representar la variable, no son precisos.

2. Cartogramas: Lo mismo que el anterior pero en vez de con dibujos con colores, se utilizan cuando se está estudiando sobre mapas. Los gráficos se pueden graficar en:

Sugerencias: El criterio del investigador junto con sus conocimientos matemáticos serán los encargados de determinar cuál es el gráfico más apropiado para cada conjunto de datos.

CAPITULO II

Medidas de Tendencia Central

2.1. Media aritmética, Mediana, Moda

El objetivo principal de las medidas de tendencia central es poder representar por medio de un solo número al conjunto de datos, es decir, dan valores representativos de la distribución de frecuencias, situados en algún lugar intermedio, alrededor del cual, se encuentran los otros valores. Nos indican dónde tienden a concentrarse los valores.

Existen tres medidas de tendencia central generales, que son, la Media aritmética, la Mediana y la Moda; así como otras que se utilizan en casos particulares como la Media ponderada, la Media Armónica, la Media Geométrica, la Media Cuadrática.

Medidas de tendencia central generales.

Media Aritmética $(\bar{x})$: Es el promedio de los datos, y su objetivo principal es encontrar el valor que debería de estar en el centro. Su ventaja principal es que es la única medida en la que $\sum(x-\bar{x})=0$, su inconveniente es que se ve influida por valores extremos.

Casi siempre, cuando nos referimos al "promedio" de algo, estamos hablando de la media aritmética. Esto es cierto en casos como la temperatura invernal promedio en la ciudad de Puno, la vida promedio de la batería del flash de una cámara o la producción promedio de maíz en una hectárea de tierra.

Propiedades de la media aritmética:

- Todo conjunto de datos de nivel de intervalo y de nivel de razón tiene un valor medio.
- Al evaluar la media se incluyen todos los valores.
- Un conjunto de datos sólo tiene una media. Esta es un valor único.
- La media es la única medida de ubicación en donde la suma de las desviaciones de cada valor con respecto a la media, siempre será cero.
- La media es una medida muy útil para comparar dos o más poblaciones.

Ventajas y desventajas de la media aritmética

La media aritmética, como un solo número que representa a un conjunto de datos completo, tiene importantes ventajas:

- Se trata de un concepto familiar para la mayoría de las personas y es intuitivamente claro.
- Cada conjunto de datos tiene una media; es una medida que puede calcularse y es única debido a que cada conjunto de datos posee una y sólo una media.
- La media es útil para llevar a cabo procedimientos estadísticos como la comparación de medias de varios conjuntos de datos

Sin embargo, como cualquier medida estadística, la media aritmética tiene desventajas que debemos conocer.

- Aunque la media es confiable en cuanto a que toma en cuenta todos los valores del conjunto de datos, puede verse afectada por valores extremos que no son representativos del resto de los datos. Observe que si los siete estudiantes resuelven un ejercicio con marcas de tiempo: **4.2 ; 4.3 ; 4.7 ; 4.8 ; 5.0 ; 5.1 ; 9.0**

  El tiempo medio es:

  $$\mathbf{M = (4.2 + 4.3 + 4.7 + 4.8 + 5.0 + 5.1 + 9.0) / 7 \quad = \quad 37.1 / 7 \quad = 5.3 \ minutos}$$

Sin embargo, si calculamos el tiempo medio para los primeros seis corredores y excluimos el valor de 9.0 minutos, la respuesta aproximada es 4.7 minutos. El valor extremo 9.0 distorsiona el valor que obtenemos para la media. Sería más representativo calcular la media sin incluir el valor extremo.

- Si alguno de los valores es extremadamente grande o extremadamente pequeño, la media no es el promedio apropiado para representar la serie de datos.
- La tercera desventaja es que somos incapaces de calcular la media para un conjunto de datos que tiene clases de extremo abierto en la parte inferior o superior de la escala.

Media Poblacional $\mu = \dfrac{\sum X}{N}$ Media muestral $\bar{x} = \dfrac{\sum X}{n}$

μ: Representa la media de la población.
N: Es el número total de elementos en la población.
x: Representa cualquier valor en particular de la población.

$\bar{X}$: Representa la media de la muestra.
n: Es el número total de elementos en la muestra.
x: Representa cualquier valor en particular de la muestra.

Datos No Agrupados (para una muestra):

$$\bar{x} = \frac{\sum x_i}{n}$$

X= cualquier dato

Número total de datos

La Media Aritmética para datos agrupados:

$$\bar{x} = \frac{\sum_{i=1}^{k} f_i * X_i}{n}$$

Frecuencia por la marca de clase de cualquier renglón

Número total de datos

donde: k = última clase

Nota: La media muestral se denota $\bar{x}$, la media poblacional se conoce como μ.

$\bar{x}$: Designa la media aritmética.

m_i : Es el valor central o punto medio de cada clase.

n : Es el número total de datos de la muestra.

f_i : Es la frecuencia en cada clase.

Ejemplo 1: Calcular la media aritmética de los números 10,12,36,25,58

$$\bar{x} = \frac{10+12+36+25+58}{5} = \frac{121}{5} = 24.2$$

Ejemplo 2. El siguiente cuadro presenta datos que describen el número de días que los generadores de una planta de energía del Cañón del Pato se encuentran fuera de servicio debido a mantenimiento normal o por alguna falla.

Tiempo sin funcionar los generadores	Generador	1	2	3	4	5	6	7	8	9	10
de la estación del Cañón del Pato	Días fuera de servicio	7	23	4	8	2	12	6	13	9	4

Solución: Para encontrar la media aritmética, sumamos los valores y dividimos el resultado entre el número de observaciones:

$$\bar{x} = \frac{\sum x_i}{n} \quad \frac{\textit{suma de todas las obsevaciones}}{\textit{número de elementos de la muestra}} \qquad \bar{x} = \frac{\sum x_i}{n} = \frac{88}{10} = 8.8$$

Ejemplo 3. Una empresa Textil las ganancias trimestrales en dólares durante los últimos tres años son las siguientes:

Tabla N°14: Tabla de ganancias trimestrales en dólares de una empresa Textil

	1er. Trimestre	2do. Trimestre	3er. Trimestre	4to. Trimestre	total
Año 1	10000	5000	25000	15000	55000
Año 2	20000	10000	20000	10000	60000
Año 3	30000	15000	45000	50000	140000
total	60000	30000	90000	75000	

a) Calcule por separado las ganancias promedio de la representante en cada uno de los cuatro trimestres.

b) Calcule por separado las ganancias trimestrales promedio en cada uno de los tres años.

c) Muestre que la media de las cuatro cantidades obtenida en el inciso a) es igual a la media de las tres cantidades que obtuvo en el inciso b). Además, muestre que estas dos cantidades son iguales a la media de los 12 números que se presentan en la tabla. (Ésta es la ganancia promedio trimestral que obtuvo la empresa durante un periodo de tres años.)

Solución:

a) 1er. Trimestre = 60000/3 =20000
   2do.Trimestre = 30000/3 =10000
   3er. Trimestre = 90000/3 =30000
   4to. Trimestre = 75000/3 =25000
b) Año 1 = 55000/4 =13750
   Año 2 = 60000/4 =15000
   Año 3 = 140000/4 =35000
c) MA= (20000+10000+30000+25000)/4=21250
   MA= (13750+15000+35000)/3=21250

MA(2#s)= (21250 + 21250)/2= 21250

MA(12#s)=(10000+5000+25000+15000+20000+10000+10000+10000+ 30000+15000+45000+50000) /12 = 30833.33(Ésta es la ganancia promedio trimestral que obtuvo la empresa durante un periodo de tres años.)

Ejemplo 4. El Servicio Postal de Lima maneja siete tipos básicos de cartas y tarjetas postales: tercera clase, segunda clase, primera clase, correo aéreo, entrega especial, correo registrado y correo certificado, se da en la siguiente tabla:

Tabla N°15: *El volumen de envíos de correo registrado durante 2017*

tipo de correo	onzas enviadas (en millones)	precio por cada onza
tercera clae	16400	0.05
segunda clase	24100	0.08
primera clase	77600	0.13
aéreo	1900	0.17
entrega especial	1300	0.35
registrado	750	0.4
certificado	800	0.45

¿Cuál es el ingreso promedio anual por cada onza de la prestación del servicio?

Tabla N°16: *El volumen promedio de envíos de correo registrado durante 2017*

	A	B	C	D	E	F	G
1	Envíos 1977 - 2019						
2	tipo de correo	onzas enviadas (en millones)	precio por cada onza	precio total de onzas			
3	tercera clae	16400	0.05	820		2019-1977 =	42
4	segunda clase	24100	0.08	1928			
5	primera clase	77600	0.13	10088			
6	aéreo	1900	0.17	323			
7	entrega especial	1300	0.35	455			
8	registrado	750	0.4	300			
9	certificado	800	0.45	360			
10	total	122850	1.63	14274			
11	anualmente	2925	0.038809524	339.8571429			

El ingreso promedio anual por onzas enviadas es de 2925 millones.

El ingreso promedio anual por cada onza de la prestación del servicio 0.038809524 dólares.

Ejemplo 4. Los gastos de publicidad son un componente importante en el costo de mercancías vendidas. A continuación, se presenta una distribución de frecuencias que muestran los gastos de publicidad de 60 compañías de productos alimenticios.

Tabla N°17: *Tabla gastos de publicidad de 60 compañías de productos alimenticios*

Gasto de publicidad (millones de $)	**Número de compañías (n)**	**Frecuencia acumulativa (F)**	**Marca de clase (m_i)**	**f_i * m_i**
25 – 35	5	5	30	150
35 – 45	10	15	40	400
45 – 55	21	36	50	1050
55 – 65	16	52	60	960
65 – 75	8	60	70	560
Total	60			3120

La media para datos agrupados es:

$$\bar{x} = \frac{\sum(f_i * m_i)}{n} = \frac{150+400+..+560}{60} = \frac{3120}{60} = 52$$

Ejemplo 5. Una empresa tabacalera tiene registros del costo de procesamiento de cada pedido. Durante los últimos 5 años, este costo fue de $55.00, $58.00, $61.00, $65.00 y $66.00. ¿Cuál fue el crecimiento porcentual promedio de la empresa durante este lapso? Si esta tasa promedio se mantiene estable durante 3 años más, ¿cuánto le costará a la empresa procesar un pedido al final de ese periodo?

Tabla N°18: *Tabla de registros del costo de cada pedido de una tabacalera*

B	C
registro	
año	costo
1	55
2	58
3	61
4	65
5	66
total	305
promedio=	61
	61.00%

E	F
registro	
año	costo
1	55
2	58
3	61
4	65
5	66
6	61
7	61
8	61
total	488
promedio=	61

El crecimiento porcentual promedio de la empresa durante este lapso fue de 61%

A la empresa le costará después de mantenerse estable durante 3 años más procesar un pedido al final de ese periodo un total de $488.

Ejemplo 6: calcular el salario promedio de:

Salario (X)	No. De emp. (F)
$15,000	18
$20,000	35
$25,000	29

Como $\sum f = 82 = n$ sustituimos en la formula y se

obtiene: $\bar{x} = \frac{(15000*18)+(20000*35)+(25000*29)}{82} = \frac{1695000}{82} = \$20,670.70$

Ejemplo 7. Si x es la media aritmética de los números r, s y t ¿Cuál (es) de las siguientes igualdades es (son) verdaderas?

I) $x = \frac{r+s+t}{3}$

II) $(x-r)+(x-s)+(x-t)=0$

III) $x+10 = \frac{r+s+t+10}{3}$

A) Solo I
B) Solo II
C) Solo III
D) Solo I y II
E) I, II y III

Solución : (Analizando cada alternativa)

I) Es verdadera por definición de media aritmética.

II) de $(x-r)+(x-s)+(x-t)=0$ se tiene: $x = \frac{r+s+t}{3}$ por lo tanto II) es verdadera.

IV) Despejando "x" en: $x+10 = \frac{r+s+t+10}{3}$ se tiene $x = \frac{r+s+t-20}{3}$ por lo tanto esta expresión no corresponde a la media aritmética por lo que III) no es verdadera.

Luego la solución es D)

Ejemplo 8. En un sector de una empresa trabajan 80 empleados cuya jubilación de su sueldo promedio es de $620 000. Si “x” empleados ganan $ 500 000 y el resto $900 000, entonces una ecuación para calcular “x” es:

A) $5x+9(80-x)=8*62$

B) $5x+8(80-x)=9*62$

C) $8x+6x(80-x)=9*5$

D) $9x+5(80-x)=8*62$

E) $62x+8(80-x)=9*5$

Solución

$\bar{x}=\dfrac{\sum_{i=1}^{n} x_i * f_i}{n}=620\ 000$, donde x_i son los grupos de sueldo y los f_i son los empleados de cada grupo, n = 80 el número total de empleados. Entonces tendremos:

$$\frac{500\ 000*x+900\ 000(80-x)}{80}=620\,000$$

Factorizamos por 100 000 y simplificamos por tal cantidad a cada lado de la igualdad obtenemos: $5x+9(80-x)=8*62$

Luego la respuesta es A)

LA MEDIANA(M_e)

Es el valor de la variable que ocupa la posición central, en un conjunto ordenado de datos. Se tiene que 50% de las observaciones se encuentran por arriba de la mediana, y 50% por debajo de ella (está en el centro de todas las observaciones).

Propiedades:

- Es única, es decir, sólo existe una mediana para un conjunto de datos.
- No se ve afectada por valores muy grandes o muy pequeños, y por lo tanto es una medida valiosa de tendencia central cuando ocurre este tipo de valores.

- Puede obtenerse para datos de nivel de razón, de intervalo y ordinal.
- En una distribución de frecuencias que tenga clases abiertas (por ejemplo: más de $65), la mediana no es afectada.

La mediana para datos agrupados:

$$Mediana = Li + \frac{\frac{N}{2} - F_{i-1}}{f_i} * C_j$$

Li : Es el límite inferior de la clase que contiene a la mediana.

F_{i-1} : Es el número acumulativo de frecuencias en todas las clases que preceden inmediatamente a la clase en cuestión con la mediana.

C_j : Es el ancho de la clase en que se encuentra la mediana.

N : Es el número total de frecuencias.

f_i : Es la frecuencia en cada clase.

Datos No agrupados: En los datos ordenados se aplica la siguiente relación, para encontrar la posición de los datos.

$posición = \frac{n+1}{2}$; en donde n = número total de datos

Entonces podemos tener sólo dos alternativas

a) El valor de la posición puede ser entero y lo único que debemos hacer es contar el número de lugares que nos indica esta fórmula.
b) El valor de la posición nos da un valor decimal (.5) y entonces debemos: sumar los valores involucrados y dividirlos entre 2. Por ejemplo; si tenemos los valores 5, 7, 8, 13 entonces la posición nos da 2.5 por que tendremos que seleccionar a los números 7 y 8 para luego sumarlos (15) y dividirlos entre 2 (7.5)

Datos Agrupados:

Se localiza la clase o renglón que contiene a la mediana, con la siguiente condición

$F_{i-1} \geq \frac{n+1}{2}$, es decir debemos encontrar la primer frecuencia acumulada que sea mayor o igual a la posición, para posteriormente aplicar la siguiente formula:

$$Me = F_i + \left(\frac{\frac{n}{2} - F_{i-1}}{f_i} \right) * IC$$ donde:

Fi	**Fi-1**	**fi**	**IC**
Frontera o límite verdadero inferior del renglón de la mediana	Frecuencia acumulada anterior al renglón de la mediana	Frecuencia del renglón de la mediana	Tamaño de intervalo de clase en el renglón de la mediana

Nota: Si la posición, en los datos no agrupados, es decimal (.5), se toma el promedio del dato anterior y el siguiente.

Ejemplo: Calcular el sueldo mediano de:

Tabla N°19: *Tabla de sueldos*

Intervalos de Clase ($)	Marca de clase (mi)	No. De emp. (fi)	Frecuencia acumulada Fi
12,500-17,500	$15,000	18	18
17,500-22,500	$20,000	35	53
22,500-27,500	$25,000	29	82
Total		82	

Primero se obtiene la posición:

$$posición = \frac{82+1}{2} = 41.5$$

Entonces buscamos el renglón de la mediana buscando la fa igual o más grande de 41.5, como 18+35 = 53, entonces decimos que es el segundo renglón o clase donde se encuentra la mediana y aplicamos la fórmula:

$$Me = F_i + \left(\frac{\frac{n}{2} - F_{i-1}}{f_i} \right) * IC = 17500 + \left(\frac{41.5\text{-}18}{35} \right) * 5000 = \$20,857.14$$

LA MODA (M_o)

La moda es el valor que aparece con mayor frecuencia.

Propiedades:

- La moda se puede determinar en todos los tipos de mediciones (nominal, ordinal, intervalo, y razón). Es especialmente útil para describir niveles nominales y ordinales de medición.
- La moda tiene la ventaja de no ser afectada por valores extremos.
- En una distribución de frecuencias que tenga clases abiertas (por ejemplo: más de $65), la moda no es afectada.

Desventajas:

- En muchas series de datos no hay moda porque ningún valor aparece más de una vez.
- En algunas series de datos hay más de una moda, en este caso uno podría preguntarse ¿Cuál es el valor representativo de la serie de datos?

Datos No Agrupados: Después de ordenar los datos buscamos el valor que más se repite.

Ejemplo: Encontrar la moda de; 47, 48, 49, 49, 49, 51, 51, 52. Podemos observar que el número que más se repite es el 49. Si ningún valor se repite, no existe moda

Datos Agrupados:

Se localiza la clase modal buscando la frecuencia más alta y después se aplica la siguiente fórmula:

$$M_o = L_i + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right) * IC$$

Li : Es el límite inferior de la clase que contiene a la mediana.

f : Es la frecuencia donde está la moda.

$\Delta_1 = f - f_{anterior}$, $\Delta_2 = f - f_{posterior}$

Nota: La distribución puede ser: amodal (no hay moda), unimodal (una sola moda), bimodal (dos modas), trimodal (tres modas),...., polimodal (muchas modas).

Ejemplo: Calcular el salario que más se repite en la tabla 16:

Intervalos de Clase ($)	Marca de clase (mi)	No. De emp. (fi)
12,500-17,500	$15,000	18
17,500-22,500	$20,000	35
22,500-27,500	$25,000	29

Observamos las frecuencias (No. de empleados) y decimos que la clase modal es la segunda, porque 35 es la frecuencia más grande y aplicamos:

$$M_o = F_i + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right) * IC = 17500 + \left(\frac{17}{17+6}\right) * 5000 = \$21{,}195.65$$

donde: $\Delta_1 = f - f_{anterior} = 35 - 18 = 17$

$\Delta_2 = f - f_{posterior} = 35 - 29 = 6$

Ejemplo 1: La edad de los residentes de mi barrio, tiene la siguiente distribución de frecuencias:

Clase	Frecuencia
47-51.9	4
52-56.9	9
57-61.9	13
62-66.9	42
67-71.9	39
72-76.9	20
77-81.9	9

Estime el valor modal de la distribución: usamos $M_o = L_i + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right) * IC$

Tabla N°20: *Tabla de las edades de los residentes*

		IC	fi	Fi		
46.9	51.9	[46.9-51.9)	4	4		
51.9	56.9	[51.9-56.9)	9	13		
56.9	61.9	[56.9-61.9)	13	26		
61.9	66.9	[61.9-66.9)	42	68	Moda	
66.9	71.9	[66.9-71.9)	39	107		
71.9	76.9	[71.9-76.9)	20	127	Δ1=	29
76.9	81.9	[76.9-81.9)	9	136	Δ2=	3
			136			
	Moda=	66.43125				

EJEMPLO 2:

Se muestra el cambio en porcentaje para el ingreso neto de 2010 a 2017, en el caso de una muestra de 12 compañías de construcción: 5, 1, -10, -6, 5, 12, 7, 8, 2, 5, -1 y 11. Calcule la mediana y la moda:

SOLUCIÓN:

Ordenando: -10, -6, -1, 1, 2, 5, 5, 5, 7, 8, 11, 12

Entonces la mediana es 5 y la moda es 5.

EJEMPLO 3:

Para que un producto sea aceptado por su cliente principal, debe cumplir con ciertas especificaciones de calidad. Una de ellas, radica en que el promedio de longitud de los 56 primeros productos este entre 20,0 y 20,9 centímetros. Evaluar la siguiente producción:

Tabla N°21: *Tabla de cantidades del producto sea aceptado por su cliente principal*

Longitud	**fi**	**Fi**	**hi**	**Hi**
Menores a 19,1	6	6	0,11	0,11
19,1 – 20,0	10	16	0,18	0,29
20,0 – 20,9	20	36	0,36	0,64
20,9 – 21,8	15	51	0,27	0,91
Mayores a 21,8	5	56	0,09	1,00

SOLUCION:

Calculamos la media y la moda; la media aritmética no se puede calcular por obvias razones. $Me = Li + \frac{\frac{n}{2} + F_{i-1}}{fi} * C_j = 20 + \frac{\frac{56}{2} - 16}{20} * 0,9 = 20,54$

$$Moda = 20 + \frac{20-10}{(20-10)+(20-15)} * 0.9 = 20,6$$

Entonces quiere decir que si pasa la inspección.

Relación entre Media Aritmética, Mediana y Moda:

Para distribuciones unimodales que sean poco asimétricas: $\overline{X} - M_o = 3(\overline{X} - M_e)$

Sus posiciones relativas, según la simetría de la distribución de frecuencias es:

Relación	Simetría
$M_o = M_e = \bar{x}$	Simétrica
$M_o < M_e < \bar{x}$	Sesgo positivo
$\bar{x} < M_e < M_o$	Sesgo negativo

Nótese que en nuestros ejemplos anteriores sobre salarios tenemos:

$M_0 > M_e > \bar{x}$ es decir $21195.65 > 20857.14 > 20670.7$ **Luego el sesgo es negativo**

Medidas de tendencia central para casos especiales

Media Aritmética Ponderada $(\bar{X}_p)$: Es el promedio de los datos en donde se le da un peso o importancia específica a cada observación. Se calcula:

Es una media aritmética que se emplea en las que se introducen unos coeficientes de ponderación, denominados wi, que son valores positivos, que representan el número de veces que un valor de la variable es más importante que otro.

$$\bar{x}_w = \frac{w_1 * X_1 + w_2 * X_2 + \ldots w_n * X_n}{w_1 + w_2 + \ldots + w_n} = \frac{\sum (wX)}{\sum w}$$

Donde:

- $\bar{x}_w$ es el símbolo de la media ponderada
- w es el peso asignado a cada observación.
- $\sum (w \times x)$ es la suma de los productos de la ponderación de cada elemento por el elemento correspondiente.
- $\sum w$ es la suma de todas las ponderaciones

$$\bar{X}_w = \frac{\sum_{i=1}^{n} W_i * X_i}{\sum_{i=1}^{n} W_i}$$

Producto de cada uno de los datos por su ponderación

Suma de las ponderaciones

Ejemplo: Se desea obtener el precio promedio de:

Precio del Producto	Cantidad en Kg.
$ 17.80	75
$ 35.90	56
$ 79.45	19

Aplicamos la fórmula:

$$\overline{X}_w = \frac{\sum_{i=1}^{n} W_i * X_i}{\sum_{i=1}^{n} W_i} = \frac{(17.8*75)+(35.9*56)+(79.45*19)}{75+56+19} = \frac{4854.95}{150} = \$32.37$$

Ejemplo 1.

Se quiere saber el promedio de un cachimbo al finalizar el primer ciclo, y sus notas son:

Tabla N°22: *Tabla de Notas de un estudiante de primer ciclo*

CURSOS	CRÉDITOS	NOTAS
Química General	4	15
Álgebra Lineal	3	16
Matemática	5.5	14
Sociología	3	11
Dibujo Técnico	4	12
Informática	3	13

Entonces la media ponderada es:

$$\overline{x}_w = \frac{4*15+3*16+5.5*14+3*11+4*12+3*13}{4+3+5.5+3+4+3} = 13.556$$

Ejemplo 2. La tienda Ripley tiene un aviso: "Si nuestros precios promedio no son iguales o menores que los de otros, usted se lo lleva gratis." Uno de los clientes de Ripley fue a la tienda un día y puso sobre el mostrador las notas de venta de seis artículos que compró a un competidor por un precio promedio menor que el de Ripley. Los artículos costaron (en dólares)

1.29 2.97 3.49 5 7.5 10.95

Los precios de Ripley de los mismos seis artículos son \$2.35, \$2.89, \$3.19, \$4.98, \$7.59 y \$11.50. Ripley le explicó al cliente: "Mi aviso se refiere a un promedio ponderado de estos artículos, nuestro promedio es menor porque nuestras ventas de estos artículos han sido:

7 9 12 8 6 3

¿Está Ripley buscando un problema o resolviéndolo al hablar de promedios ponderados?

Media Ponderada $\bar{x}_w = \dfrac{\sum_{i=1}^{n} w_i * x_i}{\sum_{i=1}^{n} w_i}$

Tabla N°23: *Tabla de promedios ponderados de aviso publicitario*

	Precios competencia	Precios de Dave (xi)	wi	xi x wi
1	1.29	2.35	7	16.45
2	2.97	2.89	9	26.01
3	3.49	3.19	12	38.28
4	5	4.98	8	39.84
5	7.5	7.59	6	45.54
6	10.95	11.5	3	34.5
	31.2		45	200.62

Prom. Ponderado Dave=	4.45822222
Prom. Articulos comprados por el cliente=	5.2

$\bar{x}_w = \dfrac{\sum_{i=1}^{n} w_i * x_i}{\sum_{i=1}^{n} w_i}$

$\bar{x} = \dfrac{\sum x_i}{n}$

Está resolviendo el problema, **pues es verdad**. El promedio ponderado de Ripley es:

$$\bar{x}_w = \frac{\sum_{i=1}^{n} w_i * x_i}{\sum_{i=1}^{n} w_i} = \frac{200.62}{45} = 4.46$$

Mientras que el promedio aritmético de los precios de la otra tienda es:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{31.2}{6} = 5.2$$

Ejemplo 3: La Línea 1 de Lima tiene, un despacho de asesoría financiera y administrativa, tiene cuatro tipos de profesionales entre su personal: asesores financieros, asociados principales, personal de campo y personal de oficina. Las tasas promedio que se cobran a los clientes por el desempeño de cada una de estas categorías profesionales son 75 dólares/hora, 40 dólares/hora, 30 dólares/hora y 15 dólares/hora, respectivamente.

Los registros de la firma indican el siguiente número de horas cobradas el año anterior en cada categoría: 8,000, 14,000, 24,000 y 35,000, respectivamente. Si La Línea 1 de Lima intenta formular una tasa de cobro promedio para estimar cuánto debe cobrar a los clientes en el año siguiente, ¿qué sugeriría que hiciera y cuál cree que sería una tasa apropiada?

Hallando el promedio ponderado de cobro/hora:

Tabla N°24: *Tabla de promedios ponderados de servicios profesionales*

	dólares/hora	#horas cobradas/año	cobro
asesor financiero	75	8000	600000
asociado principal	40	14000	560000
personal de campo	30	24000	720000
personal de oficina	15	35000	525000
	160	81000	2405000

Prom. De cobro /hora =	29.69135802
para una ganancia del 30% se debe cobrar/hora =	38.59876543

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

Ganancia de 30% = 29.69135802 + 0.3 *

$$\bar{x}_w = \frac{\sum_{i=1}^{n} w_i * x_i}{\sum_{i=1}^{n} w_i} = \frac{2405000}{8100} = 29.69$$

Por lo tanto, si se desea una ganancia del 30%, se debería cobrar/hora: $29.69 + 0.3 \times 29.69 = 38.60$

Ejemplo 4: El profesor de Estadística decide utilizar un promedio ponderado para obtener las calificaciones finales de los estudiantes que acuden a su seminario. El promedio de tareas tendrá un valor del 20% de la calificación del estudiante; el examen semestral, 25%; el examen final, 35%; el artículo de fin de semestre, 10%, y los exámenes parciales, 10%. A partir de los datos siguientes, calcule el promedio final para los cinco estudiantes del seminario.

Tabla N°25: *Calificaciones de 5 estudiantes y cálculo de promedio final*

Estudiante	Tareas	Parciales	Artículo	Ex. Semestra	Ex. Final	20%*Tareas	10% Ex. Parc.	10%*Articulo	25%*Ex. Sem	35%* Ex. Fin.	Suma de %
1	85	89	94	87	90	17	8.9	9.4	21.75	31.5	88.55
2	78	84	88	91	92	15.6	8.4	8.8	22.75	32.2	87.75
3	94	88	93	86	89	18.8	8.8	9.3	21.5	31.15	89.55
4	82	79	88	84	93	16.4	7.9	8.8	21	32.55	86.65
5	95	90	92	82	88	19	9	9.2	20.5	30.8	88.5
											441

Promedio Final = 441/ 5 = 88.2

Media Geométrica (MG): Con cierto tipo de datos, la media aritmética nos da el valor promedio correcto. La media geométrica sirve para promediar los crecimientos geométricos de una variable.

La media geométrica es útil para encontrar el promedio de porcentajes, razones, índices o tasas de crecimiento. Se utiliza ampliamente en los negocios y la economía porque frecuentemente interesa encontrar el cambio porcentual, tasas, números índices en ventas, sueldos, o cifras económicas, como el PBI. Es decir, en los casos en los que se supone que la variable presenta variaciones acumulativas

Si suponemos que **Y** representa el factor de crecimiento geométrico de la variable **X**, es decir: $Y_i = \frac{X_i}{X_{i-1}}$,entonces el factor de crecimiento geométrico promedio de la variable **X** será:

Datos No Agrupados: $G = \sqrt[n]{Y_1 * Y_2 * \cdots * Y_n}$

Ventajas e inconvenientes:

- En su cálculo intervienen todos los valores de la distribución.
- Los valores extremos tienen menor influencia que en la media aritmética.
- Además, cuando la variable toma al menos un Xi = 0, entonces se anula.
- Todos los valores de datos deben ser positivos para determinar la media geométrica.
- Si toma valores negativos se puede presentar una gama de casos particulares en los que tampoco queda determinada debido al problema de raíces de índice par de números negativos.

Ejemplo 1: Supóngase que recibe un aumento de 5% en su sueldo este año, y recibirá uno de 15% el próximo año. El aumento porcentual promedio es 9.8863% y no 10%.

Tabla N°26: *Tabla de aumentos*

	5% y 10%	**Promedio (9.8863%)**
Aumento 1	3 000 * (0.05) = 150.00	3 000 * (9.8863%) = 296.589
Aumento 2	3 150 * (0.15) = 472.50	3 296.589 * (9.8863%) = 352.911
Aumento Total	150.0 + 472.5 = 622.50	296.589 + 352.911 = 622.50

$MG = \sqrt[2]{1.05*1.15} = 1.098863$

Lo anterior se puede verificar suponiendo que su ingreso mensual inicial era de $3000 y que recibió 2 aumentos de 5% y 15%.

Otra aplicación de la media geométrica es encontrar un aumento porcentual promedio en un intervalo de tiempo.

$$MG = \sqrt[n]{\frac{Valor\_al\_final\_del\_periodo}{Valor\_al\_principio\_del\_periodo}} - 1$$

Ejemplo2. Si se gana $30 000 al año, en 2008 y $50 000 en el año 2018, ¿Cuál es la tasa de aumento anual en el periodo?

Solución: n = 2018 – 1998 = 10 años

Valor al principio del periodo = 30 000. Valor al final del periodo = 50 000.

$$MG = \sqrt[10]{\frac{50000}{30000}} - 1 = 5.24098\%$$

La tasa aumento anual promedio es de 5.24098%

Ejemplo 3. La siguiente tabla muestra la tasa de aumento en las quejas en el Municipio del Callao en el 2019 durante los últimos meses. Calcule e interprete la tasa media mensual.

Meses	**Enero**	**Febrero**	**Marzo**	**Abril**	**Mayo**
Aumento de quejas	2.6%	5.4%	3.8%	0.5%	1.4%

La tasa 2.6% también se puede expresar como 0.026, y puesto que se refiere a un aumento a partir de una base de 100%, el factor de variación será 1.026. Para los otros datos se opera igual. Por lo tanto, la media geométrica se calcula:

$$MG = \sqrt[5]{(1.026)(1.054)(1.038)(1.005)(1.014)}$$

$MG = \sqrt[5]{1.143903377}$ entonces se tiene: $MG = 1{,}0272540$ (*Factor de crecimiento medio*)

Tasa media de variación es: MG = (1,0272540 - 1) x 100 = 2,72%

Interpretación:

Si se selecciona al azar un mes entre enero y mayo, se espera que las quejas se hayan incrementado 2.72% con respecto al mes anterior.

Ejemplo 4. Si los precios de la acción del grupo "Gloria" en los últimos cuatro días fueron; 4.75, 5.23, 4.78 y 6.32 millones de dólares, calcular el factor de crecimiento promedio y el crecimiento porcentual promedio.

Existen dos formas de resolverlo:

a) De la forma más ortodoxa, es decir:

$$G = \sqrt[n]{Y_1 * Y_2 * \ldots * Y_n} = \sqrt[3]{\frac{5.23}{4.75} * \frac{4.78}{5.23} * \frac{6.32}{4.78}} = \sqrt[3]{1.330526316} = 1.099869493$$

Lo que acabamos de obtener es factor de crecimiento promedio y para obtener el crecimiento se aplica la siguiente formula:

$$crecimiento = (1 - G) * 100 = (1 - 1.099869493) * 100 = 9.9869\%$$

b) Otra forma es $G = \sqrt[\text{número de datos -1}]{\frac{\text{último}}{\text{primero}}} = \sqrt[3]{\frac{6.32}{4.75}} = \sqrt[3]{1.330526316} = 1.099869493$

Ejemplo 5.

El crecimiento en el gasto por deudores morosos del Banco WISSE durante los años 90 es el siguiente. Calcule el incremento promedio porcentual del gasto por deudores morosos durante ese periodo. Si esta tasa continúa, estime el incremento porcentual para 1997 respecto a 1995.

1989	1990	1991	1992	1993	1994	1995
0.11	0.09	0.075	0.08	0.095	0.108	0.120

Primero se determina el factor de crecimiento para cada año

$\rightarrow 1 + \frac{\text{tasa de interés}}{100}$ Luego, se halla la media Geométrica:

Tabla N°27: *Tabla de promedio porcentual del gasto por deudores*

año	gasto	tasa de interés (%)	Factor de crecimiento
1989	0.11	11	1.11
1990	0.09	9	1.09
1991	0.075	7.5	1.075
1992	0.08	8	1.08
1993	0.095	9.5	1.095
1994	0.108	10.8	1.108
1995	0.12	12	1.12
1997	0.145032	14.5032	1.145032

$1 + \frac{\text{tasa de interés}}{100}$

Media Geom. =	1.096750035
promedio=	9.68%
	13.161

c)

Número de valores

$M.G. = \sqrt[x]{\text{producto de todos los valores}}$

MG= 1.09675 *(la tasa de interés promedio anual es de 9.675%)*

Con este valor, se puede estimar el incremento porcentual para 1997 con respecto a 1995: Tasa de interés en 1995: 12%

Tasa de interés en 1997 con respecto a 1995: 12 + 0.09675*12 = 13.161%

Datos Agrupados: $G = \sqrt[n]{Y_1^{f_1} * Y_2^{f_2} * \cdots * Y_k^{f_k}}$ donde: k = última clase

Nota: Se puede demostrar que $\overline{X} \geq G$. También puede calcularse la media geométrica ponderada.

Ejemplo 6. Supóngase que se cuenta con la información diaria de los incrementos porcentuales de una acción de Telefónica y que se representan en la siguiente tabla:

Crecimiento porcentual (%)	Frecuencias en días
10	14
20	15
30	48

a) Calcular los factores de crecimiento.

Se calcula con: $y = 1 + \left(\frac{crecimiento\ porcentual}{100}\right)$

b) Calcular el factor de crecimiento promedio

$$G = \sqrt[n]{Y_1^{f_1} * Y_2^{f_2} * \ldots * Y_k^{f_k}} = \sqrt[77]{1.10^{14} * 1.20^{15} * 1.30^{48}} = 1.2415965$$

Ejemplo 4. La tasa de crecimiento en 60 empresas se muestra en el siguiente cuadro: Encontrar la media geométrica en un intervalo de tiempo.

Tabla N°28: *Tabla de tasa de crecimiento en 60 empresas*

Intervalo de clase (% de crecimiento)	**fi**	**Marca de clase (mi)**
0% - 10%	5	5%
10% - 20%	10	15%
20% - 30%	30	25%
30% - 40%	15	35%

$$MG = \sqrt[60]{1.05^5 * 1.15^{10} * 1.25^{30} * 1.35^{15}} - 1 = 23.86\%$$

Media Armónica (H): Cuando los datos a promediarse están medidos en unidades expresadas en forma de cocientes (km./hr., $/lt, etc.), lo más adecuado es utilizar la media armónica, ya que la media aritmética nos llevaría a un promedio equivocado.

Se suele utilizar para promediar variables tales como productividades, velocidades, tiempos, rendimientos, cambios, etc.

Ventajas e inconvenientes:

- En su cálculo intervienen todos los valores de la distribución.
- Su cálculo no tiene sentido cuando algún valor de la variable toma cero o próximo a cero.
- Es única.

Datos No Agrupados: $$H = \frac{n}{\sum_{i=1}^{n} \frac{1}{X_i}}$$

Ejemplo 1. Si un vehículo se mueve de la ciudad A a la B a 65 Km./hr y regresa de B a A a 98 Km./Hr a qué promedio se desplazó. $H = \frac{n}{\sum_{i=1}^{n} \frac{1}{X_i}} = \frac{2}{\frac{1}{65} + \frac{1}{98}} = 78.1595$

Ejemplo 2. Se cronometra el tiempo de un repartidor de pizzas, se toma 10 veces (en minutos):

50 60 55 69 63 52 54 58 60 67

Entonces:

$$H = \frac{10}{\frac{1}{50}+\frac{1}{60}+\frac{1}{55}+\frac{1}{69}+\ldots+\frac{1}{58}+\frac{1}{60}+\frac{1}{67}} = 58.21$$

Ejemplo 3: Los siguientes datos registran el tiempo (en minutos) que utilizan 60 médicos al realizar una cierta intervención quirúrgica.

Tabla N°29: *Tabla de datos del tiempo que usan los médicos en una intervención quirúrgica*

Tiempo de demora	Xi	Cantidad de médicos
30 – 45	37.5	20
45 – 60	52.5	30
60 – 75	67.5	10

$$\overline{X}_h = \frac{60}{\frac{20}{37.5}+\frac{30}{52.5}+\frac{10}{67.5}} = 47.89\text{min}$$

Datos Agrupados: $H = \dfrac{n}{\sum_{i=1}^{k} \frac{f_i}{X_i}}$ donde: k = última clase

Nota: Se puede demostrar que $\overline{X} \geq G \geq H$.

También puede calcularse la media armónica ponderada.

Ejemplo 4: Supóngase que una flotilla de vehículos muestra la siguiente información:

Tabla N°30: *Tabla de datos de velocidad de una flota de autos*

Velocidad promedio en km/hr	Número de vehículos
50	15
60	28
75	31

La respuesta es:

$$H=\frac{n}{\sum_{i=1}^{k}\frac{f_i}{X_i}}=\frac{74}{\frac{15}{50}+\frac{28}{60}+\frac{31}{75}}=62.711864$$

Media Cuadrática (MC):

La media cuadrática tiene como objetivo de poder obtener el promedio de valores positivos y negativos al mismo tiempo, y será una gran ayuda para poder calcular las dispersiones promedio de los datos (ver medidas de dispersión).

Datos no agrupados: $MC=\sqrt{\frac{\sum_{i=1}^{n}x_i^2}{n}}$ La media cuadrática es igual a la raíz cuadrada de la suma de los cuadrados de los valores dividida entre el número de datos:

Ejemplo:

Supóngase que se obtienen las ganancias y pérdidas del precio de una acción de ADIDAS durante una semana; - 4.00, - 3.50, 2.35, 6.20, 3.25

Calcular el promedio:

$$MC=\sqrt{\frac{\sum_{i=1}^{n}x_i^2}{n}}=\sqrt{\frac{(-4.0)^2+(-3.5)^2+2.35^2+6.2^2+3.25^2}{5}}=\sqrt{\frac{50.775}{5}}=3.186691$$

Datos agrupados: $MC=\sqrt{\frac{\sum_{i=1}^{n}f_ix_i^2}{n}}$

Ejemplo: Ahora deseamos obtener el promedio de una tabla de distribución de frecuencias, pero con datos positivos y negativos.

Tabla N°31: *Tabla de ganancias y pérdidas de una acción según los días*

Ganancias y pérdidas del precio de una acción (x)	N° de días (f)
-7.25	25
2.75	14
12.75	2

$$MC = \sqrt{\frac{\sum_{i=1}^{n} f_i x_i^2}{n}} = \sqrt{\frac{25*(-7.25)^2 + 14*2.75^2 + 2*12.75^2}{41}} = 6.5239$$

Ejercicios de Aplicación

1.- Calcule la media geométrica de los valores que siguen: 2, 8, 6, 4, 10, 6, 8, 4.

Solución

$MG = \sqrt[8]{2.8.6.4.10.6.8.4} = \mathbf{5.41}$

2. En 2008 los ingresos obtenidos en juegos de azar fueron $651 millones de dólares. En el año 2011 el ingreso aumentó a $2.4 mil millones. ¿Cuál es la media geométrica del aumento anual de este periodo?

Solución

$n = 2011 - 2008 = 3$ años

$valor\ inicial = 651$

$valor\ final = 2400$

$$MG\ aumento\ anual = \sqrt[3]{\frac{2400}{651}} - 1 = 0.55\ millones\ de\ dólares$$

El aumento anual vendría a ser del 55%.

3. En 2008 había 9.19 millones de suscriptores de televisión por cable. En el año 2000 el número había aumentado a 54.87 millones ¿Cuál es la media geométrica del aumento anual en tal periodo?

Solución

$n = 2018 - 2008 = 10$ años

$valor\ inicial = 9.19$

$valor\ final = 54.7$

$$MG\ aumento\ anual = \sqrt[10]{\frac{54.7}{9.19}} - 1 = 0.196$$

El aumento anual vendría a ser del 19.6%.

4. En el cuadro que se muestra a continuación se indica el costo (en dólares) de un año de estudio en una universidad pública y en una privada, en 2010 y en 2018. ¿Cuál es la media geométrica del incremento anual en tal periodo en ambas universidades?

Solución

Tabla N°32: *Tabla de costos de en una universidad pública y en una privada*

Tipo de Universidad	2010	2018
Publica	\$4 975	\$7 628
Privada	\$12 284	\$19 143

$n = 2018 - 2010 = 8\, años$

Para la universidad pública:

$$MG = \sqrt[8]{\frac{7\,628}{4\,975}} - 1 = 0.055$$

Para una universidad privada:

$$MG = \sqrt[8]{\frac{19\,143}{12\,284}} - 1 = 0.057$$

La tasa de incremento anual en una universidad ***pública es de 5.5%****, en tanto en una universidad* ***privada es de 5.7%.***

5. José Castillo es estudiante en la escuela de Ingeniería Industrial. El semestre pasado se escribió en cursos de estadística y contabilidad, 3 horas de cada uno, y obtuvo una calificación de A en ambos. Recibo además una B en un curso de historia de 5 horas y una B en un curso de Inglés, de dos horas. Además, tomo un curso electivo. En este curso obtuvo una A ¿Cuál fue su promedio ponderado en el semestre? supóngase que recibe 4 puntos por una A, 3 puntos por una B, etc.

Solución

$$Promedio\ Ponderado = \frac{3h*4+3h*4+5h*3+2h*3+1h*4}{14h} = \mathbf{3.5\ puntos.}$$

6. Se estima que en el Perú mostrara el mayor aumento en el número de empleos por planilla entre los años 2019 - 2030. Es de esperarse que el número de empleos aumente

de 5 164 900 hasta 6 286 800. ¿Cuál es la tasa de incremento anual media geométrica esperada?

$n = 2030 - 2019 = 21$ años

$valor\ inicial = 5\,164\,900$

$valor\ final = 6\,286\,800$

$$MG = \sqrt[21]{\frac{6\,286\,800}{5\,164\,900}} - 1 = 0.00998$$

El aumento anual vendría a ser del 0.998%.

7. El Profesor César Miranda en un estudio reciente indicó si una persona gana S/. 25000 al año el día de hoy y la tasa de inflación continua es de 3% al año, la misma persona necesitara ganar S/. 33 598 en 10 años para tener el mismo poder adquisitivo. Necesitará ganar S/. 44 771 si la tasa de inflación aumenta en 6%. Confirme que estas declaraciones son exactas calculando la media geométrica de la tasa en aumento.

Solución

Primer caso:

$$MG1 = \sqrt[10]{\frac{33598}{25000}} - 1 = 0.0301$$

Entonces se comprobaría que la tasa de inflación es de 3.01% anual.

Segundo caso:

$$MG2 = \sqrt[10]{\frac{44771}{25000}} - 1 = 0.05995$$

Del mismo modo en este caso la tasa es 5.995% anual.

8. Los rendimiento de 12 meses de cinco fondos mutualistas de crecimiento dinámico fueron 32.2%, 35.5%, 80%, 60.9% y 92.1%. Determine la media aritmética y la media geométrica de las tasas de rendimiento.

$$\bar{x} = \frac{\sum x}{n} = \frac{300.7\%}{5} = \mathbf{60.14\%}$$

$$MG = \sqrt[5]{32.2x35.5x80x60.9x92.1} = \sqrt[5]{512921772.72} = \mathbf{55.20912\%}$$

2.2. Percentiles

Medidas de posición

Los cuartiles

Estos dividen al conjunto de observaciones en 4 partes iguales, esto es cuando están ordenados.

El primer cuartil (Q1).- El 25% de las observaciones son menores y el 75% son mayores que este valor. Se tiene que Q1 puede ser considerado como la "mediana" de la mitad inferior de los datos.

El segundo cuartil (Q2).- Más conocido como la mediana, el 50% son menores y el 50% son mayores que este valor.

El tercer cuartil (Q3).- Es el valor del cual el 75% de las observaciones son menores y el 25% son mayores, que este valor. Se tiene que Q3 puede ser considerado como la "mediana" de la mitad superior de los datos.

Los valores de Q1, Q2, Q3 dividen a un conjunto de datos en 4 partes iguales.

DECILES.-

De modo similar los deciles dividen a un conjunto de observaciones en 10 partes iguales.

Deciles (D): Representan el 10%, 20%, ... , 90% de los datos acumulados respectivamente.

PERCENTILES.-

Los percentiles o centiles dividen a un conjunto de observaciones en 100 partes iguales.

Percentiles (P): Representan el 1%, 2%, ... , 99% de los datos acumulados respectivamente.

Cada cuantil delimita dos regiones:

- el p% de datos de menor valor (acumulados a la izquierda del cuantil C) cuando se quiere calcular el puntaje máximo inferior de una muestra.
- el (1-p)% de datos de mayor valor (acumulados a la derecha del cuantil C), cuando se quiere calcular el puntaje mínimo superior de una muestra.

Datos No Agrupados:

En los datos ordenados: se debe calcular la posición mediante la formula:

$$Posición = \frac{j*(n+1)}{r}$$

donde:

$j =$ *Número* de cuantil que se desea obtener

r = puede ser 4, 10 o 100 depende del cuantil que se desee obtener

n = número de datos

Después de calcular la posición se utiliza la siguiente fórmula para encontrar el cuantil deseado. A esto se le llama interpolación

$$\text{dato menor} + (\text{dato mayor - dato menor}) * \text{fracción de la posición}$$

Ejemplo 1:

- Si me dijeran que mi talla se encuentra en el octavo decil, esto quiere decir yo soy más alto que el 80% de la población y solo hay un 20% que es más alto que yo.
- Si me dijeran que mi talla se encuentra en el tercer cuartil, esto quiere decir yo soy más alto que el 75% de la población y solo hay un 25% que es más alto que yo.
- Si me dijeran que mi talla se encuentra en el centil 60, esto nos quiere decir yo soy más alto que el 60% de la población y solo hay un 40% que es más alto que yo.

Para ubicar el centil deseado, utilizaremos la siguiente fórmula:

$$\mathbf{Lp} = (\mathbf{n}+\mathbf{1}) * \frac{\mathbf{porcentaje}}{\mathbf{100}}$$

Esto quiere decir que si quisiéramos encontrar la mediana (el quinto decil, o el centil 50) tendríamos que poner así: $\mathbf{L_{50}} = (\mathbf{n}+\mathbf{1}) * \frac{\mathbf{50}}{\mathbf{100}}$ esta fórmula ayuda a localizar el valor de la variable que acumula cierto porcentaje específico de datos.

Ejemplo 1. Una muestra de familias que están suscritas a la compañía Telefónica registró los siguientes números de llamadas recibidas la semana pasada:

52	44	30	38	31	42	12	46	39	37	34	46	32	18	41	5

Entonces ordenamos:

5	12	18	30	31	32	34	37	38	39	41	42	44	46	46	52
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

Aquí interpolamos para ver la ubicación donde se encuentra los cuartiles

Primer Cuartil	**La Mediana**	**Tercer Cuartil**
$L_{25} = (16+1)*\frac{25}{100} = 4.25$	$L_{50} = (16+1)*\frac{50}{100} = 8.5$	$L_{75} = (16+1)*\frac{75}{100} = 12.75$

Para el primer cuartil, este se encuentra en 4.25, es decir entre 30 y 31. Entonces:

$$\frac{Q_1-30}{4.25-4} = \frac{31-30}{5-4}$$

Esto nos da Q1 = 30.25. Esto hubiera sido más rápido si lo halláramos así:

$$Q_1 = 30+(31-30)*0.25 = 30.25$$

Ahora para la mediana, esta se encuentra en 8.5, es decir entre 37 y 38. Entonces:

$$Mediana = Q_2 = P_{50} = 37+(38-37)*0.5 = 37.5$$

Ahora para el tercer cuartil, esta se encuentra en 12.75, es decir entre 42 y 44. Entonces:

$$Q_3 = 42+(44-42)*0.75 = 43.5$$

Ejemplo 2: Se tiene dos grupos de trabajadores textiles en dos locales el grupo 1 es de solteros y el grupo 2 es de casados y el cuadro que se da son de horas extras en un mes

G1=	14	13	20	20	19	24	16	22	20	23	23	15	15	18	24	20
G2=	29	59	35	27	37	97	70	20	41	67	94	66	62	70	37	85

El gerente quiere sacar los siguientes percentiles: 11, 14, 24 , 35, 55, 75, 85, 93 para ver el compromiso de los dos grupos con la empresa.

Solución

Tabla N°33: *Números de horas extras de dos grupos de trabajadores con Excel*

orden	G1
1	13
2	14
3	15
4	15
5	16
6	18
7	19
8	20
9	20
10	20
11	20
12	22
13	23
14	23
15	24
16	24

Ord	G2
1	20
2	27
3	29
4	35
5	37
6	37
7	41
8	59
9	62
10	66
11	67
12	70
13	70
14	85
15	94
16	97

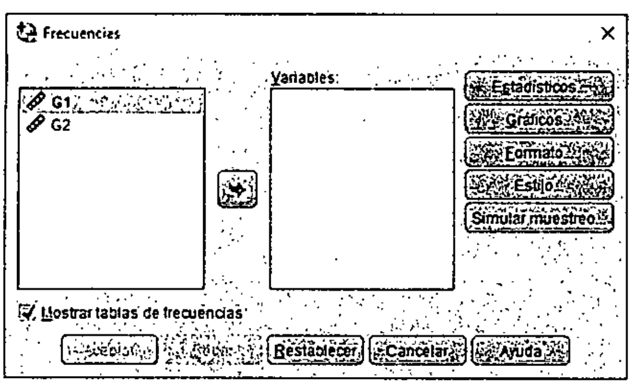
Solución Grupo 1

Percentiles		xi	Ubicación donde está el percentil xi *(n+1)/100
P85=	23.45	85	14.45
93	24	93	15.81
75	22.75	75	12.75
55	20	55	9.35
35	17.9	35	5.95
24	15.08	24	4.08
14	14.38	14	2.38
11	13.87	11	1.87

Solución Grupo 2

Percentiles		xi	Ubicación donde está el percentil xi *(n+1)/100
P85=	89.05	85	14.45
93	96.43	93	15.81
75	70	75	12.75
55	63.4	55	9.35
35	37	35	5.95
24	35.16	24	4.08
14	27.76	14	2.38
11	26.09	11	1.87

Tabla N°34: *Números de horas extras de dos grupos de trabajadores con SPSS*



Estadísticos

		G1	G2
Percentiles	11	13.8700	26.0900
	14	14.3800	27.7600
	24	15.0800	35.1600
	35	17.9000	37.0000
	55	20.0000	63.4000
	75	22.7500	70.0000
	85	23.4500	89.0500
	93	24.0000	96.4300

Obtenemos los mismos valores tanto en Excel como en SPSS.

Ejemplo 3: El Coordinador del centro de cómputo de la FIIS ha decidido desarrollar un nuevo programa de software diseñado para los docentes usuarios del Laboratorio. El Coordinador no desea desarrollar un programa que requiera demasiado espacio en el disco duro, por lo que sondearon a 36 docentes para determinar la cantidad de espacio disponible en sus computadoras. Los resultados en megabytes son los siguientes:

6.3	6.7	7.9	8.4	9.7	10.6	12.4	19.4	29.1	42.6
59.8	97.6	100.4	120.6	135.5	148.6	178.6	200.1	229.6	284.6
305.6	315.6	325.9	347.5	358.6	397.8	405.6	415.9	427.8	428.6
439.5	440.9	472.3	475.9	477.2	502.6				

Calcule el rango y el rango intercuartil agrupándolos usando la regla de Esturges. n = 36

Vmáx=	502.6	
Vmin=	6.3	
Rango=	496.3	
K=	6.135798253	7
TIC=	70.9	71
D=	0.7	71.7

Tabla N°35: *Espacio usado en las PC en el centro de cómputo por los docentes*

IC	fi	Fi
[6-77)	11	11
[77-148)	4	15
[148-219)	3	18
[219-290)	2	20
[290-361)	5	25
[361-432)	5	30
[432-504)	6	36
Total	36	

Rango

Rango = valor de la observación más grande − valor de la observación más pequeña

Rango = 504-6=498

Rango Intercuartil

Rango intercuartil = $Q_3 - Q_1$

Q3 =389.4 y Q1= 64.09090909 Rango intercuartil= 325.309

Ejemplo 4: Dados los números 3, 5, 7, 36, 45; obtener el número que represente al 75% de los datos.

Solución:

Primero obtienes la posición

Número de datos = 5; Percentil J = 75; Porciento R = 100, ubicación $\frac{75*(5+1)}{100} = 4.5$

Luego Identificamos que números están en la cuarta y quinta posición, es decir el 36 y el 45

Finalmente aplicamos la fórmula: $36+(45-36)*0.5=40.5$

es decir, el número que representa al 75% de los datos es el 40.5

Percentiles para datos agrupados:

Primero calculamos la posición en los datos, después buscamos la frecuencia acumulada anterior al reglón donde está el percentil que es F_{i-1}, y aplicamos la siguiente formula:

Posición del percentil

$$P_i = L_{inf} + \left(\frac{\left(\frac{n}{100}\right)*i - F_{i-1}}{f} \right)*IC$$

Frontera inferior   Frecuencia del renglón seleccionado

Ejemplo 1.

Encontrar el cuartil 3 de la siguiente tabla que es el percentil 75

Tabla N°36: *Tabla de datos agrupados para una muestra de 1099*

Fronteras	Frecuencia	Fa
100 - 200	389	389
200- 300	258	647
300 - 400	452	1099

$$P_{75} = L_{inf} + \left(\frac{\left(\frac{n}{100}\right)*i - F_{i-1}}{f} \right)*IC = =300+\left(\frac{\left(\frac{1099}{100}\right)*75-647}{452} \right)*100=339.214602$$

Ejemplo 2. Las edades de una muestra de estudiantes que asisten al curso de Estadística turno noche este semestre son:

19	17	15	20	23	41	33	21	18	20
18	33	32	29	24	19	18	20	17	22
55	19	22	25	28	30	44	19	20	39

a) Construya una distribución de frecuencias con intervalos:

15-19, 20-24, 25-29, 30-34 y 35 o más.

Solución: Los intervalos cerrados y los intervalos deben ser semi abiertos:

Tabla N°37: *Tabla del número de estudiantes de Estadística I según su edad*

IC	fi
[14.5-19.5)	6
[19.5-24.5)	12
[24.5-29.5)	3
[29.5-34.5)	5
[34.5-39.5)	4
	30

b) Estime el valor de la moda

$$Mo = L_{inf} + \left[\frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})} \right] C_i = 21.5$$

c) Ahora calcule las medidas de tendencia central y los percentiles 19, 22, 31, 38, 45, 64, 77, 87, 92 de los datos sin procesar. Para esto usamos SPSS

Tabla N°38: *Tabla de resultados usando SPSS*

Estadísticos		
Media		25.3333
Mediana		21.5000
Moda		19 y 20
Percentiles	19	18.0000
	22	18.8200
	31	19.0000
	38	20.0000
	45	20.0000
	64	24.8400
	77	31.7400
	87	38.8200
	92	42.5600

a. Existen múltiples modos.
b. Podemos observar que es bimodal

d) Compare sus repuestas a los incisos b) y c) y comente cuál de las dos medidas de tendencia central es más adecuada para estos datos y por qué.

Solución: La más adecuada es la Mediana, puesto a que para la media, si se utiliza la marca de clase que es la semisuma de los valores extremos de cada intervalo, se tendrían cálculos erróneos y observamos en los datos de la tabla en a) se agrupan en el segundo intervalos. Mientras que para la Moda también se da en el segundo intervalo de a), no es recomendable.

Ejemplo 3: Se dan los siguientes datos de gastos en obras del Municipio de Surco en el 2018

253	104	633	57	500	201	43	380	467	162	220	302

Calcules lo siguiente : EL rango los cuartiles la Media aritmética los percentiles 21, 23, 34, 45, 58, 65, 77, 78, 89.

Solución: Ordenando tenemos:

1	2	3	4	5	6	7	8	9	10	11	12
43	57	104	162	201	220	253	302	380	467	500	633

Respuestas

Tabla N°39: *Tabla de resultados usando SPSS*

Estadísticos

x

N	Válido	12
	Perdidos	0
Media		276,8333
Rango		590,00
Percentiles	21	91,3100
	23	103,5300
	25	118,5000
	34	178,3800
	45	217,1500
	50	236,5000
	58	279,4600
	65	337,1000
	75	445,2500
	77	467,3300
	78	471,6200
	89	575,8100

Ejemplo 4. Se tiene los siguientes datos discretos use la regla de esturges

Para Elaborar el cuadro de frecuencias absoluta simple y acumulada y las medidas de tendencia Central y los percentiles 56 y 91.

45	48	49	71	19	67	79	16	21	36
44	61	11	25	67	17	73	40	51	75
42	76	43	76	64	39	30	19	14	59
56	33	65	39	50	37	14	38	21	44
33	14	30	48	64	80	36	65	24	18
60	54	36	80	33	59	76	31	47	55
70	51	19	79	17	68	46	70	79	35
44	32	28	72	78	28	58	71	30	32
13	70	44	81	65	78	53	12	32	24
67	34	68	83	80	29	19	77	20	11
63	37	63	47	48	70	30	39	35	46
83	23	25	30	83	32	73	51	71	61

Solución :

Tabla N°40: *Tabla de resultados usando Esturges*

Vmáx=	83		
Vmin=	11		
R=	72		
K=	7.861298	8	9
TIC=	9		
D=TIC*k-R=	0	9	

Tabla N°341: *Tabla de la frecuencia absoluta simple y las respuestas en Excel*

IC	fi	Fi
[0-7)		
[7-16)	7	7
[16-25)	14	21
[25-34)	18	39
[34-43)	14	53
[43-52)	18	71
[52-61)	8	79
[61-70)	14	93
[70-79)	17	110
[79-88)	10	120
[88-97)		
	120	

MA=	48.025
Mo1=	29.5
Mo2=	45.57143
P21=	27.1
P25=Q1=	29.5
P47=	44.7
P50=Q2=Me=	46.5
P56=	50.1
P75=Q3=	68.07143
P91=	78.57647

Ejercicios de Aplicación:

1. En cada caso que se indica a continuación, indique si se trata de un dato, un estadístico o un parámetro:

Ejercicios de Aplicación:

1. En cada caso que se indica a continuación, indique si se trata de un dato, un estadístico o un parámetro:

    a) Alfonso tiene 23 años

    b) Durante el año 2018, mensualmente, mi familia gastó en promedio, $4 100,00.

    c) El promedio de notas en Matemática de los alumnos de mi clase, es de 13,6.

    d) El salario diario de Enrique, es equivalente a $ 21,00.

    e) El CI promedio de los alumnos de colegios nacionales de Lima Metropolitana es, 109.

2. Indique el tipo de variable

    a) Tipos de personalidad.

    b) Sueldos mensuales netos de los trabajadores del Estado Peruano.

    c) Número de alumnos por curso en la Facultad de Psicología de la UNMSM, el semestre anterior.

    d) Grado de instrucción de un conjunto de personas.

    e) Pesos (en Kg.) de los trabajadores del Ministerio de Salud.

3. Dé el nivel de medición para las variables dadas en la pregunta 2.

4. Determine la verdad o falsedad de las siguientes proposiciones:

    a) Para que un conjunto de individuos sea una población, éste debe tener el número mínimo de elementos establecido para cada situación.

    b) Una muestra es representativa de una población sólo si tiene el 10% de los elementos de ella.

    c) En ningún caso la variable discreta puede tomar valores decimales.

    d) La Inferencia Estadística se apoya en la Estadística Descriptiva.

    e) Es imposible pasar una medición dada en el nivel "por razones y proporciones" al nivel "por intervalos".

5. Complete:

    a) El estadístico es a la ____________ como el parámetro es a la _________________

    b) En el nivel de medición ordinal no es posible ____________ dos mediciones.

    c) La variable "Número de las camisetas de los jugadores de la selección de fútbol" está dada en el nivel de medición ____________.

    d) Las respuestas posibles a la pregunta: ¿Qué día de la semana has nacido?, son los valores de una variable ______________ en el nivel de medición ______________.

    e) La variable inteligencia puede ser medida en los niveles de medición

    ____

6. Indique el tipo de variable y el nivel de medición que corresponda a las posibles respuestas de las siguientes preguntas:

    a) ¿Cuántas horas estudias diariamente?

    b) ¿En qué distrito has nacido?

    c) ¿Qué color de zapatos prefieres?

    d) ¿Cuántos hermanos tienes?

    e) ¿Qué nota obtuviste en el último examen de Matemática

7. Calcule los límites reales de los siguientes números:

    (i) 0,23
    (ii) 1,78
    (iii) 26
    (iv) 123
    (v) 0,345
    (vi) 0,046
    (vii) 23,5

8. ¿Una variable independiente de una investigación, puede ser dependiente en otra?

9. Los siguientes datos corresponden al número de créditos aprobados en su carrera obtenidos por un conjunto de estudiantes, en la Fac. de Ingeniería Industrial:

35	37	47	32	38	39	47	52	50	52
68	64	43	38	38	42	47	48	50	31
44	48	35	32	34	46	48	49	51	56
49	51	51	39	60	64	61	36	58	59
57	58	53	64	37	54	59	66	48	42

10. Una prestigiosa universidad de 10 000 alumnos se realizó un censo de la población estudiantil, observándose que: El 35% pertenece a la Facultad de Psicología, el 45% a la Facultad de Contabilidad y el 20% a la facultad de Economía.

El ingreso familiar promedio mensual de los alumnos es S/. 3200 y el ingreso familiar promedio mensual de 180 alumnos escogidos al azar es de S/. 3100.

De la información anterior obtener:

a) ***POBLACIÓN:*** *b)* ***PARÁMETRO:***

c) MUESTRA: *d)* ***ESTADÍGRAFO:***

11.- Los siguientes datos representan la cantidad de hemoglobina (Hb) en gr/dl. Encontrados en 50 pacientes entre 5 y 16 años, atendidos en el hospital A durante los meses de Junio, Julio y Agosto del 2005.

a) ***Completar la tabla correspondiente***

Tabla N°42: *Tabla de frecuencias de la cantidad de Hemoglobina de 50 pacientes*

Hemoglobina (gr/dl.)	f_i	h_i	h_i (%)	H_i	H_i (%)	F_i
10.0 - 10.8	2					
10.9 - 11.7			10			
11.8 - 12.6	11					
12.7 - 13.5	18					
13.6 - 14.4	6					
14.5 - 15.3	2					
15.4 - 16.2			12			
Total	50		100			

b) Dar el valor y significado de $f_3, h_2, h_5\%, F_4, H_4, H_5\%$.

c) Construya el histograma y polígono para las frecuencias absolutas simples (indique correctamente en el gráfico todas las variables y el título).

CAPITULO III

CAPITULO III

Medidas de Dispersión

Valoración de los datos estadísticos de carácter cuantitativo. Sus características son:

a) Encontrar el nivel de medida de la variable de interés.
b) Elaborarla distribución de las variables.
    - Medidas de tendencia central y de dispersión para cada variable.
    - Sesgo y Kurtosis para cada variable.
    - Valoración visual de la distribución de los datos.
    - Examinar los diagramas de las probabilidades de la distribución.
    - Si se considera necesario transformar las variables.
    - Ver los resultados de la transformación.
c) Hallar la homogeneidad de las varianzas.
d) Ver el tamaño de muestra total y de los subgrupos.
e) Determinar qué prueba estadística paramétrica o no paramétrica es la más adecuada.

Por esta razón es importante estudiar las medidas de dispersión.

3.1. Rango, Desviación Cuartil, Desviación media

Rango (o Intervalo):

Es la distancia que existe entre el menor y mayor valor de los datos.

Datos No Agrupados: $\text{Rango} = V\text{max} - V\text{min}$

Datos Agrupados: $\text{rango} = LS_k - LI_1$,

donde k = última clase

Rango Semi-Inter Cuartil (Q): (o Desviación Cuartil)

Mide el rango promedio de una cuarta parte de los datos (evita los valores extremos)

$$DQ = \frac{Q_3 - Q_1}{2}$$

determinar el número de kilómetros por día que cada uno maneja en un año. Los resultados del estudio son los siguientes. Agrupe los datos usando la regla de Sturges y calcule el rango y el rango intercuartil.

3.6	4.2	4.7	4.9	5.3	5.7	6.7	7.3
7.7	8.1	8.3	8.4	8.7	8.7	8.9	9.3
9.5	9.5	9.7	10	10.3	10.5	10.7	10.8
11	11.3	11.3	11.8	12.1	12.7	12.9	13.1
13.5	13.8	14.6	14.9	16.3	17.2	18.5	20.3

Solución

Rango = 20,3 - 3,6 = 16,7 kilómetros.

Tabla N°43: *Tabla de resultados del número de kilómetros de recorrido por día*

B	C	D	E	F	G	H	I
3.6	4.2	4.7	4.9	5.3	5.7	6.7	7.3
7.7	8.1	8.3	8.4	8.7	8.7	8.9	9.3
9.5	9.5	9.7	10	10.3	10.5	10.7	10.8
11	11.3	11.3	11.8	12.1	12.7	12.9	13.1
13.5	13.8	14.6	14.9	16.3	17.2	18.5	20.3

min	3.6	
max	20.3	
r	16.7	
k	6.28679797	7
r/k	2.38571429	3
d	4.3	

		fi	Fi
3.6	6.6	5	5
6.6	9.6	7	12
9.6	12.6	15	27
12.6	15.6	8	35
15.6	18.6	3	38
18.6	21.6	2	40
21.6	24.6	0	40
		40	

10	Q1=	8.743
30	Q3=	13.725
	rango interc.	2.491

Ejemplo 2. Para los siguientes datos, calcule el rango intercuartil para datos agrupados.

99	75	84	61	33	45	66	97	69	55
72	91	74	93	54	76	52	91	77	68

Solución

Tabla N°44: *Tabla de resultados para datos agrupados usando Excel*

B	C	D	E	F	G	H	I	J	K
99	75	84	61	33	45	66	97	69	55
72	91	74	93	54	76	52	91	77	68

min	33			
max	99		33	45
R=	66		45	57
K=1+3.3LOG(N)	5.293399	6	56	68
TIC=R/K	11	12	67	79
D=K*TIC-R	6		78	90
			89	101

clase	fi	Fi
(33-45)	1	1
(45-57)	4	5
(56-68)	2	7
(67-79)	7	14
(78-90)	1	15
(89-101)	5	20
	20	

15	Q3	90
5	Q1	57

RANGO INTERCUARTIL	[illegible]

Desviación Media Absoluta (DM): (o Desviación Absoluta Promedio)

Es la distancia promedio de los datos a su media, cuyas formulas son:

Datos No Agrupados: $\mathrm{DM} = \dfrac{\sum_{i=1}^{n} |X_i - \overline{X}|}{n}$, Datos Agrupados: $\mathrm{DM} = \dfrac{\sum_{i=1}^{k} f_i |X_i - \overline{X}|}{n}$

3.2. Varianza , Desviación Estándar y Coeficiente de Variación

Varianza:

Poblacional (σ^2): Es el promedio del cuadrado de la distancia de los datos a su media

Datos No Agrupados: $\sigma^2 = \dfrac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$, Datos Agrupados: $\sigma^2 = \dfrac{\sum_{i=1}^{k} f_i (X_i - \mu)^2}{N}$

Muestral (S^2): La suma de las distancias al cuadrado se divide entre en número de datos menos uno, cuyas formulas son:

Datos No Agrupados: $S^2 = \dfrac{\sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2}{n-1}$, Datos Agrupados: $S^2 = \dfrac{\sum_{i=1}^{k} f_i \left(x_i - \overline{x}\right)^2}{n-1}$

Nota: S^2 para muestras "chicas o sea para muestras menores a 30 datos ". Para muestras grandes S^2 o σ^2 prácticamente no difieren.

Desviación Estándar:

3.3. Desviación Estándar:

Mide la variación de los datos en términos absolutos. Es la raíz cuadrada positiva de la varianza. Su notación para datos: poblacional: $\sigma = \sqrt{\sigma^2}$, muestral: $S = \sqrt{S^2}$

Ejemplo 1: La edad de los estudiantes regulares que acuden a un curso en los tumos matutino y vespertino del nivel de pre-grado de la FIIS-UNAC se describe en las siguientes dos muestras:

Turno matutino	23	29	27	22	24	21	25	26	27	24
Turno vespertino	27	34	30	29	28	30	34	35	28	29

Si la homogeneidad de la clase es un factor positivo en el aprendizaje, utilice una medida de variabilidad relativa para sugerir en cuál de los dos grupos será más fácil enseñar.

Respuesta: Para nuestro caso usamos el SPSS

Tabla N°45: *Tabla de resultados usando SPSS*

Estadísticos

		Matutino	Vespertino
N	Válido	10	10
Media		24,8000	30,4000
Mediana		24,5000	29,5000
Desv. Desviación		2.48551	2,87518
Varianza		6,178	8,267
Rango		8,00	8,00
Mínimo		21,00	27,00
Máximo		29,00	35,00
Percentiles	25	22,7500	28,0000
	50	24,5000	29,5000
	75	27,0000	34,0000

Respuesta: EL turno matutino es más fácil enseñar

Ejemplo 2. La compañía de Periódico de Trujillo "La Industria" tiene tres oficinas en tres ciudades distintas. Los niveles de salario difieren de un lugar a otro. En la oficina de Trujillo el aumento promedio a los salarios durante el año anterior fue S/ 1,500, con una desviación estándar de S/ 400. En la sucursal de Chiclayo el aumento promedio fue S/ 3,760, con una desviación estándar de S/ 622. En Lambayeque el aumento promedio fue S/ 850, con una desviación estándar de $95. Se entrevistó a tres empleados. El empleado de Trujillo recibió un aumento de S/ 1,100; el de Chiclayo, obtuvo un aumento de S/ 3,200; y el de Lambayeque uno de S/ 500. ¿Cuál de los tres tuvo el menor aumento en relación con la media y la desviación estándar de los aumentos correspondientes a su oficina?

Tabla N°46: *Niveles de salario de la Compañía "La Industria"*

sucursal	promedio	estándar	%	aumento de empleados
Trujillo	1,500	400	26.667	1,100
Chiclayo	3,760	622	16.543	3,200
Lambayeque	850	95	11.176	500

Respuesta: El empleado de Lambayeque.

La desviación estándar se interpreta construyendo intervalos alrededor del promedio:

Teorema de Chebyshev.

Para ejecutar un chebychev es necesario lo siguiente: los datos deben estar concentrados en un 75% y un 94% con relación a la desviación estándar de la media.

Si por un caso la forma de la curva tuviera forma de campana los datos deben de concentrarse en un 68%, un 95% y un 99.7% dentro de la desviación estándar.

Para encontrar los datos del chebychev se debe concentrar dentro de k desviaciones estándar y la media.

https://apaehogm-glendi.blogspot.com/2008/

Regla Empírica. Si la distribución es una curva acampanada, unimodal y simétrica:

- Aproximadamente el 68% de los datos (población) se encuentran a una desviación estándar alrededor de la media: $(\overline{X} \pm S)$

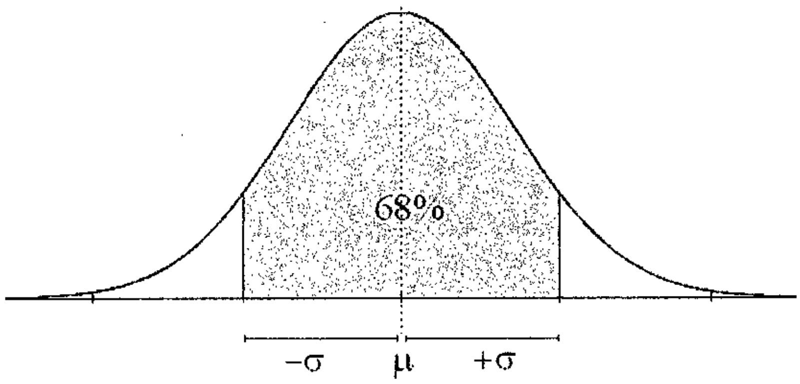


Figura N°12: Curva Normal con datos alrededor de la media más/menos una desviación estándar

- Aproximadamente el 95% de los datos (población) se encuentran a 2 desviaciones estándar alrededor de la media: $(\overline{X} \pm 2S)$

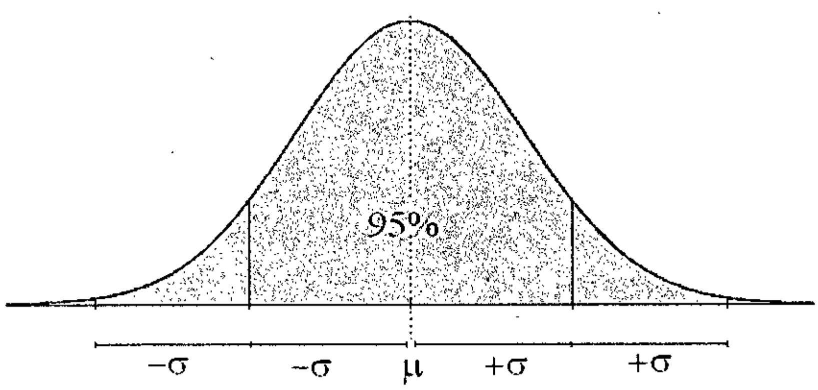


Figura N°13: Curva Normal con datos alrededor de la media más/menos dos desviaciónes estándar

- Aproximadamente el 99% de los datos (población) se encuentran a 3 desviaciones estándar alrededor de la media: $(\overline{X} \pm 3S)$

Observación: Si la distribución no es simétrica y unimodal.

- Al menos el 75% de los valores cae dentro de 2 desviaciones estándar alrededor de la media: $(\overline{X} \pm 2S)$
- Al menos el 89% de los valores caen dentro de 3 desviaciones estándar alrededor de la media: $(\overline{X} \pm 3S)$

Ejemplo 3. El Banco Continental proporciona información a sus suscriptores para permitirles evaluar el desempeño de los fondos de inversión que consideran vehículos de inversión potencial. Un estudio reciente de los fondos cuya meta de inversión establecida era crecimiento e ingreso produjo los siguientes datos de la tasa de retorno anual sobre la inversión total durante los últimos cinco años:

Red. Anual%	[11-11.9)	[12-12.9)	[13-13.9)	[14-14.9)	[15-15.9)	[16-16.9)	[17-17.9)	[18-18.9)
Frecuencia	2	2	8	10	11	8	3	1

a) Calcule la media, la varianza y la desviación estándar de la tasa de rendimiento anual para esta muestra de 45 fondos de inversión.

$$\bar{x} = \frac{\sum mi * fi}{n} = \frac{672.25}{45} = 14.9389$$

$$s^2 = \frac{\sum \left(m_i - \bar{x}\right)^2 * f_i}{n-1}$$ cuando los datos son menores de 30 datos pero como tenemos más de 30 el denominado es “n”

Tabla N°47: *Tabla de resultados de desempeño de los fondos usando Excel*

rendimiento anual	fi	mi	fi x mi	mi - media	|mi-media|	|mi-media|*f	|mi-media|²*fi
[11-11.9)	2	11.45	22.9	-3.488888889	3.48888889	6.977777778	24.34469136
[12-12.9)	2	12.45	24.9	-2.488888889	2.48888889	4.977777778	12.3891358
[13-13.9)	8	13.45	107.6	-1.488888889	1.48888889	11.91111111	17.73432099
[14-14.9)	10	14.45	144.5	-0.488888889	0.48888889	4.888888889	2.390123457
[15-15.9)	11	15.45	169.95	0.511111111	0.51111111	5.622222222	2.873580247
[16-16.9)	8	16.45	131.6	1.511111111	1.51111111	12.08888889	18.26765432
[17-17.9)	3	17.45	52.35	2.511111111	2.51111111	7.533333333	18.91703704
[18-18.9)	1	18.45	18.45	3.511111111	3.51111111	3.511111111	12.32790123
	45		672.25				109.2444444

→La varianza es: $s^2 = \frac{\sum \left(m_i - \bar{x}\right)^2 * f_i}{n} = 2.4276543$

Y la desviación Estándar es: S = 1.5580932

b) Según el teorema de Chebyshev, ¿entre qué valores debe caer al menos 75% de las observaciones de la muestra? ¿Qué porcentaje de observaciones caen de hecho en ese intervalo?

El teorema de Chebyshev establece que independientemente de la forma de la distribución, al menos 75% de los valores caen dentro de ± 2 desviaciones estándar a partir de la media de la distribución. Usamos Excel para la solución

Tabla N°48: *Resultados del 75% de las observaciones usando Chebyshev*

media=	14.93888889	$\bar{x} = \frac{\sum mi * fi}{n}$
varianza=	2.427654321	$s^2 = \frac{\sum(m_i - \bar{x})^2 * f_i}{n}$
desviación estándar=	1.558093168	= S

Chevyshev:

Intervalo 75%:	11.8227026	18.05507523	[11.822703-18.0550752)
Realmente= 87.53%			
Intervalo 68%:	13.3807957	16.49698206	[13.36-16.5146)
Realmente= 81.28%			

Procedimiento: →El teorema de Chebyshev nos dice que al menos el 75% de los valores (6 de 8) están entre $(\overline{X} \pm 2S)$=14.9388 — 2(1.558093168) = =**11.822703** y 14.9388 + 2(1.5756) = **18.055075**

Luego restamos 18.055075 - 11.822703 = 6.2323727

Luego 2*6.2323727= 12.464745, entonces tendremos 100-12.464745= 87.535255, ósea: →Realmente el 87.5% se encuentra dentro de ese intervalo.

c) Dado que la distribución es casi una campana, ¿entre qué valores se esperaría encontrar 68% de las observaciones? ¿Qué porcentaje de las observaciones de hecho caen en ese intervalo?

$(\overline{X}\text{-S})$=14.93888888-1.558= 13.38 y $(\overline{X}\text{+S})$=14.93888888+1.558 = 16.4969821

Ahora restamos 16.4969821 -13.38 = 3.116186

Luego multiplicamos 2*3.116186 = 6.23237

Ahora restamos a 75 - 6.23237 = 68.7676

Aproximadamente 68.7676% de los valores de la población cae dentro de ±1 desviación estándar a partir de la media.

3.4. Coeficiente de Variación (CV): Mide la variación relativa de la variable con respecto a su promedio. Mide la magnitud de la desviación estándar en relación con la media. Se expresa en porcentajes. Cuya fórmula es: $CV = \frac{S \times 100\%}{\bar{x}}$ para una muestra.

$CV = \frac{\sigma \times 100\%}{\bar{\mu}}$ para la población.

Observación:

- El CV es un estadístico útil para comparar la dispersión de conjuntos de datos que tienen distintas desviaciones estándar y distintos promedios.
- El CV pierde su utilidad cuando la media aritmética se acerca a cero.
- Supongamos que S = 4.8 y su media aritmética $\bar{x} = 94.8$

$$CV = \frac{4.8*100}{94.5} = 4.87309645\%$$

Interpretación: La desviación típica de la muestra es de 4.87% del valor de la media de la muestra.

Ejemplo1. ¿De qué manera cree que debe aplicarse el concepto de variabilidad a una investigación que realiza la Secretaría de Fiscalización del Callao (SFC) con el propósito de determinar la posibilidad de que un grupo de fabricantes fije los precios de los productos?

Solución: La SFC examinaría la variabilidad del precio para la industria y compararía el resultado con el de las compañías sospechosas la secretaria podría examinar las distribuciones de precios para los productos parecidos, para los mismos productos en una ciudad o para los mismos productos en ciudades diferentes. Si la variabilidad fuera significativamente diferente en cualquiera de estos casos, este resultado podría constituir una evidencia de una conspiración para llevar los precios a los mismos niveles.

Ejemplo 2: Los siguientes datos son una muestra de la tasa de producción diaria de miles de galones de pintura para autos de la fábrica CPP:

17 21 18 27 17 21 20 22 18 23

El gerente de producción de la compañía siente que una desviación estándar de más de tres botes por día indica variaciones de tasas de producción inaceptables. ¿Deberá preocuparse por las tasas de producción de la planta? Usaremos Excel

Tabla N°49: *Resultados de la producción diaria de pintura usando Excel*

	x_i	x_i^2
1	17	289
2	21	441
3	18	324
4	27	729
5	17	289
6	21	441
7	20	400
8	22	484
9	18	324
10	23	529
SUMA	204	4250
	MA=	20.4
	$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$	9.822222222
	S=Raíz de s^2	3.134042473

Ahora comprobamos la solución usando SPSS

Tabla N°50: *Resultados de la producción diaria de pintura usando SPSS*

Estadísticos		
Producción diaria		
N	Válido	10
Media		20,4000
Desv. Desviación		3.13404
Varianza		9,822
Rango		10,00

La desviación estándar de **3.13** botes representa un nivel inaceptable de variabilidad.

Ejemplo 3: La Universidad Nacional de la Molina vende tres categorías de maíz Penta 1070, que se diferencian entre sí por el nivel de consistencia de sus germinaciones. El laboratorio con que cuenta la Universidad es de última generación para la prueba de las semillas que están produciendo en un campo Agrícola y tiene una muestra de cada categoría y los resultados de las pruebas acerca de las semillas que germinan, de un paquete de 100, son los siguientes:

cat. 1 - regular	88	91	92	89	79
cat. 2 - extra	87	92	88	90	92
cat. 3 - super	90	89	79	93	88

El Ingeniero de planta cree que está mal la clasificación ¿Que le aconsejaría usted?

Solución: Primero lo hacemos mediante el excel

Tabla N°51 *Producción de semillas de maíz en tres categorías con Excel*

cat. 1 - regular	xi-MA	(xi-MA)	$(xi-MA)^2$
88	0.2	0.2	0.04
91	3.2	3.2	10.24
92	4.2	4.2	17.64
89	1.2	1.2	1.44
79	-8.8	8.8	77.44
439		17.6	106.8
Ma=	87.8		
S^2=	26.7		
S=	5.16720427		
CV=	5.88519849	5.88%	

cat. 2 - extra	xi-MA	(xi-MA)	$(xi-MA)^2$
87	-2.8	2.8	7.84
92	2.2	2.2	4.84
88	-1.8	1.8	3.24
90	0.2	0.2	0.04
92	2.2	2.2	4.84
449		9.2	20.8
Ma=	89.8		
S^2=	5.2		
S=	2.28035085		
CV=	2.5393662	2.54%	

cat. 3 - super	xi-MA	(xi-MA)	$(xi-MA)^2$
90	2.2	2.2	4.84
89	1.2	1.2	1.44
79	-8.8	8.8	77.44
93	5.2	5.2	27.04
88	0.2	0.2	0.04
439		17.6	110.8
Ma=	87.8		
S^2=	27.7		
S=	5.26307895		
CV=	5.99439516	6%	

Ahora usaremos SPSS

Tabla N°52 *Producción de semillas de maíz en tres categorías con SPSS*

Estadísticos				
		Categoría1	Categoría2	Categoría3
N	Válido	5	5	5
	Perdidos	0	0	0
Media		87,8000	89,8000	87,8000
Mediana		89,0000	90,0000	89,0000
Desv. Desviación		5,16720	2,28035	5,26308
Varianza		26,700	5,200	27,700
Rango		13,00	5,00	14,00
Mínimo		79,00	87,00	79,00
Máximo		92,00	92,00	93,00
Percentiles	25	83,5000	87,5000	83,5000
	50	89,0000	90,0000	89,0000
	75	91,5000	92,0000	91,5000

Podemos decir que no tiene sentido la clasificación ya que según la dispersión relativa la mejor categoría seria la categoría 2 pero sin embargo están considerando la categoría 3 como la mejor; además Podemos observar que la categoría 1 y tres tienen casi los mismos valores de las medidas de dispersión solo difieren en el rango y el que tiene un rango menor es la categoría 1.

Ejemplo 4. Para los siguientes datos discretos elaborados de la siguiente manera:

Primero haga el sorteo aleatorio (en Excel) hasta la tercera vez entre 16 y 71 (edades de las personas que van al cine a ver la película de estreno) entre las columnas A y J y las filas del 1 y 7 de tal manera que tenga 120 datos aleatorios.

C) Determine el número de intervalos y el tamaño del intervalo de clase, para variable cuantitativa continua, usando la regla de Sturges

D) Elabore el cuadro con todas las frecuencias estudiadas usando Excel, para las especificaciones de (A)

E) Elabore el Histograma de Frecuencias absoluta simple y el polígono de frecuencias. Use todas las especificaciones hechas en clase.

F) Encuentre La media aritmética. La Moda. La Mediana

G) Encuentre: P_{21} , P_{47} , P_{56} , P_{91} ,Q_1, Q_2, Q_3,

Imagen del desarrollo en Excel.

Solución:

Sorteo de los datos en Excel y cálculo de filas y columnas usando la regla de Sturges:

Cuadro de distribución de frecuencias

45	48	49	71	19	67	79	16	21	36
44	61	11	25	67	17	73	40	51	75
42	76	43	76	64	39	30	19	14	59
56	33	65	39	50	37	14	38	21	44
33	14	30	48	64	80	36	65	24	18
60	54	36	80	33	59	76	31	47	55
70	51	19	79	17	68	46	70	79	35
44	32	28	72	78	28	58	71	30	32
13	70	44	81	65	78	53	12	32	24
67	34	68	83	80	29	19	77	20	11
63	37	63	47	48	70	30	39	35	46
83	23	25	30	83	32	73	51	71	61

Vmáx=	83		
Vmin=	11		
R=	72		
K=	7.86129811	8	9
TIC=	9		
D=TIC*k-R=	0	9	

Tabla N°53: *Tabla de frecuencias de los 120 datos y sus resultados en Excel*

IC	fi	Fi	hi	Hi	F_i^*	H_i^*
[7-16)	7	7	0.05833333	0.05833333	120	1
[16-25)	14	21	0.11666667	0.175	113	0.94166667
[25-34)	18	39	0.15	0.325	99	0.825
[34-43)	14	53	0.11666667	0.44166667	81	0.675
[43-52)	18	71	0.15	0.59166667	67	0.55833333
[52-61)	8	79	0.06666667	0.65833333	49	0.40833333
[61-70)	14	93	0.11666667	0.775	41	0.34166667
[70-79)	17	110	0.14166667	0.91666667	27	0.225
[79-88)	10	120	0.08333333	1	10	0.08333333
	120					

MA=	48.025		
Mo1=	29.5		
Mo2=	45.5714286	k	k*n/100
P21=	27.1	21	25.2
P25=Q1=	29.5	25	30
P47=	44.7	47	56.4
P50=Q2=Me=	46.5	50	60
P56=	50.1	56	67.2
P75=Q3=	68.0714286	75	90
P91=	78.5764706	91	109.2

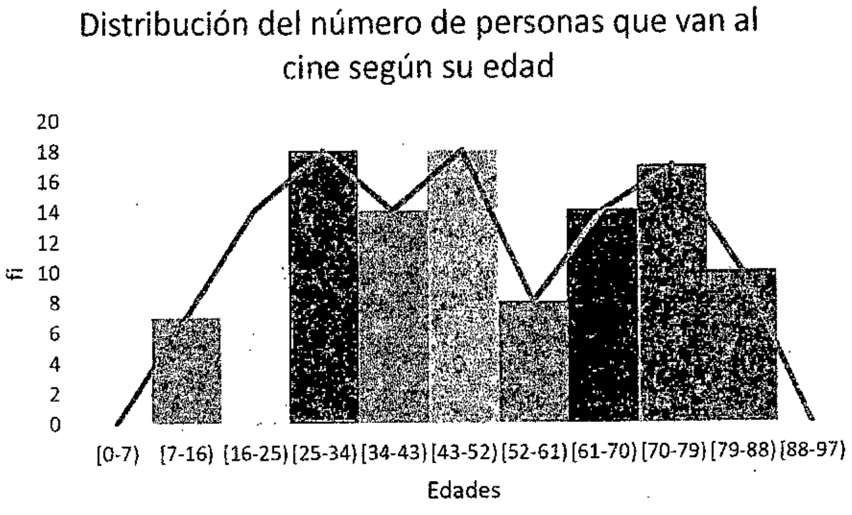


Figura N°14: Histograma y polígono de frecuencias del número de personas que van al cine a ver una película de extreno según su edad.

Ejemplo 6. Dado el siguiente conjunto de puntajes de una prueba de selección de personal :

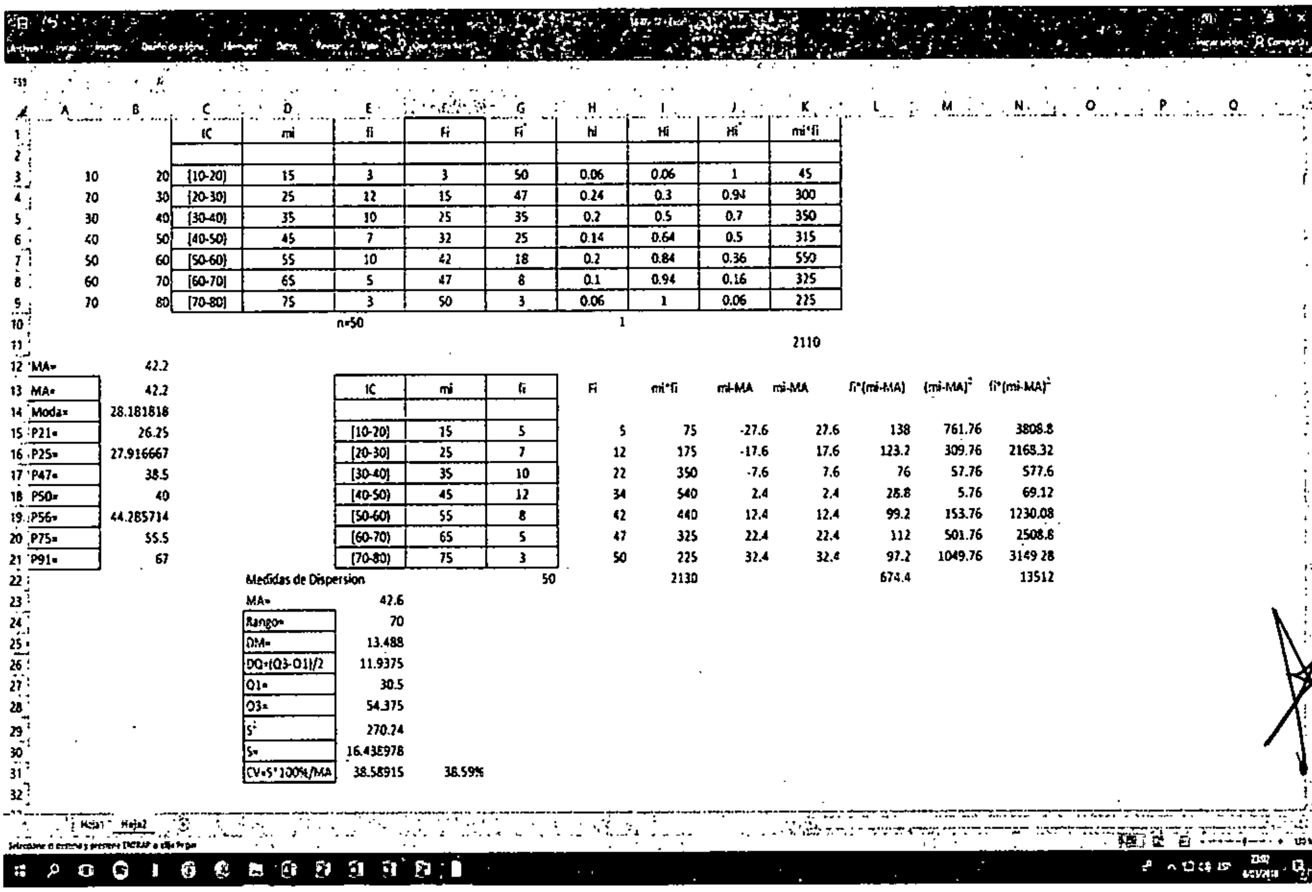
Tabla N°54: *Número de personas en la prueba de selección según su puntaje*

Puntaje	f_i
10 - 20	*3*
20 - 30	*12*
30 – 40	*10*
40 – 50	*7*
50 – 60	*10*
60 – 70	*5*
70 – 80	*3*
Total	*50*

- *Calcular:*
- *MTC: Medidas de tendencia central (Media Aritmética, Mediana, Moda, todos los cuartiles. Rango)*
- *Percentiles: 21, 47, 56, 91*
- *Desviación cuartil. Desviación media.*
- *Varianza. Desviación estándar y Coeficiente de variación.*

Solución Imagen del desarrollo en Excel:

Tabla N°55: *Tabla de frecuencias y MTC y Medidas de dispersión*



		IC	mi	fi	Fi	Fi*	hi	Hi	Hi*	mi*fi
10	20	[10-20)	15	3	3	50	0.06	0.06	1	45
20	30	[20-30)	25	12	15	47	0.24	0.3	0.94	300
30	40	[30-40)	35	10	25	35	0.2	0.5	0.7	350
40	50	[40-50)	45	7	32	25	0.14	0.64	0.5	315
50	60	[50-60)	55	10	42	18	0.2	0.84	0.36	550
60	70	[60-70)	65	5	47	8	0.1	0.94	0.16	325
70	80	[70-80)	75	3	50	3	0.06	1	0.06	225
				n=50			1			
										2110

MA=	42.2
MA=	42.2
Moda=	28.181818
P21=	26.25
P25=	27.916667
P47=	38.5
P50=	40
P56=	44.285714
P75=	55.5
P91=	67

IC	mi	fi	Fi	mi*fi	mi-MA	mi-MA	fi*(mi-MA)	(mi-MA)²	fi*(mi-MA)²
[10-20)	15	5	5	75	-27.6	27.6	138	761.76	3808.8
[20-30)	25	7	12	175	-17.6	17.6	123.2	309.76	2168.32
[30-40)	35	10	22	350	-7.6	7.6	76	57.76	577.6
[40-50)	45	12	34	540	2.4	2.4	28.8	5.76	69.12
[50-60)	55	8	42	440	12.4	12.4	99.2	153.76	1230.08
[60-70)	65	5	47	325	22.4	22.4	112	501.76	2508.8
[70-80)	75	3	50	225	32.4	32.4	97.2	1049.76	3149 28
		50		2130			674.4		13512

Medidas de Dispersion

MA=	42.6	
Rango=	70	
DM=	13.488	
DQ=(Q3-Q1)/2	11.9375	
Q1=	30.5	
Q3=	54.375	
S²	270.24	
S=	16.438978	
CV=S*100%/MA	38.58915	38.59%

Ejemplo 7. Dados el número de horas extras de un grupo de 16 trabajadores de la Empresa "COINSA" de materiales de construcción Civil: 5, 6, 7, 8, 11, 12, 13, 15, 16, 17, 18, 18, 19, 7, 13, 16. Hallar: Todas las medidas de dispersión

Tabla N°56: *Cálculo de todas las medidas de dispersión en Excel.*

N° de orden	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Suma
xi	5	6	7	7	8	11	12	13	13	15	16	16	17	18	18	19	201
xi-MA	7.5625	6.563	5.563	5.5625	4.563	1.563	0.56	0.4	0.44	2.44	3.44	3.44	4.44	5.44	5.44	6.44	63.88
(xi-MA)²	57.191	43.07	30.94	30.941	20.82	2.441	0.32	0.2	0.19	5.94	11.8	11.8	19.7	29.6	29.6	41.4	335.9

Usando Excel se tiene los siguientes resultados:

Donde MA es la Media Aritmética y x i son los datos recogidos de la muestra

Tabla N°57: *Tabla de respuestas de todas las medidas de dispersión*

MA=	12.5625	
DM=	3.9921875	
DQ=	4.75	
Q1=	7.25	
Q3=	16.75	
S^2	22.3958333	22.3958333
S=	4.73242362	4.73242362
CV=S*100/MA	37.6710338	37.67%

Ejemplo 8: Dado los datos x_i de las ventas de en miles de soles de un negocio se pide todas las medidas de dispersión.

Tabla N°58: *Cálculo de todas las medidas de dispersión de las ventas en Excel.*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
xi	13	14	18	19	25	26	27	31	34	35	35	36	37	38	39	39
xi-MA	16.13	15.13	11.13	10.13	4.13	3.13	2.13	1.88	4.88	5.88	5.88	6.88	7.88	8.88	9.88	9.88
$(xi-MA)^2$	260.02	228.77	123.77	102.52	17.02	9.77	4.52	3.52	23.77	34.52	34.52	47.27	62.02	78.77	97.52	97.52

Solución

Usando Excel se tiene los siguientes resultados:

Donde MA es la Media Aritmética y x_i son los datos recogidos de la muestra

Tabla N°59: *Tabla de respuestas de todas las medidas de dispersión*

Rango=	26		Q_3=	36.75
Media Aritmética	29.125			
Desviación Media	7.73438		Q_1=	20.5
Desviación cuartil $DQ=(Q_3-Q_1)/2$	8.125			
Varianza: S^2=	81.7167			
Desviación Estándar: S=	9.03973			
Coeficiente de Variación: CV=	31.0377	31.04%		

EJERCICIOS DE APLICACIÓN

1.- Los siguientes datos corresponden a los Jornales Semanales de una Empresa de confecciones Textil, los datos incluyen las horas extras y están redondeados, así mismo están expresados en dólares.

116	**102**	**141**	**86**	**123**	**73**	**99**	**133**	**110**
144	**137**	**151**	**79**	**72**	**112**	**154**	**75**	**149**
66	**132**	**138**	**123**	**79**	**131**	**130**	**93**	**105**
75	**105**	**131**	**145**	**141**	**119**	**68**	**112**	**105**
155	**116**	**81**	**96**	**73**	**132**	**146**	**103**	**74**
139	**140**	**77**	**94**	**142**	**95**	**73**	**93**	**67**
71	**65**	**91**	**141**	**83**	**69**	**122**	**102**	**64**
138	**144**	**95**	**145**	**120**	**74**	**81**	**110**	**67**
77	**148**	**105**	**67**	**79**	**102**	**104**	**65**	**121**

1.1.1 Hallar la Tabla de Distribución de Frecuencias,

1.1.2 Que significa $\mathbf{f_3, F_5, F_3^*, h_5, h_4, H_4, H_3}$ **y** $\mathbf{H_4^*}$

1.1.3 Graficar el Histograma de Frecuencias Relativas.

1.1.4 Graficar el Polígono de Frecuencias absolutas

1.1.5 Graficar el histograma y polígono de Frecuencias Absolutas.

2.- Los datos que se muestran corresponden a la evaluación de una prueba psicotécnica que se ha aplicado a un grupo de trabajadores, la Tabla de Distribución de Frecuencias – TDF nos muestra los siguientes datos:

Tabla N°60: *Tabla de distribución de frecuencias*

CLASE	INTERVALOS	mi	fi	Fi	Fi*	hi	Hi	Hi*
1	[52 -)	54.5	5			0.1		
2	[-)			9				
3	[-)						0.26	
4	[-)					0.12		
5	[-)		15					
6	[-)	79.5						
7	[-)							0.2
8	[-)					0.14		
	TOTAL							

2.1 Completar totalmente la Tabla de Distribución de Frecuencias.

2.2 Que significa $\mathbf{f_1, f_3, f_5, f_7}$

2.3 Que significa $\mathbf{F_4, F_5^*, F_7, F_3^*, F_4^*}$ **y** $\mathbf{F_6}$

2.4 Que significa $\mathbf{h_4, h_5}$ **y** $\mathbf{h_7}$

2.5 Que significa H_3, H_5, H_4^*, H_5^* y H_6

2.6 Graficar el Histograma de Frecuencias Absolutas.

2.7 Graficar el Histograma de Frecuencias Acumuladas Descendente.

2.8 Graficar el Polígono de Frecuencias Relativas.

3.- La Gerencia de Planta de Aceros Arequipa, tomó una prueba de destreza a sus operarios de planta, registrando para cada uno de ellos el tiempo, en minutos, que emplearon para culminar cierta tarea. Los datos obtenidos se resumieron en la siguiente tabla:

Tabla N°61: *Prueba de destreza en Aceros Arequipa y el tiempo de culminación*

CLASE	INTERVALOS	mi	fi	FI	FI*	hi	HI	HI*
1	[20 - 25)						0.04	
2	[-)		6			0.12		
3	[-)							
4	[-)	37.5						0.64
5	[-)		8					
6	[-)		7					
	TOTAL							

3.1 Completar totalmente la Tabla de Distribución de Frecuencias.

3.2 Que significa f_2, f_6, F_3, F_5, F_3^*, F_4^*, F_6^*, h_4, h_6, H_4, H_5, H_2^*, H_3^* y H_4^*

3.3 Graficar el Histograma de Frecuencias Absolutas Acumuladas

3.4 Graficar el Histograma de Frecuencias Acumuladas Descendente

3.5 Graficar el Polígono de Frecuencias Absolutas

4.- El Gerente de Recursos Humanos de una Empresa Textil ha solicitado las horas extras que han trabajado los obreros en el I Semestre del año en curso, el empleado que estaba elaborando el cuadro lo ha dejado inconcluso, solo registra algunos datos como se indican a continuación: Limite inferior del primer intervalo **375**, el límite inferior del sexto intervalo y último es **875**, $f_1=12$, $F_2=30$, $f_3=28$, $F_4=80$, $H_4=0.80$ y $h_5=0.15$:

4.1 Completar totalmente la Tabla de Distribución de Frecuencias.

4.2 Que significa f_3, f_4, F_3, F_4 , F_3^*, F_4^*, h_2, H_4, H_2^*, H_3^* y H_4^*

4.3 Graficar el Histograma de Frecuencias Absolutas

4.4 Graficar el Histograma de Frecuencias Relativas Acumuladas Descendente.

4.5 Graficar el Polígono de Frecuencias Absolutas.

4.6 Graficar la Ojiva de las Frecuencias Absolutas Ascendente.

4.7 Graficar la Ojiva de las Frecuencias Absolutas Descendente.

5. Dado el siguiente cuadro de puntuaciones:

Tabla N°62: *Tabla de distribución de frecuencias absoluta relativa*

Puntuaciones	h_i
15-19	0,02
20-24	0,09
25-29	0,12
30-34	0,21
35-39	0,11
40-44	0,19
45-49	0,16
50-54	0,10

- Hallar la media aritmética, mediana y moda

6.- La siguiente información es la que ha dejado un trabajador que estaba elaborando la Tabla de Distribución de Frecuencias, las mismas que corresponden a una evaluación psicotécnica de un grupo de trabajadores de una empresa de seguridad, los datos son los siguientes:

El número de intervalos de clase son siete con igual amplitud; el Límite inferior del tercer intervalo de frecuencia es **452, m_5=672; f_1=20; f_3=25; f_5=4 f_7=1; h_2=0.40; h_4=0.08 y (h_5+h_6+h_7=0.07)**; con estos datos calcular lo siguiente:

6.1 Completar totalmente la Tabla de Distribución de Frecuencias.

6.2 Que significa **f_3, f_4, F_3, F_4 , F_3^*, F_4^*, h_2, H_4, H_2^*, H_3^* y H_4^***

6.3 Graficar el Histograma de Frecuencias Absolutas

6.4 Graficar el Histograma de Frecuencias Relativas Acumuladas Descendente.

6.5 Graficar el Polígono de Frecuencias Absolutas.

6.6 Graficar la Ojiva de las Frecuencias Absolutas Ascendente.

6.7 Graficar la Ojiva de las Frecuencias Absolutas Descendente.

7.- En una empresa de Metal Mecánica las Remuneraciones de 80 trabajadores de la Planta de producción se encuentran ordenados de acuerdo a la Tabla de Distribución de Frecuencias que se muestra, además se tiene conocimiento que los trabajadores que tiene un ingrese entre S/. 1500 y menos de S/. 1600 nuevos soles su frecuencia relativa es mayor en un 5% con respecto a los que perciben entre S/.1700 y menos de S/. 1800 nuevos soles.

Tabla N°63: *Tabla de Remuneraciones de una empresa Metal Mecánica*

Clase	Intervalo	fi	Fi
1	[1100-1200)	4	
2	[1200-1300)		
3	[1300-1400)	24	44
4	[1400-1500)		60
5	[1500-1600)		
6	[1600-1700)	8	
7	[1700-1800)		

7.1 Completar la Tabla de Distribución de Frecuencias, así como Interpretar que significa **f_3, F_4, H_3, F_4* y H_5***

7.2 Graficar el Histograma de Frecuencias Relativas.

7.3 Graficar el Histograma de Frecuencias Relativas Acumuladas Descendente.

7.4 Graficar el Polígono de Frecuencias Absolutas.

7.5 Graficar la Ojiva de las Frecuencias Absolutas Ascendente.

7.6 Graficar la Ojiva de las Frecuencias Absolutas Descendente.

8.- La siguiente Distribución de frecuencias muestra las notas de 100 estudiantes del curso de matemáticas, los mismos que se muestran a continuación:

Tabla N°64: *Tabla de frecuencia de notas de 100 estudiantes*

Clase	Intervalo	fi	Fi
1	[5-7)	4	
2	[7-9)		
3	[9-11)	24	44
4	[11-13)		60
5	[13-15)		

8.1 Completar la Tabla de Distribución de Frecuencias, así como Interpretar que significa **f_3, F_4, H_3, F_4* y H_5***

8.2 Graficar el Histograma de Frecuencias Relativas.

8.3 Graficar el Histograma de Frecuencias Relativas Acumuladas Descendente.

8.4 Graficar el Polígono de Frecuencias Absolutas.

8.5 Graficar la Ojiva de las Frecuencias Absolutas Ascendente.

8.6 Graficar la Ojiva de las Frecuencias Absolutas Descendente.

9.- La siguiente Distribución de frecuencias muestra las notas de alumnos del curso de matemáticas, los mismos que se muestran a continuación:

Tabla N°65: *Tabla de frecuencia de notas de estudiantes de matemáticas*

INTERVALO	hi
[05 - 07)	k
[07 - 09)	0.17
[09 - 11)	2k
[11 - 13)	4k
[13 - 15)	0.13

9.1 Con los datos completar la Tabla de Distribución de Frecuencias.

9.2 Graficar el Histograma de Frecuencias Relativas.

9.3 Graficar el Histograma de Frecuencias Relativas Acumuladas Descendente.

9.4 Graficar la Ojiva de las Frecuencias Absolutas Ascendente.

9.5 Graficar la Ojiva de las Frecuencias Absolutas Descendente.

10. La siguiente información es la que ha dejado un trabajador que estaba elaborando la Tabla de Distribución de Frecuencias, las mismas que corresponden a una evaluación psicotécnica de un grupo de trabajadores de una empresa de seguridad, los datos son los siguientes:

    El número de intervalos de clase son siete con igual amplitud; el Límite superior del tercer intervalo de frecuencia es **540, m_6=630; f_1=30; f_3=60; f_5=30 f_7=10; h_2=0.20; h_4=0.10 y (h_5+h_6+h_7=0.25)**; con estos datos calcular lo siguiente:

    10.1Construir la Tabla de Distribución de Frecuencias.

    10.2 Graficar el Histograma de Frecuencias Relativas.

    10.3 Graficar el Histograma de Frecuencias Relativas Acumuladas.

    10.4 Graficar la Ojiva de las Frecuencias Absolutas Ascendente.

    10.5 Graficar la Ojiva de las Frecuencias Absolutas Descendente.

11. Se tiene a continuación las notas de unos alumnos del curso de matemáticas, las mismas que son las siguientes:

    14, 16, 07, 11, 13, 09, 12, 15, 16, 08, 09, 16, 12, 13, 15

    Calcular lo siguiente:

    a) Percentil 25, 30, 33, 50, 75 y 80. Comente su respuesta para cada caso.
    b) Decil 3, 5, 6, 7 y 8. Comente su respuesta para cada caso.
    c) Quartil 1, 2 y 3. Comente su respuesta para cada caso.

12. El Gerente de Recursos Humanos de una Empresa Textil ha solicitado las horas extras que han trabajado un grupo de obreros en el I Semestre del año 2018, el

2018, el empleado que estaba elaborando el cuadro lo ha dejado inconcluso, solo registra algunos datos como se indican a continuación: Límite inferior del primer intervalo **300**, el límite máximo del sexto intervalo **600**, **f_1=2, F_2=5, f_3=4, F_4=14, H_4=0.70 y h_5=0.20**:

a. Completar la Tabla de Distribución de Frecuencias, y calcular P_{32}, D_7 y Q_2 comente su resultado que significa cada uno de ellos.
b. Calcular la desviación media
c. Calcular la desviación estándar y el coeficiente de variación, comente que significa el coeficiente de variación.
d. Calcular su Coeficiente de Asimetría a través de los Percentiles y comente su respuesta.

13. El Administrador de la Empresa J&E estaba procesando la información de las horas extras que han trabajado un grupo de obreros, el cuadro lo ha dejado inconcluso, solo registra algunos datos como se indican a continuación: Se sabe que tiene siete intervalos de frecuencia, de amplitud igual a 12, **F_1=8, f_7=8, f_6=20, ($m_3 \times f_3$=1260), ($f_2 + f_5$=62), h_3=0.21, h_5=0.21 y H_6=0.96** con estos datos desarrolle lo siguiente:

a. Construya la Tabla de Distribución de Frecuencias y calcule la Moda y Mediana y la varianza.
b. Calcule su Coeficiente de variación y comente su respuesta.
c. Calcule el Decil 6, Quartil 1, Percentil 25 comente sus respuestas.
d. Calcular su Coeficiente de Asimetría a través de los Quartiles y comente su respuesta.

14. En una empresa de Agroindustrial las Remuneraciones de sus 80 trabajadores de la Planta de producción tienen como sueldo mínimo $150 y su límite superior del intervalo máximo es de $300. Las remuneraciones se encuentran distribuidas en cinco intervalos de frecuencia de igual amplitud, además se sabe que el 10% de los trabajadores ganan al menos $150 pero menos de $180, el 30% gana menos de $210, el 55% gana menos de $240 y el 35% gana al menos $240 y menos de $270.

a. Completar la Tabla de Distribución de Frecuencias, y calcular P_{20}, P_{77}, Q_3 y Q_2 comente su resultado que significa cada uno de ellos.

c. El Sindicato de los trabajadores propone un aumento del 5% más una bonificación de \$20; cual sería el nuevo resultado de la Varianza, comente su respuesta.

d. Calcular su Coeficiente de Asimetría a través de los Media y Mediana y comente su respuesta.

12. En una Cia. Minera los Ingresos de 160 empleados administrativos están ordenados en los siguientes intervalos y se tiene conocimiento de que los trabajadores que tienen un Ingreso desde \$450 y menos de \$500 dólares su frecuencia relativa son mayores en 5% de los que perciben desde \$550.

Tabla N°66: *Tabla de frecuencia de ingresos de 160 empleados administrativos*

Clase	INTERVALOS	m_i	f_i	F_i	h_i
1	[250-300)		8		
2	[300-350)				
3	[350-400)		48	86	
4	[400-450)			120	
5	[450-500)				
6	[500-550)		16		
7	[550-600)				
			160		

Completar totalmente la Tabla de Distribución de Frecuencias y calcular los siguientes:

a) Percentil 25, 30, 33, 50, 75 y 80.
b) Decil 3, 5, 6, 7, 8 y 9.
c) Quartil 1, 2 y 3.

Fórmulas de Percentiles

Para datos no agrupados

$$P_k = x_{\frac{k(n+1)}{100}}$$

Para datos Agrupados

$$P_k = L_{inf} + \left(\frac{\frac{kn}{100} - F_{i-1}}{f_i} \right) c_i$$

Fórmulas de Deciles

Fórmulas de Deciles

Para datos no agrupados

$$D_k = x_{\frac{k(n+1)}{10}}$$

Para datos Agrupados

$$D_k = L_{inf} + \left(\frac{\frac{kn}{10} - F_{i-1}}{f_i} \right) c_i$$

Fórmulas de Quartiles para datos no agrupados

$$Q_k = x_{\frac{k(n+1)}{4}}$$

FORMULAS DE INTERES

Intervalos de Frecuencias

K = 1 + 3.3 log(n)

Media Aritmética

$$\overline{x} = \frac{\sum_{i=1}^{n} m_i f_i}{n}$$

$$\overline{x} = \frac{\sum_{i=1}^{n} m_i f_i}{n} = \frac{m_1 f_1 + m_2 f_2 + \ldots + m_{n-1} f_{n-1} + m_n f_n}{n}$$

Mediana: $$Me = L_{\inf} + \frac{\left(\frac{n}{2} - F_{i-1} \right) C_i}{f_i}$$

Moda: $$Mo = L_{\inf} + \left[\frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})} \right] C_i$$

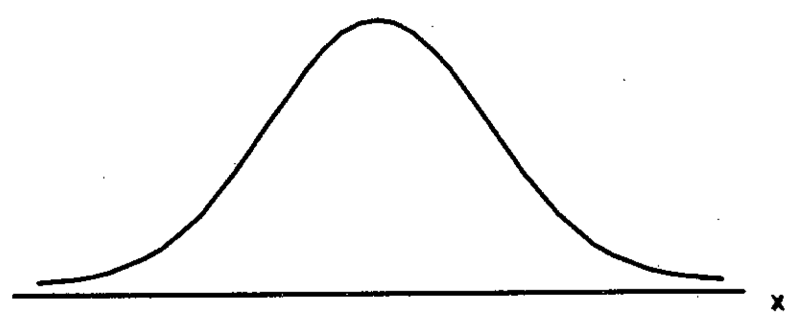
CAPITULO IV.

Puntuaciones Estándar, Asimetría y Kurtosis

4.1. Puntuaciones estándar y la curva normal.

DISTRIBUCIÓN NORMAL

Distribución Normal: Distribución en forma de campana que se logra con muestras de 100 o más unidades muestrales y que es útil y necesaria cuando se hacen inferencias estadísticas. Se caracterizar por ser un proceso que opera en condiciones normales, si tiene los materiales dentro de especificaciones y del mismo lote, un método consistente, un medio ambiente adecuado, el operador capacitado, y el equipo ajustado correctamente, si se toman mediciones en alguna característica del producto, mostrará el siguiente comportamiento:



$$-\infty < x < \infty$$

Figura N°15. Curva de la distribución normal

La siguiente función da origen a la gráfica de la curva normal

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

Propiedades

a) Es simétrica con respecto al eje que pasa por la media aritmética.
b) Los puntos de inflexión de la curva están en los puntos $\mu \pm s$

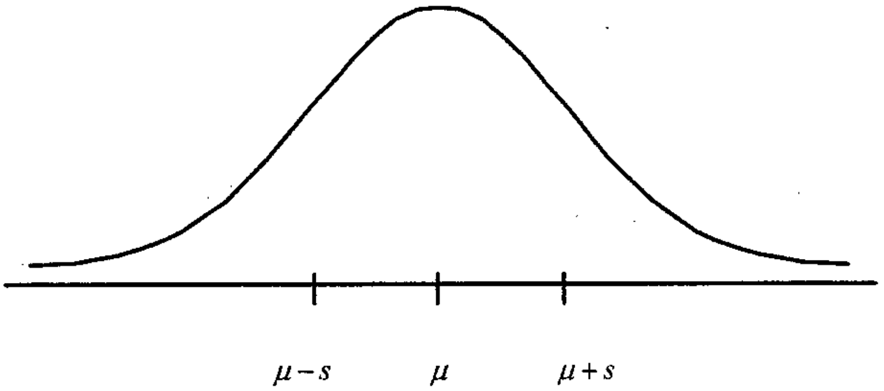


Figura N°16. La distribución normal con la media y desviación estándar

c) La curva es asintótica con respecto al eje horizontal.

d) $f(x) > 0$

e) El área bajo la curva es 1.

Distribución Normal Estándar

Esta distribución tiene $\mu = 0$ y desviación estándar $\sigma = 1$, la media aritmética, la mediana y la moda coinciden, y el valor de Z esta entre -3 y 3 pero pueden tomar valores mas extremos como +/- 4, +/- 5. aquí su área se pierde por ser muy baja.

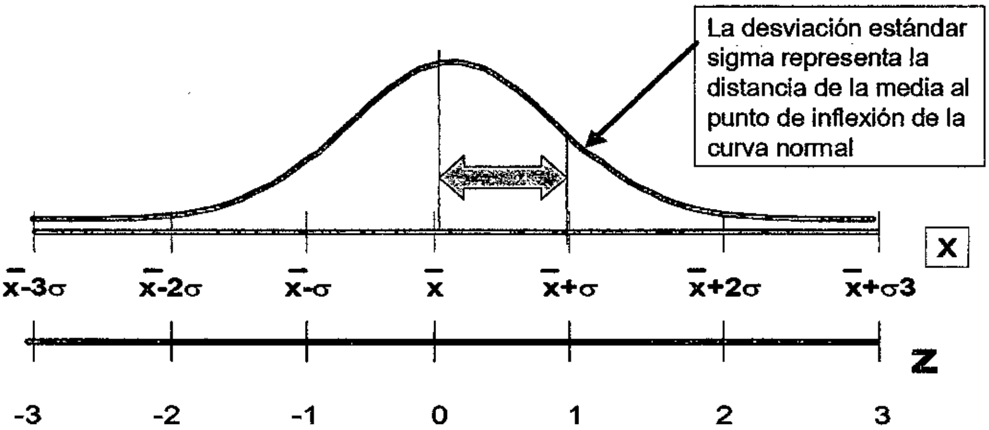


Figura N°17. La distribución normal con media 0 y desviaciones estándar

Puntuaciones estándar "Z"

Es una puntuación que se expresa en unidades de desviación estándar, su ubicación relativa de una puntuación bruta o directa es en relación a la media aritmética.

Su fórmula es: $Z = \frac{x_i - \bar{x}}{s}$

Ejemplo

Si la media aritmética de un grupo es 60 y la desviación estándar es 8, entonces, la puntuación estándar "z" corresponde a la puntuación bruta 68, e "1" puesto que la puntuación 69 se desvia una desviación estándar de la media aritmética 60. esto es:

$$Z = \frac{x_i - \bar{x}}{s} = \frac{68 - 60}{8} = 1$$

4.2. Asimetría y Kurtosis

Cálculo de Resultados de datos estadísticos.

En Asimetría analizamos la manera de informar resultados estadísticos con ética e imparcialidad. Aunque está aprendiendo a organizar, resumir e interpretar datos empleando la estadística, también es importante que comprenda la estadística con el fin de que se pueda procesar la información en forma inteligente usando el Excel aquí estamos aprendiendo la forma de calcular estadísticas descriptivas de naturaleza numérica. Específicamente la manera de calcular e interpretar medidas de ubicación para un conjunto de datos: (medidas de tendencia central) la media, la mediana y la moda, (medidas de dispersión) Rango, Desviación media, desviación cuartil, varianza y coeficiente de variación, así como también se calcula las asimetrías y la kurtosis. También podemos ver las ventajas y desventajas que nos brinda cada estadístico. Conocer las ventajas y desventajas de la media, la mediana y la moda es importante al dar un informe estadístico y cuando se emplea información estadística para tomar decisiones. Recuerde que el rango proporciona información sobre la dispersión total de una distribución. Sin embargo, no proporciona información sobre la forma en que se acumulan los datos o se concentran en torno al centro de la distribución. Conforme aprenda más estadística, necesitará recordar que cuando emplea estadísticas debe mantener un punto de vista independiente y con principios. Cualquier informe estadístico requiere la comunicación honesta y objetiva de los resultados

Kurtosis: Es la mayor o menor concentración de los datos en torno de media aritmética

$$K = \frac{(Q_3 - Q_1)}{2x(P_{90} - P_{10})}$$

Son las deformaciones graficas verticales que presentan las distribuciones de frecuencias. Las clases de kurtosis son:

Platicurtica: Si $K > 0.263$ Mesocurtica si: $K = 0.263$

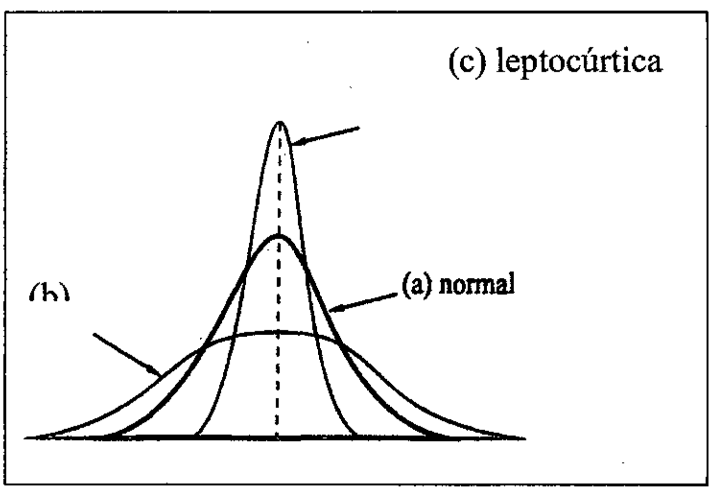
Leptocurtica si: $K < 0.263$

DISTRIBUCIONES SIMÉTRICAS

Las curvas simétricas son de 3 tipos:

Normal o mesocúrtica ; **Platicúrtica** y **Leptocúrtica**

Curvas simétricas; (a) normal, (b) platicúrtica, (c) leptocúrtica



Sesgo cero: moda = mediana = media

Figura N°18. La distribución normal y sus deformaciones graficas verticales

DISTRIBUCIONES ASIMÉTRICAS

Asimetría son las deformaciones graficas horizontales que presentan las distribuciones de frecuencias; las cuales se obtienen de unir las marcas de clases.

la distribución de frecuencias puede mostrar las siguientes asimetrías:

Las curvas asimétricas pueden ser de dos tipos:

Asimétricas positivas: Asimetría positiva: si la media aritmética se encuentra a la derecha de las otras medidas de tendencia central, cuando su sesgo es hacia la derecha.

Sesgo a la derecha: media y mediana se encuentran a la derecha de la moda.

O sea que se cumple: Moda < mediana < media

(o de cola al lado derecho), figura (a)

Asimetría negativa: si la media aritmética se encuentra a la izquierda de la mediana y la moda, cuando su sesgo es hacia la izquierda.

O sea que se cumple la desigualdad: Media < mediana < moda

(o de cola a la izquierda), fig. (b)

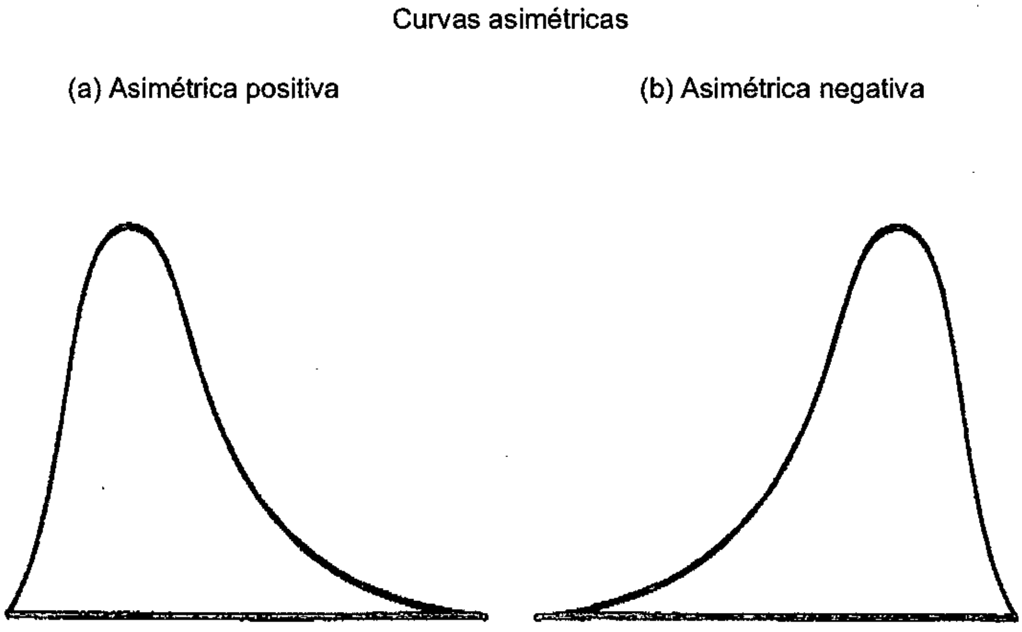


Figura N°19. La distribución normal y sus sesgos positivo y negativo

Coeficiente de Asimétrica (As)

Sirve para determinar el tipo de asimetría que presenta una distribución de frecuencias:

Si: As = 0 es simétrica

As > 0 es asimétrica positiva

As < 0 es asimétrica negativa

Coeficientes de Pearson

$$As_1 = \frac{\bar{X} - Mo}{s} \qquad As_2 = \frac{3(\bar{X} - Me)}{s}$$

$$As_3 = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1} \qquad As_4 = \frac{P_{90} - 2P_{50} + P_{10}}{P_{90} - P_{10}}$$

Donde: P_{10} percentil 10 ; P_{50} percentil 50 ; P_{90} percentil 90 y

Q_1 cuartil 1; Q_2 cuartil 2; Q_3 cuartil 3,

Ejemplo1. Para los siguientes datos discretos elaborados de la siguiente manera:

Primero haga el sorteo aleatorio (en Excel) hasta la tercera vez entre 16 y 86 (edades de las personas que van al cine a ver la película de estreno) entre las columnas A y J y ocho filas de tal manera que tenga 80 datos aleatorios.

H) Determine el número de intervalos y el tamaño del intervalo de clase, para variable cuantitativa continua, usando la regla de Sturges

I) Elabore el cuadro con todas las frecuencias estudiadas usando Excel, para las especificaciones de (A) y su grafica de un histograma de Frecuencia abs. Simple

J) Halle las medidas de tendencia central

K) Halle las medidas de dispersión

L) Halle todas las asimetrías estudiadas y la kurtosis.

Solución

Tabla N°67: *Tabla de 80 datos aplicando la regla de Sturges en Excel*

G26 f_x

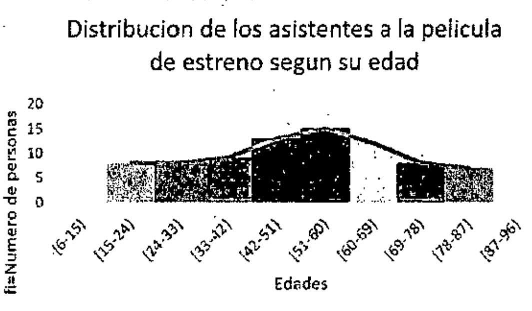
	A	B	C	D	E	F	G	H	I	J
16	70	47	77	34	77	60	30	28	27	56
17	51	23	84	65	86	51	79	44	64	37
18	54	78	19	67	62	47	69	16	44	17
19	58	41	49	60	20	27	61	21	81	77
20	43	68	37	52	40	47	26	40	34	44
21	59	66	37	32	42	84	19	28	78	77
22	56	55	16	75	49	49	66	56	77	55
23	45	55	64	63	28	48	51	59	59	38
24	Vmax=	86								
25	Vmin=	16								
26	Rango=	70								
27	K=1+3.3*log(n)	7.28019696	8							
28	TIC=R/K	8.75	9							
29	D=TIC*K-R	2								

Tabla N°68: *Tabla de distribución de los datos en frecuencias en Excel*

IC	fi	hi	Fi	Hi	Fi*	Hi*	hi*%	Hi*%	mi	mi*fi	mi-MA	(mi-MA)	fi*(mi-MA)	$(mi-MA)^2$	$fi*(mi-MA)^2$
[6-15)			0												
[15-24)	8	0.1	8	0.1	80	1	10	10	19.5	156	-31.725	31.725	253.8	1006.47563	8051.805
[24-33)	8	0.1	16	0.2	72	0.9	10	20	28.5	228	-22.725	22.725	181.8	516.425625	4131.405
[33-42)	9	0.1125	25	0.3125	64	0.8	11.25	31.25	37.5	337.5	-13.725	13.725	123.525	188.375625	1695.38063
[42-51)	13	0.1625	38	0.475	55	0.6875	16.25	47.5	46.5	604.5	-4.725	4.725	61.425	22.325625	290.233125
[51-60)	15	0.1875	53	0.6625	42	0.525	18.75	66.25	55.5	832.5	4.275	4.275	64.125	18.275625	274.134375
[60-69)	12	0.15	65	0.8125	27	0.3375	15	81.25	64.5	774	13.275	13.275	159.3	176.225625	2114.7075
[69-78)	8	0.1	73	0.9125	15	0.1875	10	91.25	73.5	588	22.275	22.275	178.2	496.175625	3969.405
[78-87)	7	0.0875	80	1	7	0.0875	8.75	100	82.5	577.5	31.275	31.275	218.925	978.125625	6846.87938
[87-96)															
	80	1								4098			1241.1		27373.95

Tabla N°69: *Resultados de las MTC, MD , gráfica y Asimetría en Excel*

B	C	D	E	F	G	H	I	J
MTC			MD			Simetria		
MA=	51.225		Rango=	72		As1=	-0.1824526	
Me=Q2=P50	52.2		DQ=	13.1875		As2=	-0.1581256	
Mo=	54.6		DM=	15.51375		As3=	-0.1526066	
Q1=	37		S^2=	342.174375		As4=	-0.0666667	
Q2=	52.2		S=	18.49795597		K=	0.24940898	Platicurtica
Q3=	63.375	Ubicación Percen	C.V=	36.11118783	36.11%			
P23=	35.4	18.4						
P44=	49.0615385	35.2						
P58=	56.04	46.4						
P78=	67.05	62.4						
P69=	61.65	55.2						
P83=	70.575	66.4						
P90=	76.875	72						
P10=	24	8						
P50=	52.2	40						



Ejemplo 2. Dado el siguiente conjunto de datos:

Tabla N°70: *Tabla de frecuencia de 50 datos acumulados en intervalos*

C.I.	f_i
10 - 20	*3*
20 - 30	*10*
30 – 40	*9*
40 – 50	*11*
50 – 60	*8*
60 – 70	*7*
70 – 80	*2*
Total	*50*

Calcular: Rango , Desviación cuartil. Desviación media. Varianza. Desviación estándar. Coeficiente de variación. Hallar Todos los coeficientes de Asimetría y la Kurtosis

Solución: Trabajando en Excel:

Primero en insertar función concatenamos los intervalos luego en la columna de la frecuencia absoluta simple ingresamos los datos que estamos analizando (fi) que son los que nos dan. Luego elaboramos un plan de completar en la tabla las fórmulas de las medidas de dispersión (Desviación cuartil, desviación media, varianza, desviación estándar y coeficiente de variación). Luego usamos las fórmulas de las medidas de Asimetría y Kurtosis.

Tabla N°71: *Tabla de distribución de los datos en frecuencias y solución en Excel*

		IC	fi	Fi	mi	mi*fi	mi-MA	mi-MA	fi*(mi-MA)	$(mi-MA)^2$	$fi*(mi-MA)^2$
10	20	[10-20)	3	3	15	45	-28	28	84	784	2352
20	30	[20-30)	10	13	25	250	-18	18	180	324	3240
30	40	[30-40)	9	22	35	315	-8	8	72	64	576
40	50	[40-50)	11	33	45	495	2	2	22	4	44
50	60	[50-60)	8	41	55	440	12	12	96	144	1152
60	70	[60-70)	7	48	65	455	22	22	154	484	3388
70	80	[70-80)	2	50	75	150	32	32	64	1024	2048
			50			2150			672		12800

Mo=	44
MA=	43
Q3=	55.625
Q1=	29.5
Q2=Me=P50=	42.7272727
P10=	22
P90=	65.7142857

Medidas de Dispersión

DQ=	13.0625	
DM=	13.44	
S^2=	256	
S=	16	
CV=	37.2093023	37.21%

Asimetrias		
As1=	-0.0625	
As2=	0.05113636	
As3=	-0.0126142	
As4=	0.0516934	
K=	0.29881536	Leptocurtica

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

$$As = \frac{P_{90} - 2P_{50} + P_{10}}{P_{90} - P_{10}} \qquad As_2 = \frac{3(\overline{X} - Me)}{s} \qquad K = \frac{(Q_3 - Q_1)}{2 \times (P_{90} - P_{10})} \qquad As_1 = \frac{\overline{X} - Mo}{s}$$

Ejemplo 3. Dado el siguiente conjunto de datos en la tabla:

Tabla N°72: *Tabla de frecuencia de 50 datos agrupados en intervalos*

C.I.	f_i
5 - 10	*2*
10 - 15	*7*
15 – 20	*11*
20 – 25	*9*
25 – 30	*10*
30 – 35	*8*
35 – 40	*3*
Total	*50*

Calcular: *Rango . Desviación cuartil. Desviación media.*

- *Varianza. Desviación estándar. Coeficiente de variación.*
- *Hallar Todos los coeficientes de Asimetria y la Kurtosis*

Solución en Excel

Tabla N°73: *Tabla de distribución de los datos en frecuencias y solución en Excel*

		IC	fi	Fi	mi	mi*fi	mi-MA	mi-MA	fi*(mi-MA)	$(mi-MA)^2$	$fi*(mi-MA)^2$
5	10	[5-10)	2	2	7.5	15	-15.4	15.4	30.8	237.16	474.32
10	15	[10-15)	7	9	12.5	87.5	-10.4	10.4	72.8	108.16	757.12
15	20	[15-20)	11	20	17.5	192.5	-5.4	5.4	59.4	29.16	320.76
20	25	[20-25)	9	29	22.5	202.5	-0.4	0.4	3.6	0.16	1.44
25	30	[25-30)	10	39	27.5	275	4.6	4.6	46	21.16	211.6
30	35	[30-35)	8	47	32.5	260	9.6	9.6	76.8	92.16	737.28
35	40	[35-40)	3	50	37.5	112.5	14.6	14.6	43.8	213.16	639.48
			50			1145			333.2		3142

Mo=	18.3333333
MA=	22.9
Q3=	29.25
Q1=	16.5909091
Q2=Me=P50=	22.7777778
P10=	12.1428571
P90=	33.75

Medidas de Dispersión		
DQ=	6.32954545	
DM=	6.664	
S^2=	62.84	
S=	7.92716847	
CV=	34.6164562	34.62%

Asimetrias		
As1=	0.57607791	
As2=	0.04625443	
As3=	0.02254139	
As4=	0.01561065	
K=	0.29293764	Leptocurtica

Ejemplo 4. (Para los siguientes 18 datos discretos:

19, 10, 9, 11, 12, 8, 11, 14, 3, 17, 13, 6, 10, 7, 16, 9, 18, 9.

Calcular: Las Medidas de Tendencia Central (MTC), Las Medidas de Dispersión (MD) y Todas las Asimetrías

Tabla N°74: *Tabla de distribución de los datos discretos y solución en Excel*

fx

B	C	D	E	F	G	H	I	J
xi	Orden	xi	xi-MA	xi-MA	$(xi-MA)^2$	Mo=	9	
3	1	19	-8.222222	8.222222	67.60493462	MA=	11.2222222	
6	2	10	-5.222222	5.222222	27.27160262	Me=Q2=P50=	10.5	
7	3	9	-4.222222	4.222222	17.82715862	Q1=	8.75	
8	4	11	-3.222222	3.222222	10.38271462	Q2=	10.5	
9	5	12	-2.222222	2.222222	4.938270617	Q3=	14.5	
9	6	8	-2.222222	2.222222	4.938270617	P10=	5.7	
9	7	11	-2.222222	2.222222	4.938270617	P90=	18.1	
10	8	14	-1.222222	1.222222	1.493826617		Med. Dispersión	
10	9	3	-1.222222	1.222222	1.493826617	Rango=	16	
11	10	17	-0.222222	0.222222	0.049382617	DQ=(Q3-Q1)/2	2.875	
11	11	13	-0.222222	0.222222	0.049382617	DM=	3.382716	
12	12	6	0.777778	0.777778	0.604938617	S2=	18.5359477	
13	13	10	1.777778	1.777778	3.160494617	S=	4.30533944	
14	14	7	2.777778	2.777778	7.716050617	CV=	38.3644109	0.3836
16	15	16	4.777778	4.777778	22.82716262	Asimetrias		
17	16	9	5.777778	5.777778	33.38271862	As1=	0.51615494	
18	17	18	6.777778	6.777778	45.93827462	As2=	0.50325107	
19	18	9	7.777778	7.777778	60.49383062	As3=	0.39130435	
				60.888888	315.1111111	As4=	0.22580645	
						K=	0.23185484	Platicurtica

Ejercicios de Aplicación

1.- En una empresa de Metal Mecánica se ha aplicado una prueba psicotécnica a 46 trabajadores, habiendo obtenidos las siguientes puntuaciones en el cuadro que se presenta a continuación: La distribución es simétrica con 7 intervalos de frecuencia, $\mathbf{f_1=10}$, $\mathbf{f_3=50}$, $\mathbf{F_3=90}$, $\mathbf{H_3=0.36}$, $\mathbf{m_4=65}$y el Límite inferior del tercer intervalo es igual a **50**, completar la Tabla de distribución de frecuencias.

1.1 Completar la Tabla de Distribución de Frecuencias

1.2 Calcular el Coeficiente de Asimetría a través de la Media y Mediana.

1.3 Calcular el Coeficiente de Asimetría a través de los Percentiles.

1.4 Calcular el Coeficiente de Asimetría a través de los Cuartiles.

1.5 Modificar las siguientes frecuencias absolutas $f_5=12$, $f_6=9$ y $f_8=8$ y calcular el coeficiente de Asimetría a través de de la Media y Mediana,

1.6 Calcular el coeficiente de Kurtosis.

2.- El Gerente de Recursos Humanos de una Empresa Textil ha solicitado las horas extras que han trabajado los obreros en el I Trimestre del año en curso, el empleado que estaba elaborando el cuadro lo ha dejado inconcluso, solo registra algunos datos como se muestran en el cuadro siguiente:

Tabla N°75: *Tabla de frecuencia de horas extras de una empresa Textil*

CLASE	INTERVALOS	m_i	f_i	h_i	F_i	H_i
1	[250-300)				2	
2	[300-350)				5	
3	[350-400)				9	
4	[400-450)				14	0.7
5	[450-500)			0.2		
6	[500-550)					

2.1 Calcular el Coeficiente de Asimetría a través de la Media y Moda.

2.2 Calcular el Coeficiente de Asimetría a través de la Media y Mediana.

2.3 Calcular el Coeficiente de Asimetría a través de los Percentiles.

2.4 Calcular el Coeficiente de Asimetría a través de los Cuartiles.

2.5 Calcular la Kurtosis.

CAPITULO V.

Correlación y Regresión

5.1. Introducción al Análisis Multivariado

Se estudia la interacción de "n" variables: Ejemplo: Regresión Múltiple, Análisis Discriminante, etc.

Nosotros en este trabajo analizaremos la correlación y regresión bivariado, la multivariabilidad de materia de un estudio más elaborado con procesos estocásticos complejos. Ahora entramos a:

5.2. Correlación

La correlación es una técnica estadística que mide la relación (covariación) entre dos o más variables cuantitativas continuas (x, y). La manera más sencilla de saber si dos variables están correlacionadas es determinar si co-varían (varían conjuntamente) los datos son bivariados. Es importante hacer notar que esta covariación no implica necesariamente causalidad.

Coeficiente de correlación

Estadístico que cuantifica la correlación cuyos valores están comprendidos entre: -1 y 1

Fuente:
http://www.ccg.unam.mx/~vinuesa/R4biosciences/docs/Tema8_correlacion.pdf

- La correlación r puede variar entre -1 y +1, ambos extremos indicando correlaciones perfectas, negativa y positiva respectivamente.
- Un valor de r = 0 indica que no existe relación lineal entre las dos variables.
- Una correlación positiva indica que ambas variables varían en el mismo sentido.

- Una correlación negativa significa que ambas variables varían en sentidos opuestos.
- Se interpreta de la siguiente manera:
  1. **Si $r = 1$,** existe una correlación positiva perfecta. El índice indica una dependencia total entre las dos variables denominada *relación directa*: cuando una de ellas aumenta, la otra también lo hace en proporción constante.
  2. **Si $0 < r < 1$,** existe una correlación positiva.
  3. **Si $r = 0$,** no existe relación lineal. Pero esto no necesariamente implica que las variables son independientes: pueden existir todavía relaciones no lineales entre las dos variables.
  4. **Si $-1 < r < 0$,** existe una correlación negativa.
  5. **Si $r = -1$,** existe una correlación negativa perfecta. El índice indica una dependencia total entre las dos variables llamada *relación inversa*: cuando una de ellas aumenta, la otra disminuye en proporción constante

Tabla N°76: *Tabla de interpretación de correlaciones entre dos variables*

CORRELACIÓN NEGATIVA	
-0.95	Muy Fuerte
-0.75	Considerable
-0.5	Media
-0.1	Débil
0	Inexistente

CORRELACIÓN POSITIVA	
1	Perfecta
0.95	Muy Fuerte
0.75	Considerable
0.5	Media
0.1	Débil
0	Inexistente

Tabla N°77: *Tabla de uso del estadístico para el nivel de medición*

	Nominal	Ordinal	Intervalo	Razón	Dicotómica	Rangos
Nominal	V de Cramer Tschuprow's *T*			Poliserial		
Ordinal		Policórica -Likert Spearman	Spearman Kendall-tau Poliserial		Rango biserial	
Intervalo			Pearson	Pearson		
Razón				Pearson	Punto-biserial (natural) Coeficiente biserial (creada)	
Dicotómica					Coeficiente Phi (natural) Tetracórica (creada)	
Rangos						Spearman-Rangos

Fuente: . http:// www.ccg.unam.mx/ ~vinuesa/

Coeficientes de correlación por nivel de medición de las variables

1. Coeficiente de Pearson: r_P , es para dos variables cuantitativas.
2. Coeficientes de rangos de Spearman r_S , para dos variables de rangos.
3. Tau de Kendall: tau , para dos variables de rango.
4. Coeficiente phi: φ , para dos variables dicotómicas naturales.
5. V de Cramer: V , para dos variables nominales.
6. Coeficiente tetracórico: r_t , para dos variables dicotómicas creadas.

7. Coeficiente punto-biseral: r_{pb} , para una variable cuantitativa con una variable dicotómica natural.
8. Coeficiente biseral: r_b , para una variable cuantitativa con una variable dicotómica creada.

Correlación Policórica

Calcula la correlación policórica (y su error estándar) entre dos variables ordinales o de su tabla de contingencia, bajo el supuesto de que las variables ordinales analizan las variables continuas latentes que son bivariadas normales. O bien está disponible el estimador de máxima verosimilitud o una aproximación de "dos pasos" (posiblemente mucho) más rápida. Para el estimador de ML, las estimaciones de los umbrales y la matriz de covarianza de las estimaciones también están disponibles.

Fuente:https://www.rdocumentation.org/packages/DescTools/versions/0.99.19/topics/CorPolychor

Coeficiente de correlación de Pearson

La correlación más usada y más popular es el Coeficiente de Correlación de Pearson. Para su aplicación es indispensable que la correlación sea lineal cuando ambas variables son cuantitativas y siguen una distribución normal.

El coeficiente de correlación de Pearson, que se simboliza con la letra minúscula r, se calcula dividiendo la suma de los productos de las desviaciones de cada variante de X e Y, con respecto a sus medias (suma que se denomina covarianza de X e Y), por el producto de las desviaciones estándar de ambas variables. En forma práctica, el coeficiente de correlación de Pearson es:

$$r=\frac{n\sum_{i=1}^{n}(xy)-\left(\sum_{i=1}^{n}x\right)\left(\sum_{i=1}^{n}y\right)}{\sqrt{\left[n\sum_{i=1}^{n}x^2-\left(\sum_{i=1}^{n}x\right)^2\right]\left[n\sum_{i=1}^{n}y^2-\left(\sum_{i=1}^{n}y\right)^2\right]}}$$

Coeficiente de correlación de Spearman

Se usa solo cuando se han logrado mediciones en el nivel ordinal o siendo las mediciones cuantitativas se desea determinar la correlación entre los rangos correspondientes, se utiliza el coeficiente por rangos de Spearman.

$$r_S = 1 - \frac{6\sum d_i^2}{n(n^2-1)}, \text{ donde: } d = R_x - R_y$$

Donde R_x = Rango de los valores de "x"

R_y = Rango de los valores de "y"

Ejemplo: Se han evaluado el coeficiente de inteligencia y su rendimiento académico del ciclo 2017-B, de 10 estudiantes.

Tabla N°78: *Tabla de cálculo de la correlación de Pearson y Spearman*

x=Integ	y=RA	x*y	x^2	y^2	R_x	R_y	d	d^2
114	7	798	12996	49	7	5.5	1.5	2.25
118	8	944	13924	64	5.5	3	2.5	6.25
103	5	515	10609	25	9	9.5	-0.5	0.25
124	7	868	15376	49	4	5.5	-1.5	2.25
137	9	1233	18769	81	2	2	0	0
126	7	882	15876	49	3	5.5	-2.5	6.25
112	6	672	12544	36	8	8	0	0
142	10	1420	20164	100	1	1	0	0
118	7	826	13924	49	5.5	5.5	0	0
99	5	495	9801	25	10	9.5	0.5	0.25
Suma:1193	71	8653	143983	527				17.5

$$r_S = 1 - \frac{6\sum d_i^2}{n(n^2-1)} = 0.8939\ ,\ r_P = 0.937595354$$ luego podemos decir que hay una

Correlación Positiva muy fuerte entre la inteligencia y el RA.

5.3. Regresión

Coeficiente de Determinación

Es el cuadrado del coeficiente de correlación. Expresado en tanto por ciento mide el grado de información compartida entre dos variables continuas

Coeficientes de Regresión

Determinación de la ecuación matemática

En la regresión, los valores de y son predichos a partir de valores de x dados o conocidos. La variable "y" recibe el nombre variable dependiente y la variable "x" el de variable independiente.

En el ajuste de funciones de regresión simple, se puede utilizar diversas funciones matemáticas, tales como:

a) La línea Recta $Y^* = a + bX$

b) La Parábola $Y^* = a + bX + cX^2$

c) La curva Potencial $Y^* = bX^a$

d) La curva Exponencial $Y^* = a\,b^X$

e) La Hipérbola Equilátera $Y^* = \frac{a}{X}$

f) La curva Logística $\frac{1}{Y} = a + b\,c^X$

g) La Curva de Gompertz $Y^* = a\,b^{c^X}$

Cada una de estas funciones tiene una forma particular para un conjunto de puntos.

El problema de ajuste de una función de regresión a un conjunto de n valores (X,Y) comprende tres pasos:

i) Graficar el diagrama de esparcimiento o una nube de puntos (X,Y).

ii) Definir la forma de la función de regresión (recta , parábola, exponencial, etc.)

iii) Determinar el valor numérico de los parámetros de la función elegida. Loa parámetros de la función de regresión se obtienen a partir de las Ecuaciones normales obtenidos por el método de los Mínimos cuadrados.

En un modelo de regresión lineal son los valores de a y b que determinan la expresión de la recta de regresión $y = a + b\ x$

La regresión lineal simple comprende el intento de ajustar una nube de puntos en una línea recta o ecuación matemática lineal que describe la reacción entre dos variables.

La regresión puede ser utilizadas de diversas formas. Se emplean en situaciones en la que las dos variables miden aproximadamente lo mismo, pero en las que una variable es relativamente costosa, o, por el contrario, es poco interesante trabajar con ella, mientras que con la otra variable no ocurre lo mismo.

La finalidad de una ecuación de regresión seria estimar los valores de una variable con base en los valores conocidos de la otra.

Otra forma de emplear una ecuación de regresión es para explicar los valores de una variable en término de otra. Es decir, se puede intuir una relación de causa y efecto entre dos variables. El análisis de regresión únicamente indica qué relación matemática podría haber, de existir una. Ni con regresión ni con la correlación se pude establecer si una variable tiene "causa "ciertos valores de otra variable.

Ecuación Lineal

Dos características importantes de una ecuación lineal

- la independencia de la recta
- la localización de la recta en algún punto. Una ecuación lineal tiene la forma

$$\mathbf{y = a + bx}$$

En la que a y b son valores que se determina a partir de los datos de la muestra; a indica la altura de la recta en x= 0, y b señala su pendiente. La variable y es la que se habrá de predecir, y x es la variable predictora.

Regresión lineal múltiple

El modelo de regresión lineal múltiple sirve para explicar una variable respuesta continua en términos de varios factores o variables explicativas continuas en donde el SPSS nos ayuda mucho en este modelo.

Regresión polinómica

Es un tipo especial de regresión múltiple donde aparecen como variables independientes un única variable y potencias de ésta (al cuadrado, al cubo.)

El análisis de regresión sirve para predecir una medida en función de otra medida (o varias).

Y = Variable dependiente

A esta variable se le denomina variable predicha, también se le considera como variable explicada

X = Variable independiente

A esta variable se le denomina variable predictora, también se le considera como variable explicativa.

¿Es posible descubrir una relación?

$$Y = f(x) + \text{error}$$

Donde f es una función de un tipo determinado. Podemos decir que el error es aleatorio, pequeño, y no depende de x..

El modelo de Regresión presenta:

- La bondad de un ajuste de un modelo de regresión se mide usando el coeficiente de determinación R^2
- R^2 es una cantidad adimensional que sólo puede tomar valores en [0, 1]

El estudiante podría hacerse algunas preguntas

¿Cuándo un ajuste es bueno?

La respuesta más acertada es R^2 será cercano a uno.

¿por qué? El porcentaje de eficiencia del modelo se acerca a 100%

¿Cuando un ajuste es malo?

Cuando R^2 será cercano a cero.

¿por qué? El porcentaje de eficiencia del modelo se acerca a 0%

En realidad, a R^2 también se le denomina porcentaje de variabilidad explicado por el modelo de regresión.

¿por qué es difícil calcular R^2 ?

Podemos decir que R^2 puede ser pesado de calcular en modelos de regresión general, pero en el modelo lineal simple, la expresión es de lo más sencilla: $\mathbf{R^2=r^2}$

Otros modelos de Regresión

- Se pueden considerar otros tipos de modelos, en función del aspecto que presente el diagrama de dispersión (regresión no lineal)
- Incluso se puede considerar el que una variable dependa de varias (regresión múltiple).

Recta o parábola

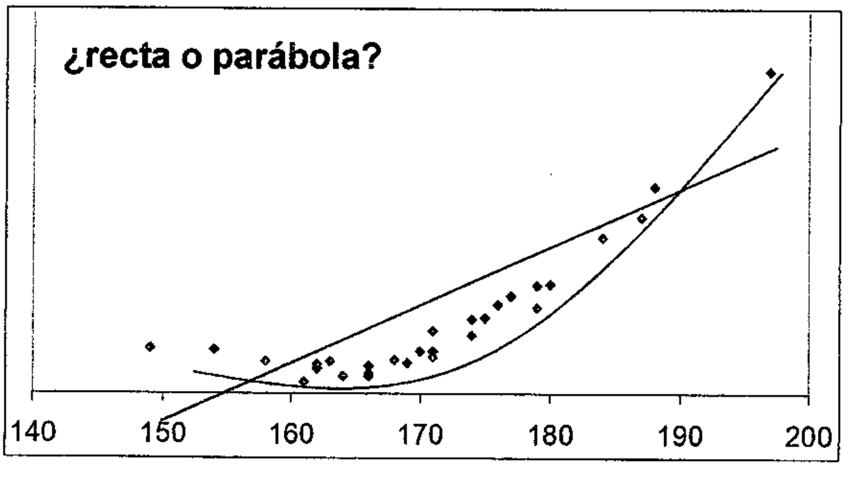


Figura N°20. Grafica de una recta y parábola de ajuste a la dispersión de puntos

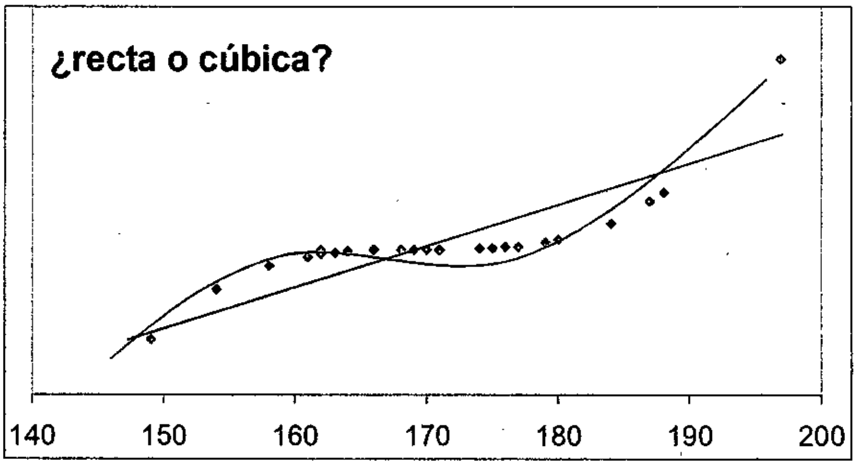


Figura N°21. Grafica de una recta y cúbica de ajuste a la dispersión de puntos

¿Cómo podemos medir la bondad de una regresión?

Imaginemos un diagrama de dispersión, y vamos a tratar de comprender en primer lugar qué es el error residual, su relación con la varianza de Y, y de ahí, cómo medir la bondad de un ajuste.

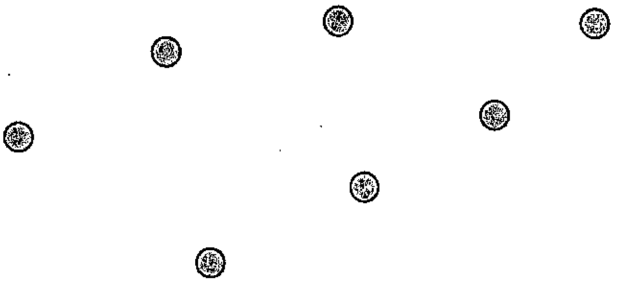


Figura N°22. Grafica de la dispersión de puntos

Interpretación de la variabilidad en Y

Aquí veamos cuál es la variabilidad en el eje Y.

La franja sombreada indica la zona donde varían los valores de Y.

Proyección sobre el eje Y

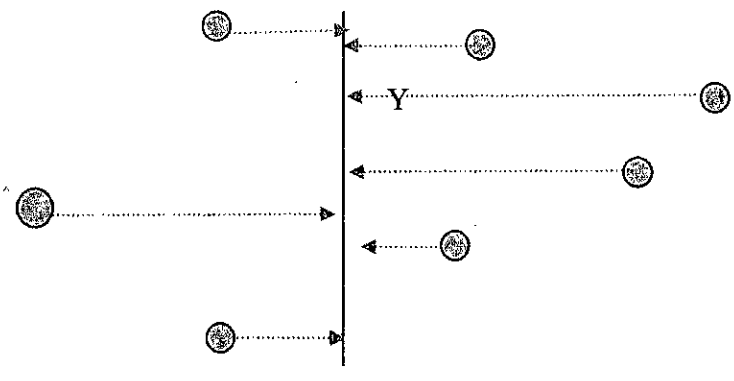


Figura N°23. Grafica de la dispersión de puntos y la variabilidad en Y

Cuantos menos dispersos sean los residuos, mejor será la bondad del ajuste

Interpretación del residuo

Veamos ahora en los errores de predicción (líneas verticales). Los proyectamos sobre el eje Y.
Se observa que los errores de predicción, residuos, están menos dispersos que la variable Y original.

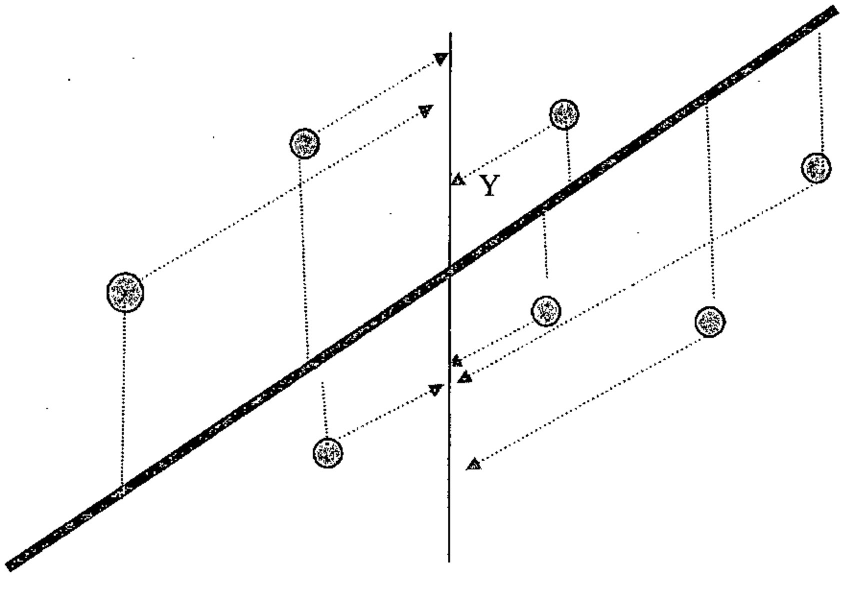


Figura N°24. Grafica de la dispersión de puntos y recta de ajuste

Bondad de ajuste

Se tiene que:

- La dispersión del error residual será una fracción de la dispersión original de Y
- Cuanto menor sea la dispersión del error residual mejor será el ajuste de regresión.

Luego definimos como medida de bondad de ajuste de regresión, o coeficiente de determinación a:

Ejemplo El consumo de cosméticos en el Callao expresado por las ventas, es evaluado en función de la cantidad de distribuidores de los mismos. La evolución en el tiempo se refleja en las siguientes cantidades:

Tabla N°79: *Ventas de Cosméticos en el Callao en función de los distribuidores*

Cantidad de Distribuidoras	Ventas de Cosméticos
x	y
1	8
3	104
7	1311
11	5338
15	13511
18	23310
20	31994

Se pide: a) Graficar los valores.

b) Ajustar dichos valores empleando diferentes modelos (lineal y cuadrático). Mostrar sus gráficas.

c) Indicar cuál es el modelo más adecuado para el análisis de regresión. Fundamente su respuesta.

d) Calcular la función y = f(x) que mejor represente el modelo.

e) Estimar cuanto será la venta para 25 distribuidoras del producto.

Solución: Usaremos Excel para resolver nuestro problema a) y b) :

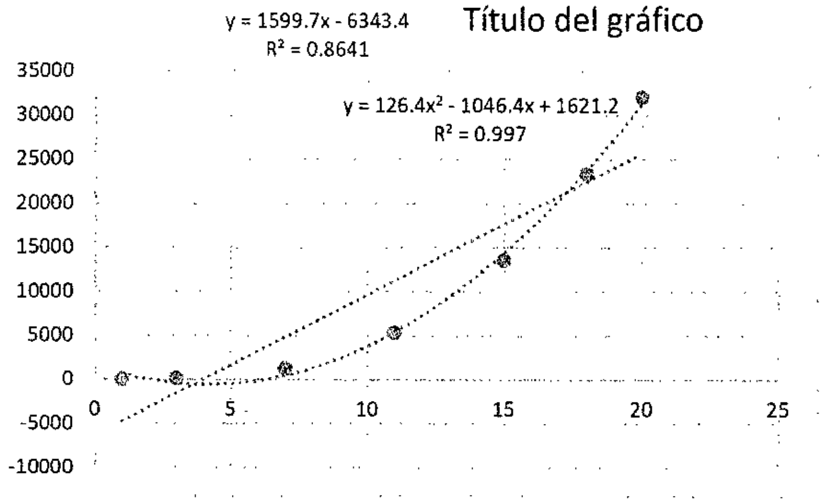


Figura N°25. Gráfica de Regresión Lineal y Cuadrática

Solución (c) El modelo adecuado es la parabólica porque su R^2 se acerca a 1.

Solución (d) $y = 126.4x^2 - 1046.4x + 1621.2$

Solución (e) la venta para 25 distribuidoras del producto es 33649.1

Ejemplo: La siguiente tabla muestra las cotizaciones alcanzadas por las acciones de una empresa durante un mes, en relación a la cantidad de días consecutivos que se mantuvieron en alza:

Tabla N°80: *Tabla de cotizaciones alcanzadas por las acciones de una empresa*

N° de días	Cotización
x	y
3	90
4	96
5	98
8	103
10	109

a) Construir una gráfica de dispersión de los datos. ¿Se puede usar un modelo de regresión simple?

b) Trazar la recta de regresión, y la regresión cuadrática y presentar las ecuaciones y el coeficiente de terminación en el gráfico. Interprete el significado de r^2 para la ecuación lineal.

c) Calcular la cotización cuando el número de días es 7.

Solución: Usaremos Excel para resolver nuestro problema

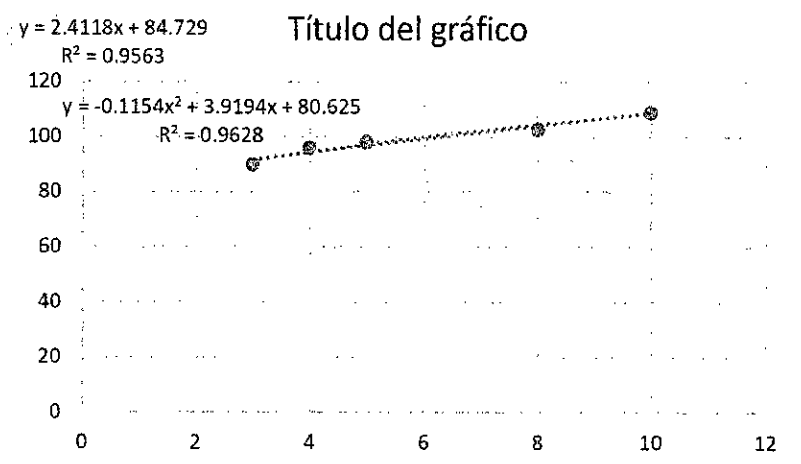


Figura N°26. Gráfica de Regresión Lineal de las cotizaciones.

Solución (a) No porque la cuadrática el R^2 es mayor que el R^2 de la regresión lineal.

Solución (b) $R^2 = 0.9563$, significa que el 95.63% influye el número de días en la cotización en alza.

Solución (c) la cotización en el séptimo día es: 101.6116

Ejemplo: Se realizó un estudio de tallas de hijos y padres de 30 alumnos del 2do. Ciclo 2017-B de Psicología de la URP y estos fueron los resultados.

a) ¿Qué correlación hay entre la talla de padres e hijos? **de su interpretación.**

b) Halle la ecuación de regresión simple entre las: TP y TH

c) ¿Qué talla se espera de un hijo cuyo padre mide 179 cm. de estatura?

A que conclusión puede llegar Usted

Tabla N°81: *Cálculo de la correlación de Pearson y Spearman de TP y TH*

TP=x	TH=y	x^2	y^2	x*y	Rx	Ry	d=Rx-Ry	d^2
177	170	31329	28900	30090	9.5	3.5	6	36
174	171	30276	29241	29754	5.5	6	-0.5	0.25
173	174	29929	30276	30102	3.5	10	-6.5	42.25
179	169	32041	28561	30251	15	1	14	196
174	170	30276	28900	29580	5.5	3.5	2	4
179	175	32041	30625	31325	15	13.5	1.5	2.25
180	173	32400	29929	31140	18.5	8	10.5	110.25
173	175	29929	30625	30275	3.5	13.5	-10	100
178	170	31684	28900	30260	11.5	3.5	8	64
185	179	34225	32041	33115	25.50	22	3.5	12.25
177	170	31329	28900	30090	9.5	3.5	6	36
172	174	29584	30276	29928	2	10	-8	64
178	183	31684	33489	32574	11.5	26.5	-15	225
181	178	32761	31684	32218	20.5	19	1.5	2.25
175	178	30625	31684	31150	7	19	-12	144
179	175	32041	30625	31325	15	13.5	1.5	2.25
181	178	32761	31684	32218	20.5	19	1.5	2.25
185	181	34225	32761	33485	25.5	24	1.5	2.25
188	189	35344	35721	35532	29.5	29	0.5	0.25
182	177	33124	31329	32214	22.5	16	6.5	42.25
185	178	34225	31684	32930	25.5	19	6.5	42.25
182	183	33124	33489	33306	22.5	26.5	-4	16
179	181	32041	32761	32399	15	24.00	-9	81
179	172	32041	29584	30788	15	7	8	64
170	174	28900	30276	29580	1	10	-9	81
176	175	30976	30625	30800	8	13.5	-5.5	30.25
185	178	34225	31684	32930	25.5	19	6.5	42.25
188	189	35344	35721	35532	29.5	29	0.5	0.25
187	181	34969	32761	33847	28	24	4	16
180	189	32400	35721	34020	18.5	29	-10.5	110.25
Suma:5381	5309	965853	940457	952758				1571

Nota: TP; Talla de padres.   TH: Talla de hijos

Soluc. a) $r_S = 1 - \frac{6\sum d_i^2}{n(n^2-1)} = 0.65050056$, $r_P = 0.62508392$luego podemos decir que hay una Correlación Positiva considerable entre la la talla de hijos y la talla de padres.

Solución b) Usamos Excel para graficar y hallar la regresión simple de la talla de padres e hijos

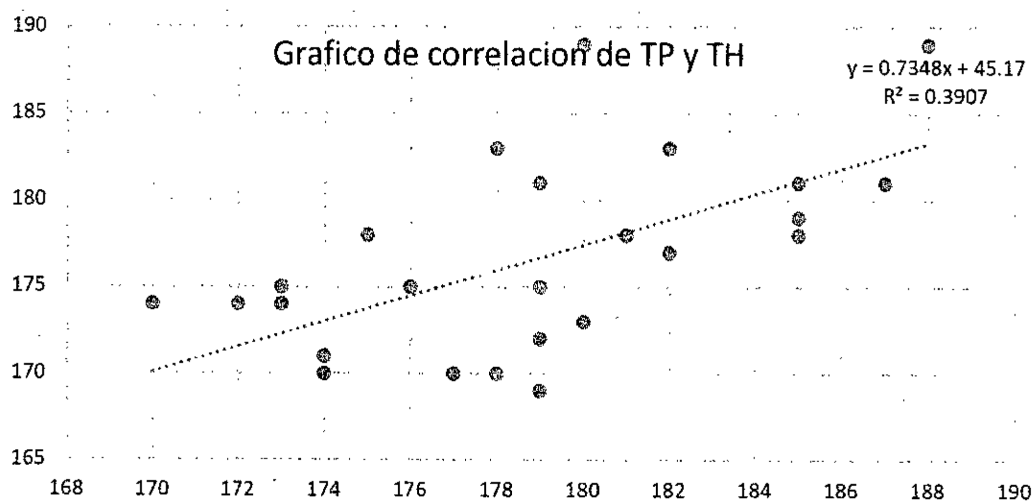


Figura N°27. Gráfica de Regresión Lineal de las Tallas de hijos y padres.

Solución c) La talla del hijo es 176.6992 centímetros.

Niveles de análisis de las relaciones entre las variables

Podemos establecer tres grandes niveles de análisis, por lo que la utilización de la Estadística y el análisis de datos se refiere, o diferentes maneras de aproximación a los datos de una investigación:

Aproximación Univariada: Ordenar, resumir, tabular, graficar, describir, estimar parámetros, evaluar ajustes, contrastar hipótesis univariadas (media proporción).

En este primer nivel el investigador establece como objetivo conocer bien todas las variables objeto de estudio, tomadas de una en una. A partir de aquí resume la información y lo ilustra de manera sencilla y clara. Para obtener dicha información recurre a la estadística descriptiva elaborando tablas, gráficos y los estadísticos que se necesitan (media, varianza, etc.). Con este análisis el investigador puede observar el comportamiento de las variables analizadas en la población de estudio.

Aproximación Bivariada: Contrastar diferencias entre grupos, analizar relaciones entre variables.

Aquí el investigador se sitúa en el nivel bivariado, bajo el cual trata de obtener evidencias acerca de la relación de influencia que hay en una determinada variable de estudio con alguna otra variable de interés. Construye así hipótesis bivariadas que debe contrastar con datos empíricos, haciendo uso para ello de la estadística Inferencial.

Aproximación Multivariada: Elaborar modelos y contrastarlos empíricamente, realizar pronósticos, identificar grupos homogéneos, factores o dimensiones subyacentes.

Si nuestra investigación guarda una relación significativa con diferentes variables debemos plantear la elaboración de modelos multivariados con la intensión de explicar el comportamiento y poder llegar a predecirlo. Tratará en un primer momento de identificar cual es el modelo que mejor se ajusta a los datos empíricos de que dispone (la realidad observada) y tratara de inferir en qué medida podría ajustarse también a futuras observaciones.

Ejemplo: Correlación de las variables aborto eugenésico y el derecho a la vida, resuelto con ayuda del SPSS Para los siguientes datos supuestos recogidos mediante un cuestionario aplicado a 20 madres jóvenes de entre 20 años y 30 años.

Cuyos puntajes arrojados son:

Tabla N°82: *Puntajes de aborto eugenésico y el derecho a la vida*

Aborto Eugenesico	Derecho a la Vida
12,00	11,00
9,00	10,00
11,00	17,00
11,00	14,00
10,00	15,00
12,00	14,00
10,00	14,00
12,00	15,00
12,00	13,00
11,00	14,00
12,00	14,00
12,00	14,00
13,00	15,00
12,00	15,00
13,00	16,00
11,00	14,00
10,00	14,00
10,00	14,00
14,00	11,00
13,00	11,00

Prueba

Para correlacionar las variables Aborto Eugenésico y la variable derecho a la vida usaremos la regresión lineal usando el SPSSv.24: Una vez analizado cada una de las variables y su comportamiento, reemplazamos las constantes en la ecuación de regresión: cuyas variables independientes son

- x : Aborto Eugenésico (AE)
- y : Derecho a la Vida (DV)

Se presentará la siguiente ecuación a resolver:

$$Y = a + bx$$

Y la variable dependiente es: Y : Derecho a la Vida

Aquí presentamos un resumen del modelo

Tabla N°83: *Correlaciones entre aborto eugenésico y el derecho a la vida*

Correlaciones

		Aborto al Eugenésico	Derecho a la Vida
Aborto Eugenésico	Correlación de Pearson	1	**-0,035**
	Sig. (bilateral)		,884
	N	20	20
Derecho a la Vida	Correlación de Pearson	**-0,035**	1
	Sig. (bilateral)	0,884	
	N	20	20

Tabla N°84: *Resumen del modelo aborto eugenésico y el derecho a la vida*

Resumen del modelo[b]

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	0,035[a]	0,001	-0,054	1,82083

a. Variables predictoras: (Constante), Aborto a Eugenésico

b. Variable dependiente: Derecho a la Vida

Tabla N°85: *Anova del modelo aborto eugenésico y el derecho a la vida*

ANOVA[a]

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	0,073	1	,073	,022	,884[b]
	Residual	59,677	18	3,315		
	Total	59,750	19			

a. Variable dependiente: Derecho a la Vida

b. Variables predictoras: (Constante), Aborto Eugenésico

Tabla N°86: *Coeficientes de la recta de regresión*

Coeficientes[a]

Modelo		Coeficientes no estandarizados		Coeficientes tipificados	t	Sig.
		B	Error típ.	Beta		
1	(Constante)	14,306	3,783		3,782	0,001
	Aborto Eugenésico	-0,048	0,327	-0,035	-0,148	0,884

a. Variable dependiente: Derecho a la Vida

Por lo tanto, podemos construir la ecuación de regresión que buscamos:

$$Y = 14.306 - 0.048\,x$$

Podemos ver que el p-valor arrojado es 0.0010 < 0.05 lo que no hay evidencias suficientes para rechazar la hipótesis nula por lo que: "La expectativa del derecho a la vida es mayor que el aborto Eugenésico" .

El coeficiente de determinación múltiple (r^2)

Utilizaremos para determinar la tasa porcentual de Y para ser explicados las variables múltiples, utilizando la si siguiente formula:

$$r^2 = \frac{SCregresiòn}{SCTotal} \qquad r^2 = \frac{0.073}{59.750} = 0.001221757322 \approx 0.001$$

Estadísticas de la Regresión:

Coeficientes de la correlación simple: - 0.035, es una correlación Pearson negativa muy baja entre las variables: El Aborto Eugenésico y el derecho a la Vida

Coeficiente de determinación R^2: 0.001225, R^2 ajustado: 0.001

Error Típico: 1.82083

Como podemos ver en las conclusiones del estudio estadístico de las variables El Aborto Eugenésico y el derecho a la Vida, el aborto Eugenésico influye negativamente con respecto al derecho a la Vida y además como podemos ver en la ecuación de regresión: $y = 14.306 - 0.048\,x$

x : Aborto Eugenésico (AE) , y : Derecho a la Vida (DV)

El valor de "x" tendría que ser muy alto, para que el derecho a la vida sea vulnerado, esta es la razón por la cual rechazamos la hipótesis nula y aceptamos la hipótesis alterna: "La expectativa del derecho a la vida es mayor que el aborto Eugenésico". Además, el coeficiente de determinación R^2: 0.001225, luego el 0.12% determina como influye una variable en la otra.

Gráfico de Correlación de las Variables: Aborto Eugenésico y el Derecho a la Vida

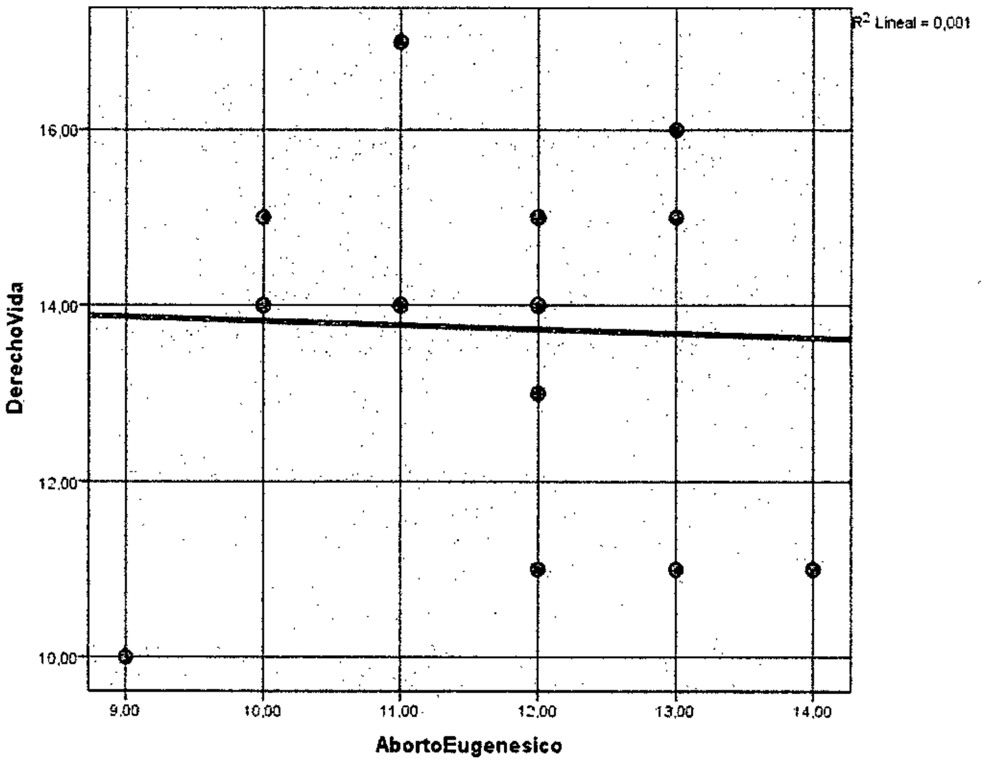


Figura N°28. Regresión Lineal Aborto Eugenésico y el Derecho a la Vida

$$Y = 14.306 - 0.048\,x$$

Podemos observar que la pendiente es negativa por lo que se confirma lo que se ha analizado el SPSS.

Por lo que podemos concluir que: "La expectativa del derecho a la vida es mayor que el aborto Eugenésico".

Ejercicios de Aplicación

1. La siguiente tabla muestra un registro de las cotizaciones alzadas por las acciones de una empresa durante un mes, en relación a la cantidad de días consecutivos que se mantuvieron en alza

Número de días (x)	3	4	5	8	10
Cotización (y)	91	93	97	107	109

a) *Usando el método de mínimos cuadrados halle la ecuación de regresión lineal*

b) Realizar una predicción (función Pronostico) sobre la cotización cundo el número de días es x = 7.

2. Se realizó un estudio de tallas de hijos y padres de 30 alumnos del 2do. Ciclo 2018-B de Ing. Industrial de la UNAC y estos fueron los resultados:

TP	170	171	173	170	170	173	172	171	172	173
TH	175	175	177	175	173	178	171	174	170	179

T P	173	175	170	171	174	172	181	182	175	180
TH	175	174	173	178	168	175	179	187	177	184

T P	182	180	179	178	175	179	184	187	187	177
TH	183	181	180	179	172	173	179	178	188	178

d) ¿Qué correlación hay entre la talla de padres e hijos? **de su interpretación.**

e) Halle la ecuación de regresión simple entre las: TP y TH

f) ¿Qué talla se espera de un hijo cuyo padre mide 179 cm. de estatura?. **A que conclusión puede llegar Usted. Quienes tienen mejores tallas.**

g) Calcule la media aritmética, mediana, primer cuartil, tercer cuartil y la varianza, desviación estándar para las tallas de padres e hijos.

h) Construya un diagrama de caja para cada talla y compare cual tiene mayor variabilidad y como es la asimetría.

Nota: TP; Talla de padres.
TH: Talla de hijos

4. El consumo de cosméticos en una gran ciudad expresado por las ventas, es evaluado en función de la cantidad de distribuidores de los mismos. La evolución en el tiempo se refleja en el siguiente cuadro elaborado por el gerente de ventas que es un Psicólogo.

a) Grafique los valores (elabore la gráfica dispersión de puntos) y ajustar dichos valores empleando los modelos (Lineal, Cuadrático y logarítmico). Mostrar los Modelos en graficas individuales.
b) Indicar cuál es el modelo más adecuado y halle la ecuación que lo representa para el análisis de regresión. Fundamente su respuesta.
c) Estimar cuánto será la venta para 25 distribuidoras.

Tabla N°87: *Consumo de cosméticos en función de la cantidad de distribuidores*

Cantidad de Distribuidoras (x)	Ventas de Cosméticos
1	5
3	205
7	1459
11	6548
15	19568
18	32487
20	95478

2. La fuerza en kilogramos y la longitud alcanzada por una fibra plástica en centímetros están relacionadas. Se realiza una prueba aplicando distintas fuerzas y calculando el estiramiento de la fibra y se observa que es exponencial según muestra la siguiente tabla

   Halle: a) Grafique los valores en un diagrama de dos dimensiones

   b) Ajustar dichos valores empleando el modelo exponencial
   c) Cuál es la ecuación del modelo exponencial y = f (x)
   d) Calcular el valor de la correlación de Pearson e indique que tipo de correlación existe.
   e) Calcular el coeficiente de determinación e indique que representa.
   f) Según el modelo exponencial cuanto será la venta de 25 distribuidoras.

Tabla N°88: *Tabla de la fuerza y la longitud alcanzada por una fibra plástica*

Fuerza (Kg)	Longitud
1	5
3	36
5	127
8	441
11	636
15	2571
17	12440

"Para que una persona descubra su visión, primero la persona tiene que descubrir su don".

La matemática es el alfabeto con que Dios escribió el mundo ¡Buen Trabajo!

CAPITULO VI:

Análisis Combinatorio

6.1. Permutaciones

Se entiende por *análisis combinatorio*, el estudio de los subconjuntos de p elementos tomados sin repetición de un conjunto de n elementos.

El análisis combinatorio es esencial para el cálculo de probabilidades en experimentos complejos, y su conocimiento es necesario y fundamental para todo estudiante de educación superior.

El símbolo $n!$ (léase ***n factorial***) siendo n un número entero positivo, indica el producto de los enteros positivos sucesivos de 1 a n.

$n! = n(n-1)(n-2)(n-3)\cdots\times 5\times 4\times 3\times 2\times 1.$, ó

$n! = 1\times 2\times 3\times 4\times \ldots \times n,\ n\geq 1,\ n\in \mathbb{N},$

donde si $n = 0$: $0! = 1$, por definición

Ejemplos: $5! = 1\times 2\times 3\times 4\times 5 = 120$

$6! = 1\times 2\times 3\times 4\times 5\times 6 = 720$

Propiedades de los factoriales

i.- Si $m! = n!$ Entonces $m = n$, $\forall\, m, n \in \mathbb{N}$,.

ii.- $m! = m(m-1)(m-2)! = \ldots$, $\forall\, n > 1$

iii.- $(m\pm n)! \neq m! \pm n!$

iv.- $(m\, n)! \neq m!\, n!$

v.- $\left(\dfrac{m}{n}\right)! \neq \dfrac{m\,!}{n\,!}$

Ejemplo. Simplificar: 1. $\dfrac{(x+2)x!-(x+1)!}{2(x+1)!}$; $x \in \mathbb{N}$.

Solución

$$\frac{(x+2)x!-(x+1)!}{2(x+1)!}=\frac{x\,x!+2x!-(x+1)x!}{2(x+1)x!}=\frac{x\,x!+2x!-x\,x!-x!}{2(x+1)x!}$$

$$=\frac{x!}{2(x+1)x!}=\frac{1}{2(x+1)}$$

2. $\dfrac{2(n-2)!+(n-1)!}{(n+1)!}$; $n \geq 2$. $n \in \mathbb{N}$.

Solución

$$\frac{2(n-2)!+(n-1)!}{(n+1)!}=\frac{(n-2)!+(n-1)(n-2)!}{(n+1)(n)(n-1)(n-2)!}$$

$$=\frac{n+1}{(n+1)(n)(n-1)}=\frac{1}{n(n-1)}$$

Los arreglos (sin repetición) de orden n formados a partir de los ***n*** elementos de un conjunto E; es decir las diversas formas de ordenar los ***n*** elementos del conjunto E, se llaman ***permutaciones.*** *Notación:* $\boxed{P_n = n!}$

6.1. Combinaciones

Un conjunto E de n elementos distintos, se llama ***combinación*** (sin repetición) de estos ***n*** elementos tomados de ***p*** a la vez (con $p \leq n$) a todo subconjunto que se puede construir seleccionando ***p*** elementos entre los ***n***, sin que se considere el orden de colocación de los elementos.

Es muy importante notar la diferencia que existe entre arreglos y combinaciones. En un arreglo, el orden de colocación de los elementos es determinante; en una combinación, el orden de colocación no tiene ninguna importancia.

El número de combinaciones de n elementos tomados de p en p se indica por:

$$C_{n,p} = \binom{n}{p} = \frac{n!}{p!\,(n-p)!}.$$

NÚMEROS COMBINATORIOS

Se denomina así, al número total de los diferentes grupos que se pueden formar con "m" elementos tomando todos a la vez o de k en k , de modo que los grupos se diferencien por lo menos en un elemento (aquí el orden no es necesario).

Lectura: Combinaciones de m elementos tomados de k en k. Para el cual usaremos la siguiente fórmula.

$$C_k^m = \binom{m}{k} = \frac{m\ (m-1)\ (m-2)\ \dots\ (m-k+1)}{k!}\ ,\ 0 \le k \le m,$$

Nota.- $\binom{m}{k}$ también se le denomina coeficiente binomial y se lee número combinatorio de m elementos tomados en grupos de k, en k (m , k $\in$ N) .

Teorema 1: Si m es un entero positivo y k es un entero no negativo, y $k \le m$,

entonces: $\binom{m}{k} = \frac{m\,!}{k\,!\,(m-k)!}$, m, $k \le$ N.

Demostración.

Usando la diferencia de número combinatorio. Si $0 \le k \le m$, m , k $\in$ N.

$$\binom{m}{k} = \frac{m\ (m-1)\ (m-2)\ \dots\ (m-k+1)}{k!}$$

$$= \frac{m\ (m-1)\ (m-2)\ \dots\ (m-k+1)}{k!} \cdot \frac{(m-k)!}{(m-k)!} = \frac{m\,!}{k\,!\,(m-k)!}$$

Ejercicios de Aplicación.

1. De un grupo de 15 estudiantes mujeres y 12 varones se quiere elegir dos delegados para el curso de matemática de modo que sean un hombre y una mujer. ¿De cuántas maneras se puede hacer ésta elección?

Solución

Se tiene: estudiantes mujeres: $C_1^{15} = \frac{15!}{14!\,1!} = 15$ Estudiantes varones: $C_1^{12} = \frac{12!}{11!\,1!} = 12$

Total: $C_1^{15} \times C_1^{12} = (15)(12) = 180$. Luego existen 180 formas de elegir a los delegados.

2. De un grupo de 5 matemáticos, 3 físicos y 4 ingenieros se debe elegir un comité de 6 personas, de modo que se incluyan 3 matemáticos, 1 físicos y 2 ingenieros. ¿De cuántas formas se puede hacer ésta elección?

Solución

Se tiene: Matemáticos: $C_3^5 = \frac{5!}{2!\,3!} = \frac{4 \cdot 5}{2} = 10$

Físicos: $C_1^3 = \frac{3!}{2!\,1!} = 3$ y el número de Ingenieros: $C_2^4 = \frac{4!}{2!\,2!} = 6$

Total: $C_3^5 \times C_1^3 \times C_2^4 = (10)(3)(6) = 180$

Es decir existen 180 formas de elegir este comité.

Teorema 2: Suma de combinaciones.

$$C_k^m \; + \; C_{k+1}^m \; = \; C_{k+1}^{m+1}$$

Demostración. Usando el teorema 1.

$$C_k^m + C_{k+1}^m = \frac{m!}{k!(m-k)!} + \frac{m!}{(k+1)!\,\left[m-(k+1)\right]!}$$

$$= \frac{m!}{k!(m-k)!} \cdot \frac{(k+1)}{(k+1)} + \frac{m!}{(k+1)!\,\left[m-k-1\right]!} \cdot \frac{(m-k)}{m-k}$$

$$= \frac{(k+1)\ m!}{(k+1)\ k!\ (m-k)!} + \frac{(m-k)\ m!}{(k+1)!\ (m-k)!}$$

$$= \frac{(m-k+1+k)\ m!}{(k+1)!\ (m-k)!} = \frac{(m+1)\ m!}{(k+1)!\ (m-k)!}$$

$$= \frac{(m+1)!}{(k+1)!\ \left[(m+1)-(k+1)\right]!} = C_{k+1}^{m+1}$$

Propiedades de las combinaciones

i) $C_0^m = 1$

ii) $C_m^m = 1$

iii) $C_1^m = m = C_{m-1}^m$

iv) $C_k^m = C_{m-k}^m$

v) $C_k^m = C_p^m \Leftrightarrow \mathrm{k} = \mathrm{p} \vee \mathrm{k} + \mathrm{p} = \mathrm{m}$

Degradación de índices; consiste en descomponer un número combinatorio en otro equivalente. a) ambos índices $C_k^m = \frac{m}{k}\ C_{k-1}^{m-1}$

b) sólo índice superior $C_k^m = \frac{m}{m-k}\ C_k^{m-1}$

c) sólo índice inferior $C_k^m = \frac{m-k+1}{k}\ C_{k-1}^m$

6.2. Variaciones

Variaciones son las ordenaciones diferentes (es decir, importa el lugar que un elemento ocupe) que se pueden lograr tomando para ello, grupos, de determinado tamaño de un total dado de elementos.

El número de ***Variaciones*** de n elementos tomados de ***p*** en ***p*** se indica por:

$$V_{n,p} = \frac{n!}{(n-p)!}$$

Cuadro de características de las **Permutaciones, las Combinaciones y *Variaciones***

Tabla N°89: Características de Permutaciones, Combinaciones y *Variaciones*

	Orden de los elementos		Cantidad de los elementos	
	importa	No importa	Entran todos	Entran algunos
Permutaciones	X		X	
Variaciones	X			X
Combinaciones		X	X	X

Ejercicios con soluciones y explicaciones

1. En una carrera de 15 caballos se supone que se apuesta sobre una terna de caballos. Escogiendo una sola terna:
   - i) ¿Cuál es la probabilidad de que esta terna llegue primera sin consideración del orden de llegada entre los tres caballos?
   - ii) ¿Cuál es la probabilidad de que esta terna llegue primera en el orden indicado?
   - iii) ¿ Cuál es la probabilidad de que esta terna llegue primera en otro orden que el indicado?
   - iv) Escogiendo cinco caballos, ¿cuál es la probabilidad de que estos cinco caballos, tres lleguen primero?

Solución

Si n es probabilidad de que un suceso ocurra. La probabilidad de que A ocurra es $\Pr(A) = \frac{|A|}{n}$,

i) La probabilidad de escoger la terna que ganará sin consideración del orden de llegada es: $\frac{1}{C_{15,3}} = \frac{(15-3)!\ 3!}{15!} = \frac{1}{455}$

ii) La probabilidad de escoger la terna que ganará en su orden de llegada es:

$$\frac{1}{V_{15,3}} = \frac{(15-3)!}{15!} = \frac{1}{2730}$$

iii) La probabilidad de que esta terna llegue primera en otro orden que el de llegada es: $\frac{1}{C_{15,3}} - \frac{1}{V_{15,3}} = \frac{1}{455} - \frac{1}{2730} = \frac{1}{546}$

iv) La probabilidad de escoger la terna que ganará sin consideración de su orden de llegada entre los cinco caballos es:

$$\frac{C_{5,3}}{C_{15,5}} = \frac{10}{3003}$$

2. En un salón de clases hay 9 niños y 3 niñas. ¿De cuántas maneras puede escoger un comité de 4 alumnos?

Solución

Este problema se trata de combinaciones: $C_4^{12} = \frac{12!}{8!\ 4!} = 495$

3. ¿Cuántos objetos distintos deben existir para que el número de combinaciones que se puedan formar, tomándolos 2 a 2, sea igual a 6 veces el número de objetos?

Solución

De acuerdo con el enunciado se tiene:
$$C_2^x = 6x \quad \Rightarrow \frac{x!}{(x-2)!\,2!} = 6x \quad \Rightarrow x = 13$$

4. El número de variaciones de "x" objetos tomados 6 a 6 es 720 veces el número de combinaciones de esos mismos objetos tomados de 4 en 4. Hallar "x"

Solución

De acuerdo con el enunciado se tiene: $V_6^x = 720\ C_4^x$, que desarrollando se tiene:

$$(x)(x-1)(x-2)(x-3)(x-4)(x-5) = \frac{720(x)(x-1)(x-2)}{4\times 3\times 2\times 1}$$

Luego se tiene que: x = 10

5. En un taller de acabado de la empresa " Tornos de Alta Calidad S.A.", es necesario despedir a 4 personas por razones de reestructuración del taller; 8 personas trabajan en el taller, y entre ellas el jefe del taller dio siempre los mejores resultados a la compañía y se excluye la posibilidad de despedirlo. ¿De cuántas maneras se puede despedir a estas 4 personas del taller?

Solución

Se pueden despedir a los empleados de $C_4^7 = 35$ maneras distintas.

6. En Matemática Discreta se tiene que elegir un delegado y un subdelegado. Hay 7 candidatos. ¿Cuántas **combinaciones** se pueden hacer con los candidatos para realizar la selección?

a) 21
b) 49
c) 42

Solución: V(7,2)=7·6=42

Hay otra posible interpretación se deriva del significado matemático de combinaciones

7. A un restaurante un grupo de tres chicos y dos chicas son colocados al azar en una mesa circular. Si a es el número de colocaciones diferentes en las que se sientan dos chicas juntas y b es el número de colocaciones diferentes en las que no se sientan dos chicas juntas (dos colocaciones serán iguales si una puede ser obtenida de la otra mediante una rotación apropiada). Entonces:
   a) a=12 y b=12
   b) a=14 y b=12
   c) a=15 y b=10

   Solución: Al ser circular, fijamos uno como referencia, supongamos un chico: O_1*, los otros chicos los llamamos:* O_2*,* O_3*. Las chicas:* A_1 *y* A_2

   Colocaciones con chicas juntas:

   $O_1AAOO \rightarrow 2!\cdot 2!=4$

   $O_1OAAO \rightarrow 2!\cdot 2!=4$

   $O_1OOAA \rightarrow 2!\cdot 2!=4$

   Total: 12

   Colocaciones con chicas separadas:

   $O_1AOAO \rightarrow 2!\cdot 2!=4$

   $O_1AOOA \rightarrow 2!\cdot 2!=4$

   $O_1OAOA \rightarrow 2!\cdot 2!=4$

   Total: 12

8. ¿Cuál es el número de colocaciones diferentes de 7 libros en mi estante de modo que tres libros de Cálculo estén siempre separados entre sí?

a) 1520

b) 1634

c) 1440

Solución: □□□□□□□

Hay 10 formas de escoger 3 casillas separadas

Hay 3! maneras de permutar 3 elementos

Hay 4! maneras de permutar 4 elementos

En total: 10·3!·4!=1440

Otra forma de enfocarlo:

Hay un total de 7! maneras de colocar los 7 libros.

Hay 3!·5·5·4! maneras de colocar 2 libros juntos.

Total: 7! – 3!·5·5·4!

9. ¿Cuántos números de cinco cifras se pueden escribir con cuatro dos y cuatro cincos?

a) 30

b) 50

c) 36

Solución: un número de cinco cifras se puede obtener:

4 dos y 1 cinco$\rightarrow$*22225*$\rightarrow P^5_{4,1} = \dfrac{5!}{4!\cdot 1!} = 5$

3 dos y 2 cincos$\rightarrow$*22255*$\rightarrow P^5_{3,2} = \dfrac{5!}{3!\cdot 2!} = 10$

2 dos y 3 cincos $\rightarrow$*22555*$\rightarrow P^5_{2,3} = \dfrac{5!}{2!\cdot 3!} = 10$

1 dos y 4 cincos→25555→ $P_{1,4}^{5}=\frac{5!}{1!\cdot 4!}=5$

Total de números: 5+10+10+5=30

10. ¿Cuál es el tamaño mínimo de una población estudiantil para que exista al menos un día al año (de 365 días) donde coincidan la fecha del aniversario de nacimiento de al menos nueve estudiantes?
    a) 2921
    b) 2633
    c) 3025

    Solución: colocando 8 estudiantes por día, de forma que su aniversario sea ese día, tenemos un total: 8·365=2920

    Si añadimos un estudiante más, se colocará en uno de los 365 días, día que pasará a tener 9 estudiantes.

    La respuesta es 2921

11. ¿Cuál es el número de soluciones enteras no negativas de la ecuación: $x_1+x_2+x_3+x_4+x_5=30$?
    a) 60211
    b) 46376
    c) 48520

    Solución: el problema es similar a las permutaciones con repetición de treinta 1 y cuatro separadores:

$$P_{30,4}^{34}=\frac{34!}{30!\cdot 4!}=\frac{34\cdot 33\cdot 32\cdot 31}{4\cdot 3\cdot 2\cdot 1}=46376$$

12. En una carrera de largo aliento intervienen 4 españoles, 4 italianos, 4 ingleses y 4 franceses. Supuesto que terminan la carrera todos los corredores, cuántos podios distintos pueden darse al acabar la carrera en los cuales no hay españoles.
    a) 1348

b) 1320
c) 1570

Solución: El oro, la plata y el bronce lo obtienen tres personas distintas. Si no pueden ser españoles, hay 12 personas no españolas.

El oro lo pueden obtener 12 personas

La plata 11 personas

El bronce 10 personas

Total: $12 \cdot 11 \cdot 10 = 1320$

13. ¿Cuántas permutaciones del conjunto de números 1, 2, 3, 4, 5 y 6, satisfacen la condición: el 1 está en primera posición y el 4 en la tercera?

a) 23
b) 24
c) 26

Solución: Colocando fijos el 1 en la primera y el 4 en la tercera, los cuatro números restantes: 2,3,5,6 se pueden colocar de 4! formas distintas (permutaciones).

Total: $4! = 24$

15. cuántas formas 5 hombres y 3 mujeres se pueden sentar alrededor de una mesa redonda de modo que dos mujeres no se encuentren juntas. (Dos formas son iguales si se llega de una a otra por rotación. No importa únicamente el sexo sino también que persona es:

a) 1440
b) 6520
c) 1100

Solución: dado que es son permutaciones circulares, fijamos un hombre como referencia relativa.

Hay 10 maneras de escoger los tres sitios para las mujeres de forma que no se sienten juntas.

Hay 3! formas distintas de colocar las tres mujeres en tres sitios.

Hay 4! formas distintas de colocar los cuatro hombres en los sitios restantes.

Total: $10 \cdot 3! \cdot 4! = 1440$

16. Un candidato a una Magistratura ha estudiado 120 horas a lo largo de 14 días (se supone que cada día lo ha hecho un número entero de horas). Entonces hubo necesariamente un par de días consecutivos en los que estudió al menos

a) 19 horas en total

b) 18 horas en total

c) 20 horas en total

Solución: repartiendo 119 horas entre 14 días, puede quedar por día:

98989898989898 ó 98989898989889

*(obsérvese que **no hay** una pareja consecutiva con más de 17 horas, aunque todas las parejas tienen 17 horas salvo una que tiene 16).*

Si añadimos 1 hora más para obtener los 120, habrá necesariamente una pareja consecutiva con 18 horas.

17. Con las cifras 0,1,2,3,4,5,6,7,8 se forman números de cinco cifras, ¿Cuántos números diferentes pueden formarse sin repetir cifras?

a) 15120

b) 13144

c) 12882

Solución: entendiendo que "01234" es un número de cinco cifras, lo que nos piden serán variaciones sin repetición de 9 elementos tomados de 5 en 5.

$V(9,5)=9 \cdot 8 \cdot 7 \cdot 6 \cdot 5=15120$

18. En una pollería hay 4 tipos de bocadillos para comer. ¿De cuántas maneras distintas se pueden elegir seis bocadillos de entre los 4 tipos?

a) 81

b) 87

c) 84

Solución: el problema es similar a repartir 6 bolas idénticas en cuatro casillas, donde cada casilla representa un tipo de bocadillo. También es similar a las distintas permutaciones que se pueden realizar con: 1/11/11/1, donde hay 6 unos y 3 separadores.

El nº de unos hasta el primer separador indica en número de bocadillos escogidos del primer tipo.

El nº de unos entre el primero y segundo separador nos indica el número de bocadillos escogidos del segundo tipo.

Total: $P^{9}_{6,3} = \dfrac{9!}{6!\,3!} = \dfrac{9{\cdot}8{\cdot}7}{3{\cdot}2{\cdot}1} = 3{\cdot}4{\cdot}7 = 84$

19. ¿Cuántas sucesiones de n dígitos se pueden formar con los elementos {0,1,2}, que posean al menos un '0', un '1' y un '2'?

a) 3^n

b) $3^n - 3\cdot 2^n + 3$

c) $3^n - 2^n + 1$

Solución:

Total de sucesiones de n dígitos son: 3^n

Total de sucesiones que no poseen "0": 2^n

Total de sucesiones que no poseen "1": 2^n

Total de sucesiones que no poseen "2": 2^n

Total de sucesiones sin "0" ni "1": 1

Total de sucesiones sin "0" ni "2": 1

Total de sucesiones sin "1" ni "2": 1

Resumiendo: $3^n - 3\cdot 2^n + 1 + 1 + 1$

20. Sea E un alfabeto con 5 vocales y 21 consonantes. ¿Cuántas palabras de 5 letras pueden formarse con las letras de E, tales que la primera y la última letras sean vocales distintas y las otras tres sean consonantes distintas?

a) $26!/(3!\cdot 2!)$

b) $2^5\cdot 3^{21}$

c) $V(5,2)\cdot V(21,3)$

Solución: formando series $V_1V_2C_1C_2C3$ (donde V=vocal, C=consonante)

Para V_1V_2 tenemos: $V(5,2)=5\cdot 4$ posibilidades

Para $C_1C_2C_3$ tenemos: $V(21,3)=21\cdot 20\cdot 19$

Total=$V(5,2)\cdot V(21,3)=5\cdot 4\cdot 21\cdot 20\cdot 19$

21. Con los dígitos 1,2,3,4,5 se forman números de tres cifras. ¿Cuántos números diferentes pueden formarse sin repetir cifras que sean múltiplos de 3?

a) 60

b) 24

c) 20

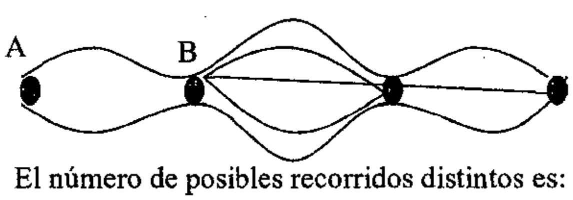
Solución: escogemos primeramente los subconjuntos de tres elementos que dan lugar a números múltiplos de 3:

123, 135, 234, 345 →4 subconjuntos

Ahora obtenemos todas las permutaciones de estos tres elementos → 3!=6 por cada subconjunto

Total=$4\cdot 3!=24$

22. Para ir de la ciudad A a la ciudad D hay que pasar por las ciudades B y C a través de las carreteras que se indican en la figura



El número de posibles recorridos distintos es:

a) 10
b) C(10,2)·C(10,5)·C(10,3)
c) 30

Solución: aplicando el principio multiplicativo

Para ir de A a B hay: 2 posibilidades

Para ir de B a C hay: 5 posibilidades

Para ir de C a D hay: 3 posibilidades

Total=2·5·3=30

23. ¿Cuántas permutaciones del conjunto de números {1,2,3,4,6,9} satisfacen la condición de que en la primera posición y en la última haya un múltiplo de 3?

a) 360
b) 24
c) 144

Solución: cifras múltiplos de 3 son: 3,6,9

En la primera y en la última deben estar ocupadas por dos de estas cifras, lo que tenemos: V(3,2)=3·2=6 posibilidades

Las otras cuatro posiciones pueden ser ocupadas por las cifras restantes de $V(4,4)=P_4=4\cdot 3\cdot 2\cdot 1=24$

Total=6·24=144

24. En una carrera de 100 metros planos intervienen 4 corredores por cada uno de los 4 equipos. Supuesto que terminan la carrera todos los corredores, ¿cuántos resultados distintos pueden darse al acabar la carrera en los cuales no hay ningún corredor del equipo A entre los tres primeros?

a) 1348
b) 1320
c) 1570

Solución: no pueden quedar en las tres primeras posiciones los 4 corredores del equipo A, pero sí los 12 restantes.

La 1ª posición puede ser ocupada por 12 corredores.

Por cada ocupación de la primera, la segunda puede ser ocpuada por 11.

Y por cada ocupación de la primera y segunda la tercera puede ser ocupada por 10.

Total=12·11·10=1320

25. ¿Cuántas permutaciones del conjunto de números 1,2,3,4,5 y 6, satisfacen la condición: el 1 está en primera posición y el 4 en la tercera?

a) 23
b) 24
c) 26

Solución: si el 1 ocupa la primera posición y el 4 la tercera, quedan 4 elementos por colocar en las restantes 4 posiciones, lo que hace un total de 4!=24 permutaciones.

26. Se tienen "cadenas" formadas por dos letras seguidas de cuatro dígitos y otras tres letras más. No están permitidas las repeticiones de letras y dígitos dentro de cada grupo, pero el último grupo de tres letras puede contener una o dos de las utilizadas al principio de la cadena. ¿Cuántas cadenas distintas se pueden formar si el número de letras disponibles es 26?

a) 560.000.000
b) 720.100.029
c) 51.105.600.000

Solución: para obtener todas las seires de la forma: $L_1L_2D_1D_2D_3D_4L_3L_4L_5$ *(donde L=letra y D=dígito).*

Para L_1L_2 *tenemos 26·25 posibilidades*

Para $D_1D_2D_3D_4$ *tenemos 10·9·8·7 posibilidades*

Y para $L_3L_4L_5$ tenemos $26\cdot25\cdot24$

$Total=26\cdot25\cdot10\cdot9\cdot8\cdot7\cdot26\cdot25\cdot24=51.105.600.000$

27. ¿De cuántas maneras se pueden ordenar la palabra EXAMENES si no puede haber dos "E" adyacentes?

a) 2100

b) 2400

c) 5400

Solución: hay tres E, que de forma no adyacente se pueden colocar de 20 formas distintas. Las restantes cinco letras se pueden colocar de 5! maneras distintas.

$Total=20\cdot5!=20\cdot120=2400$

28. Un deportista ha entrenado 42 horas a lo largo de 8 días consecutivos (se supone que cada día lo ha hecho un número entero de horas). Entonces hubo necesariamente un par de días consecutivos en los que entrenó, al menos, un total de horas de:

a) 13

b) 12

c) 11

Solución: si repartimos 40 horas en ocho días obtenemos una distribución equitativa:

55555555

Podemos así garantizar que no hay pareja de días con más de 10horas. Si añadimos 2 horas, pueden quedar en la forma:

55655565

Entonces habrá al menos una pareja con 11 días.

CAPITULO VII.

Probabilidades

7.1. Probabilidades: nociones preliminares, reglas.

Las probabilidades se ocupan de los fenómenos y/o experimentos aleatorios o no aleatorios, el cálculo de probabilidades nos suministra las reglas para el estudio de los experimentos aleatorios o de azar, constituyendo la base para la estadística inductiva o inferencial.

Experimentos y sucesos aleatorios

Experimentos: conjunto de acciones reales o imaginarias que se realizan con la finalidad de obtener resultados.

Diremos que *un experimento es aleatorio* si se verifican las siguientes condiciones:

a. Se puede repetir indefinidamente, siempre en las mismas condiciones;
b. Antes de realizarlo, no se puede predecir el resultado que se va a obtener;
c. El resultado del experimento no se conoce, pero puede conocerse los posibles resultados del mismo. El resultado que se obtenga, *e*, pertenece a un conjunto conocido previamente de resultados posibles. A este conjunto, de resultados posibles, lo denominaremos **espacio muestral** y lo denotaremos normalmente mediante la letra *E*. Los elementos del espacio muestral se denominan **sucesos elementales**.

$e_1, e_2 \in E \Rightarrow e_1, e_2$ son sucesos elementales

Cualquier subconjunto de *E* será denominado **suceso aleatorio**, y se denotará normalmente con las letras $A, B, \ldots$

$A, B \subset E \Rightarrow A, B$ son sucesos aleatorios.

En conclusión, en un experimento aleatorio no sabemos que resultado saldrá, pero si qué valores puede tomar.

Las variables aleatorias se clasifican en:

- Discretas: Aquellas que resultan de contar el número de casos en los que el evento de interés ocurre, por ejemplo: número de hijos de una familia, número de veces que llega una paciente al servicio de emergencia, etc.
- Continuas: Aquellas que resultan producto de una medición, por ejemplo: el peso, el nivel de hemoglobina, etc.

Probabilidad: Es un número que esta entre 0 y 1 que permite cuantificar la posibilidad de ocurrencia de un evento.

Sea $\xi \rightarrow \Omega = \{w_1, w_2, \dots\dots\dots, w_n, \dots\}$

y sea **E** $\subset$ **W**, un evento de **W**

Con cada evento **E**, asociamos un número real **P(E)**, llamado ***Probabilidad de E***, que satisface:

i) $0 \leq P(E) \leq 1$

ii) $P(W) = 1$ $\quad\quad$ $P(\Phi) = 0$

iii) Si E y F son eventos mutuamente excluyentes ($E \cap F = \Phi$),

entonces: $P(E \cup F) = P(E) + P(F)$

7.2. Sucesos independientes, probabilidad total y teorema de Bayes.

La distribución de probabilidad de una variable aleatoria es una función que asigna a cada suceso definido sobre la variable aleatoria la probabilidad de que dicho suceso ocurra. La distribución de probabilidad está definida sobre el conjunto de todos los eventos rango de valores de la variable aleatoria. Para el teorema de Bayes tenemos que los sucesos A, A y B, obtén P(A/B) en función de los valores de P(A), P(A), P(B/A) y P(B/A)

$$P(A \wedge B) = \begin{Bmatrix} P(A).P(B/A) \\ P(B).P(A/B) \end{Bmatrix}$$

$$P(A).P(B/A) = P(B).P(A/B)$$

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

$$P(A/B) = \frac{P(A).P(B/A)}{P(A).P(B/A) + P(\bar{A})P(B/\bar{A})}$$

El cálculo de probabilidades es bastante amplio pero nuestro objetivo es llevarlo a la práctica solucionando ejercicios usando ya sea el Excel y/o el SPSS en forma directa, razón por la cual entramos directamente a

7.3. Distribución Binomial

Un experimento aleatorio que consiste de n ensayos repetidos tales que:

a) Los ensayos son independientes
b) Cada ensayo tiene solo dos resultados llamados éxito y fracaso
c) La probabilidad de éxito en cada ensayo, denotada por p, permanece constante, recibe el nombre de experimento binomial.

$$b(x;n,p)=\binom{n}{x}p^{x}(1-p)^{n-x}=\frac{n!}{x!(n-x)!}p^{x}(1-p)^{n-x} \quad ; \; x=1,2,\ldots\ldots,n$$

Donde: x = Es el número de éxitos en el ensayo

n = Es el Número de ensayos independientes

p = Es la Probabilidad de éxito y q = Es la Probabilidad de fracaso

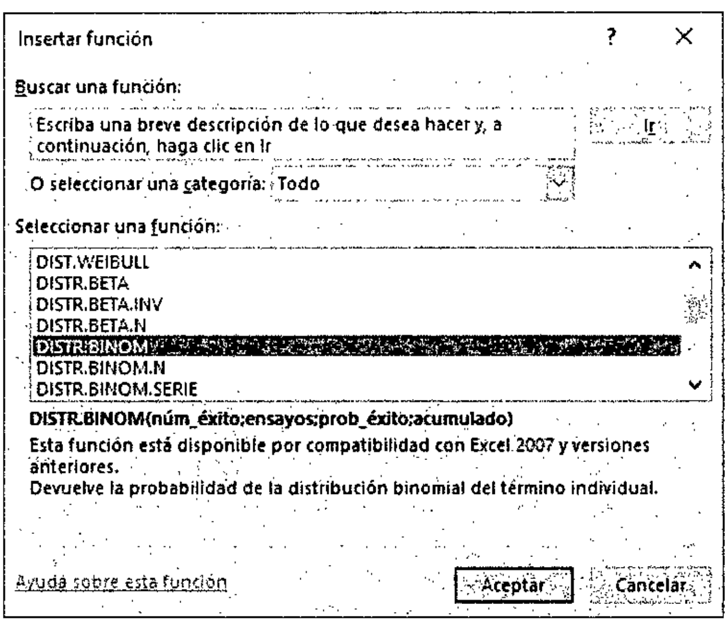


Figura N°29. La Distribución Binomial dentro del Excel

Ejemplo: ¿Cuál es la probabilidad de obtener 6 caras al lanzar una moneda 10 veces? Notación: X ~ B (x; n ; p), Distribución Binomial (DB)

x = número de aciertos, de obtener 6 caras.

n = número de ensayos. Y p = probabilidad de éxito salga cara.

Solución: usamos: $b(x;n,p)=\binom{n}{x}p^{x}q^{n-x}=\dfrac{n!}{x!(n-x)!}p^{x}q^{n-x}$

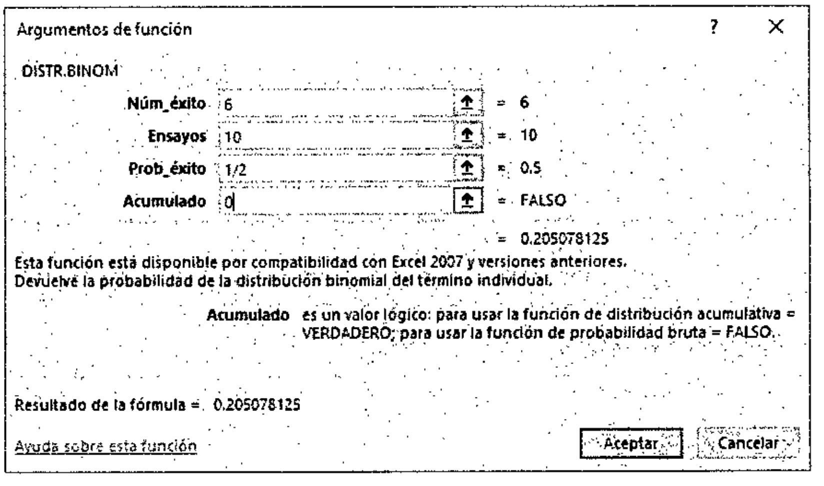


Figura N°30. La DB para obtener 6 caras al lanzar una moneda 10 veces

$$b(x=6;n=10,p=1/2)=\binom{10}{6}0.5^{6}0.5^{10-6}$$

$$\binom{10}{6}0.5^{6}0.5^{10-6}=\frac{10!}{6!(10-6)!}0.5^{6}0.5^{10-6}$$

P (x=6) = 0. 205

2. ¿Cuál es la probabilidad de obtener 4 veces el número 3 al lanzar un dado 8 veces? Notación: X ~ B (x; n ; p)

   x = número de aciertos, de obtener 4 veces el número 3.

   n = número de ensayos.

   P = probabilidad de éxito salga un 3 al lanzar el dado.

Solución

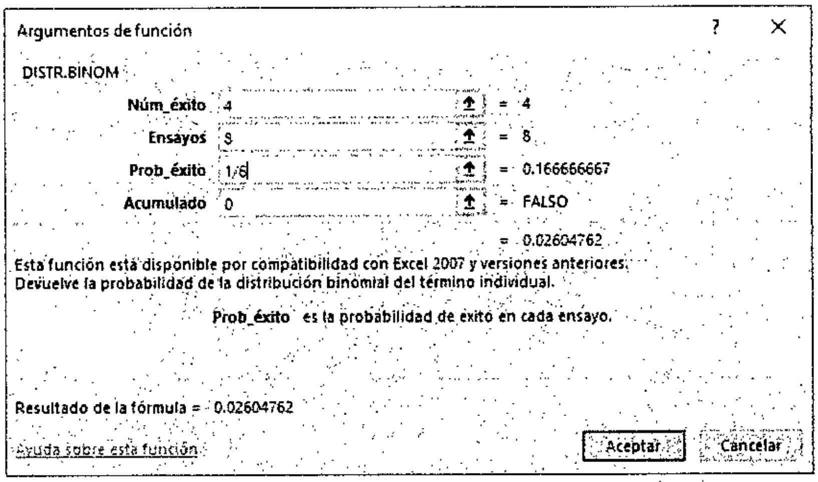


Figura N°31. La DB para obtener 4 veces el número 3 al lanzar un dado 8 veces

$$b(x;n,p)=\binom{n}{x}p^{x}q^{n-x}=\frac{n!}{x!(n-x)!}p^{x}q^{n-x}$$

$$b(x=4;n=8,p=1/6)=\binom{8}{4}0.166^{4}\times 0.8333^{8-4}$$

$$\binom{8}{4}0.1666^{4}\times 0.8333^{8-4}=\frac{8!}{4!(8-4)!}0.1666^{4}\times 0.8333^{8-4}$$

P = 0.026

Ejercicios de Aplicación

1.- Construir el cuadro de la Distribución Binomial al lanzar 15 veces una moneda y determinar lo siguiente:

a) La probabilidad de que salgan 11 caras y cuantos casos cumplen esta condición.

b) La probabilidad de que salgan entre 4 y 6 caras inclusive y cuantos casos cumplen esta condición.

c) La probabilidad de que salgan más de 12 caras y cuantos cumplen esta condición.

Tabla N°90: *Distribución Binomial al lanzar 15 veces una moneda en Excel*

TOTAL DE CARAS (xi)	PROBABILIDAD CARAS p(xi)	N° CARAS REPITEN	P(xi = 11)	P(4 ≤ xi ≤ 6)	P(12 < xi)
15	3.05E-05	1			3.05E-05
14	0.00045776	15			0.00045776
13	0.00320435	105			0.00320435
12	0.0138855	455			
11	0.04165649	1365	0.04165649		
10	0.09164429	3003			
9	0.15274048	5005			
8	0.19638062	6435			
7	0.19638062	6435			
6	0.15274048	5005		0.15274048	
5	0.09164429	3003		0.09164429	
4	0.04165649	1365		0.04165649	
3	0.0138855	455			
2	0.00320435	105			
1	0.00045776	15			
0	3.05E-05	1			
TOTAL	1	32768	0.04165649	0.28604126	0.00369263
CASOS QUE CUMPLEN			1365	9373	121

a) La probabilidad que salgan 11 cara es 0.04165649 y los casos que se cumplen es 1365.

b) La probabilidad de que salgan entre 4 y 6 caras es 0.28604126 y los casos que se cumplen es 9373

c) La probabilidad de que salgan más de 12 caras es 0.00369263 y los casos que se cumplen es 121

2.- La probabilidad de que un paciente se recupere de una rara enfermedad de la sangre es 0.4 Si se sabe que 15 personas han contraído esta enfermedad. ¿Cuál es la probabilidad de que:

a) Al menos 10 sobrevivan,
b) Sobrevivan entre 3 y 8 personas, y
c) Sobrevivan exactamente 5 personas

Tabla N°91: *Distribución de probabilidad Binomial en Excel*

xi	p(xi)	P(10 < xi)	P(3 < xi < 8)	P(xi = 5)
0	0.00047018			
1	0.00470185			
2	0.02194197			
3	0.0633879			
4	0.1267758		0.1267758	
5	0.18593784		0.18593784	0.18593784
6	0.20659761		0.20659761	
7	0.17708366		0.17708366	
8	0.11805577			
9	0.06121411			
10	0.02448564	0.02448564		
11	0.00741989	0.00741989		
12	0.00164886	0.00164886		
13	0.00025367	0.00025367		
14	2.42E-05	2.42E-05		
15	1.07E-06	1.07E-06		
TOTAL	1	0.0338333	0.69639491	0.18593784

a) Que al menos 10 sobrevivan es 0.0338333,
b) Que Sobrevivan entre 3 y 8 personas es 0.69639491
c) Que sobrevivan exactamente 5 personas es 0.18593784

3.- Una máquina produce cierto tipo de piezas, de las cuales el 5% en promedio son defectuosas. En una muestra aleatoria de 5 piezas ¿cuál es la probabilidad de obtener :

a) Exactamente dos piezas defectuosas?

b) Por lo menos una pieza defectuosa?

c) Exactamente tres piezas no defectuosas

d) Por lo menos 2 piezas no defectuosas

Tabla N°92: *DB de la producción de piezas defectuosas y no defectuosas*

			DEFECTUOSOS		NO DEFECTUOSOS	
xi	Defectuosos p(xi)	No Defectuosos q(xi)	P(xi = 2)	P(0 < xi)	P(xi = 3)	P(1 < xi)
0	0.77378094	3.13E-07				
1	0.20362656	2.97E-05		0.20362656		
2	0.02143438	0.00112813	0.02143438	0.02143438		0.00112813
3	0.00112813	0.02143438		0.00112813	0.02143438	0.02143438
4	2.97E-05	0.20362656		2.97E-05		0.20362656
5	3.13E-07	0.77378094		3.13E-07		0.77378094
TOTAL	1	1	0.02143438	0.22621906	0.02143438	0.99997

a) La probabilidad que salgan exactamente dos piezas defectuosas es 0.02143438

b) La probabilidad que salgan por lo menos una pieza defectuosa es 0.22621906

c) La probabilidad que salgan que exactamente tres piezas no defectuosas es 0.02143438

d) La probabilidad que por lo menos 2 piezas no defectuosas0.99997

4.- Sea x una variable aleatoria con distribución binomial, cuya media es 12 y varianza 4.8, Calcular lo siguiente:

a) $P(x = 5)$ b) $P(5 < x < 9)$ c) $P(x < 4)$

Solución:

Sabemos que: "p" es el éxito y que "q" es el fracaso y que : $p + q = 1$

Además que: la media es $\mu = n\,p = 12$

La varianza es $\sigma^2 = n\,p\,q = 4.8$ aquí reemplazamos la media y se tiene: $12\,q = 4.8$ luego $q = 0.4$ luego $p = 0.6$ y $n = 20$

Con estos datos ya podemos usar la distribución Binomial y usando Excel se tiene:

Tabla N°93: *Tabla de Distribución de probabilidad Binomial*

xi	p(xi)	P(xi = 5)	P(5 < xi < 9)	P(xi < 4)
0	1.10E-08			1.10E-08
1	3.30E-07			3.30E-07
2	4.70E-06			4.70E-06
3	4.23E-05			4.23E-05
4	0.00026969			
5	0.00129449	0.0012945		
6	0.00485435		0.00485435	
7	0.01456305		0.01456305	
8	0.03549744		0.03549744	
9	0.07099488			
10	0.11714155			
11	0.15973848			
12	0.17970579			
13	0.16588227			
14	0.1244117			
15	0.07464702			
16	0.03499079			
17	0.01234969			
18	0.00308742			
19	0.00048749			
20	3.66E-05			
TOTAL	1	0.0012945	0.05491484	4.73E-05

Luego: a) $P(x = 5) = 0.0012945$

b) $P(5 < x < 9) = 0.05491484$

c) $P(X < 4) = 0.000047345$

5.- Un examen consta de 20 preguntas; cada una de ellas tiene 5 respuestas posibles de las cuales sólo una es la respuesta correcta. Si un estudiante que desconoce el curso contesta la prueba aleatoriamente calcular:

a) ¿Cuál es la probabilidad de que acierte en más de 10 respuestas correctas?

b) ¿Cuál es el número esperado de respuestas correctas?

Solución: Sabemos que: "p" es el éxito y que "q" es el fracaso y que para nuestro caso : $p = 1/5 = 0.2$ luego $q = 4/5 = 0.8$

Además que: la media es $\mu = n\,p$ o sea $E(xi) = \sum xi\, p(xi) = 4$

Tabla N°94: *Distribución Binomial de un examen de 20 preguntas*

xi	p(xi)	P(10 < xi)	E(xi) = ∑xi p(xi)
0	0.01152922		0
1	0.05764608		0.05764608
2	0.13690943		0.27381886
3	0.20536414		0.61609243
4	0.2181994		0.87279761
5	0.17455952		0.87279761
6	0.1090997		0.65459821
7	0.05454985		0.38184895
8	0.02216088		0.17728701
9	0.00738696		0.06648263
10	0.00203141		0.02031414
11	0.00046168	0.00046168	0.00507853
12	8.66E-05	8.66E-05	0.00103879
13	1.33E-05	1.33E-05	0.00017313
14	1.66E-06	1.66E-06	2.33E-05
15	1.66E-07	1.66E-07	2.50E-06
16	1.30E-08	1.30E-08	2.08E-07
17	7.65E-10	7.65E-10	1.30E-08
18	3.19E-11	3.19E-11	5.74E-10
19	8.39E-13	8.39E-13	1.59E-11
20	1.05E-14	1.05E-14	2.10E-13
TOTAL	1	0.00056341	4

a) La probabilidad de que acierte en más de 10 respuestas correctas es 0. 0005634

b) El número esperado de respuestas correctas es: 4

Ejercicios Propuestos

1.- Se ha elaborado un examen de selección de personal consistente en 30 preguntas, para cada una de ellas hay 7 posibles respuestas, pero se tiene conocimiento que por cada pregunta dos respuestas son correctas, pero se ha establecido que para aprobar esta prueba tiene que contestar correctamente 16 preguntas.

a) Cual es la probabilidad de que el postulante que rinda la prueba salga aprobado.

d) Cual es la probabilidad de que el postulante acierte más de 12 y menos 30.

e) Calcular el promedio y su desviación estándar.

2.- Un empresario de metal mecánica produce candados, de los cuales se tiene conocimiento que el 8.5% son defectuosos. En una selección aleatoria de 8 candados se pide lo siguiente:

a) Cual es la probabilidad de encontrar 6 candados defectuosos.

b) Cual e la probabilidad de encontrar entre 4 y 7 candados defectuosos aceptables.

c) Cual es la probabilidad de encontrar 3 candados aceptables.

d) Cual es su promedio y Desviación estándar de los candados defectuosos.

e) Cual es su promedio y Desviación estándar de los candados aceptables.

3.- Las observaciones durante un largo periodo muestran que un vendedor determinado puede concluir una venta en una sola entrevista con una probabilidad del 20% Supóngase que el vendedor entrevista a cuatro compradores.

a) Cuál es la probabilidad de que exactamente 2 compradores compren el producto.

b) Cuál es la probabilidad de que al menos 2 compradores compren el producto.

7.4. Distribución Hipergeométrica

$$h(x;n,M,N)=\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

Donde:

x = Es el número de éxitos de la muestra

n = Es el Tamaño de la muestra

M = Es la Población éxito

N = Es el tamaño de la población

Ejemplo

1. Una caja contiene 20 CD, de los cuales 8 están en blanco y los 12 restantes contienen información.

Si una persona selecciona 4 CD aleatoriamente. Se pide calcular:

a) Cuál es la probabilidad de que 1 CD este en blanco.
b) Cuál es la probabilidad de que 2 CD estén en blanco.
c) Cuál es la probabilidad de que 1 CD estén con información.

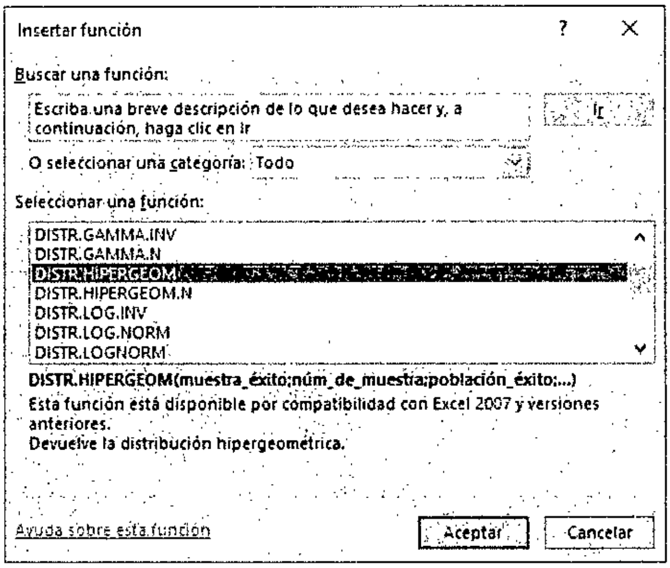


Figura N°32. La Distribución Hipergeométrica dentro del Excel

Solución con calculadora

$$p(x)=h(x=1;n=4,M=8,N=20)=\frac{\frac{8!}{1!7!}\times\frac{12!}{3!9!}}{\frac{20!}{4!16!}}=0.3632611$$

Solución a: com Excel

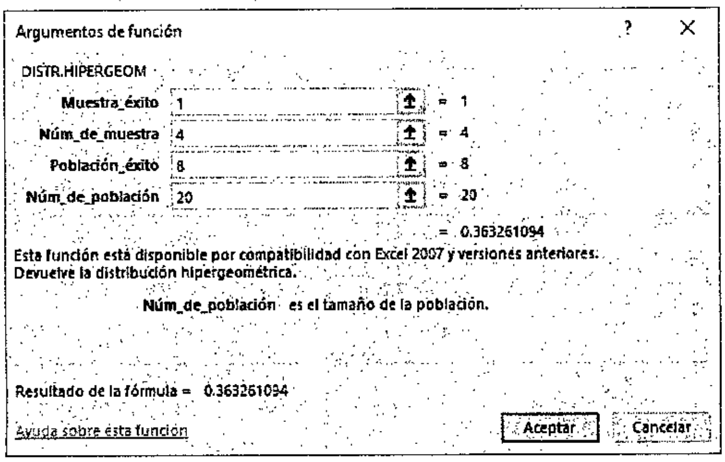


Figura N°33. La Distribución Hipergeométrica solución en Excel

x = 1 , n = 4 , M = 8 , N = 20

Solución b: para x = 2 n = 4 M = 8 N = 20

$$p(x)=h(x=2;n=4,M=8,N=20)=\frac{\frac{8!}{2!6!}\times\frac{12!}{2!10!}}{\frac{20!}{4!16!}}=0.3814241$$

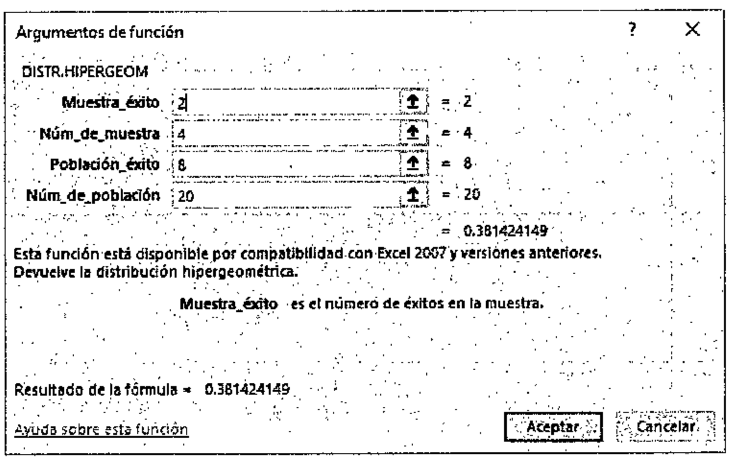


Figura N°34. La Distribución Hipergeométrica solución en Excel para x = 2

Solución c:

x = 1 n=4 M=12 N=2

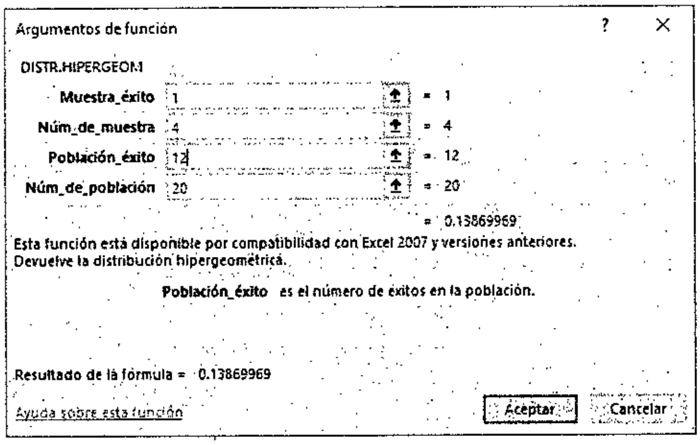


Figura N°35. La Distribución Hipergeométrica solución en Excel para x = 1

La probabilidad es: P = 0.13869696969

2. En una Clínica hay 20 enfermos, de los cuales se sabe que el 30% tiene cáncer, se extrae aleatoriamente 4 pacientes para el despistaje de cáncer.

a) ¿Cuál es la probabilidad de que al menos uno tenga cáncer?

b) ¿Cuál es la probabilidad de que exactamente 3 tengan cáncer?

c) ¿Cuál es la probabilidad de que más de uno tenga cáncer?

d) ¿Cuál es la probabilidad de que ninguno tenga cáncer?

e) ¿Cuál es su promedio y varianza?

Solución a: x =1, 2, 3, 4. n = 4 M = 6 N =20

Solución b: x = 3 n = 4 M = 6 N = 20

Solución c: x = 2, 3, 4. n = 4 M = 6 N = 20

3. Un fabricante de neumáticos para automóviles reporta que de un embarque de 5000 enviados a un gran distribuidor local, 1000 están ligeramente manchados.

Si pequeños comerciantes compran al distribuidor 10 de estos neumáticos aleatoriamente:

a) ¿Cuál es la probabilidad de que exactamente 5 neumáticos estén manchados?

b) ¿Cuál es la probabilidad de encontrar más de 1 neumáticos manchados?

Solución a: x = 5 n = 10 M = 1000 N = 5000

Solución b: x=2, 3, 4, ...10 , n=10 , M=1000 , N=5000

4. Una caja contiene 20 CD, de los cuales 8 están en blanco y los 12 restantes contienen información.

Si una persona selecciona aleatoriamente sin reposición 4 CD. Se pide calcular lo siguiente:

a) ¿Cuál es la probabilidad de que 1 CD este en blanco?
b) ¿Cuál es la probabilidad de que 2 CD estén en blanco?
c) ¿Cuál es la probabilidad de obtener más de 1 CD en blanco?
d) ¿Cuál es la probabilidad de que 3 CD tengan información?
e) ¿Cuál es la probabilidad de obtener más de 2 CD con información?
f) ¿Cuál es su promedio de los CD en blanco?
g) ¿Cuál es su varianza de los CD en blanco?

Solución

Total de CD en blanco =	8	= M
Total de CD con Información =	12	= N - M
Total de CD =	20	= N
Tamaño Muestra =	4	= n

Tabla N°95: *Distribución Hipergeométrica, con CD en Blanco*

xi	p(xi)	p(xi = 1)	p(xi = 2)	p(1 < xi)	E(xi) = ∑xi p(xi)	V(xi) = ∑(xi-E(xi))^2 p(xi)
4	0.01444788			0.01444788	0.05779154	0.083219814
3	0.13869969			0.13869969	0.41609907	0.271851393
2	0.38142415		0.38142415	0.38142415	0.7628483	0.061027864
1	0.36326109	0.36326109			0.36326109	0.130773994
0	0.10216718				0	0.261547988
Total	1	0.36326109	0.38142415	0.53457172	1.6	0.808421053

a) La probabilidad de que 1 CD este en blanco es 0.36326109

b) La probabilidad de que 2 CD estén en blanco es 0.38142415

c) La probabilidad de obtener más de 1 CD en blanco 0.53457172

Total de CD con Información =	12	= M
Total de CD en Blanco =	8	= N - M
Total de CD =	20	= N
Tamaño Muestra =	4	= n

Tabla N°96: *Distribución Hipergeométrica, con CD con información*

x	p(x)	p(x = 3)	p(1 < x)
4	0.10216718		0.10216718
3	0.36326109	0.36326109	0.36326109
2	0.38142415		0.38142415
1	0.13869969		
0	0.01444788		
Total	1	0.36326109	0.84685243

d) La probabilidad de que 3 CD tengan información es 0.36326109

e) La probabilidad de obtener más de 2 CD con información es 0.84685243

f) El promedio de los CD en blanco es 1.6

g) La varianza de los CD en blanco es 0.808421053

Ejercicios de Aplicación

1.- Una urna contiene 5 bolas blancas y 6 rojas. SI se extraen 4 bolas sin reposición. Hallar la distribución de probabilidad del número de bolas extraídas y calcular lo siguiente:

a) ¿Cuál es la probabilidad de extraer exactamente 3 bolas rojas?

b) ¿Cuál es la probabilidad de extraer exactamente 3 bolas blancas?

c) ¿Cuál es la probabilidad de extraer exactamente 4 bolas blancas?

d) ¿Cuál es la probabilidad de extraer exactamente 4 bolas rojas?

f) ¿Cuál es su promedio y varianza de las bolas blancas?

g) ¿Cuál es su promedio y varianza de las bolas rojas?

2.- En una Clínica hay 20 enfermos, de los cuales se sabe que el 30% tiene cáncer, si se extrae aleatoriamente 4 pacientes para el despistaje de cáncer, calcular lo siguiente:

a) ¿Cuál es la probabilidad de que al menos uno tenga cáncer?

b) ¿Cuál es la probabilidad de que exactamente 3 tengan cáncer?

c) ¿Cuál es la probabilidad de que más de uno tenga cáncer?

d) ¿Cuál es la probabilidad de que ninguno tenga cáncer?

e) ¿Cuál es la probabilidad de que más de 3 no tenga cáncer?

f) ¿Cuál es la probabilidad de que menos de 3 no tenga cáncer?

g) ¿Cuál es su promedio y varianza de los enfermos de cáncer?

f) ¿Cuál es su promedio y varianza de los enfermos que no tienen cáncer?

3.- Una empresa produce candados y los agrupa en Lotes de 40 componentes, se considera aceptables si cada uno de ellos no contienen más de 3 componentes defectuosos. El procedimiento de muestreo del lote consiste en seleccionar 5 componentes aleatoriamente y sin reemplazo.

a) ¿Cuál es la probabilidad de que exactamente 1 componente defectuoso se encuentre en la muestra si hay 3 componentes defectuosos en todo el lote?

b) ¿Cuál es la probabilidad de encontrar menos de 3 componentes defectuosos en la muestra si hay 3 componentes defectuosos en todo el lote?

c) ¿Cuál es su promedio y varianza de los componentes defectuosos?

4.- Un fabricante de neumáticos para automóviles reporta que de un embarque de 5000 neumáticos enviados a un gran distribuidor local, 1000 están ligeramente manchados.

Si pequeños comerciantes compra al distribuidor 10 de estos neumáticos aleatoriamente:

a) ¿Cuál es la probabilidad de que exactamente 5 neumáticos estén manchados?

b) ¿Cuál es la probabilidad de encontrar más de 4 neumáticos manchados?

c) ¿Cuál es la probabilidad de encontrar más de 3 y menos de 10 neumáticos manchados?

d) ¿Cuál es su promedio y varianza de los neumáticos manchados?

7.5. Distribución de Probabilidades: variable aleatoria, esperanza matemática, varianza y desviación estándar de la población.

VALOR ESPERADO:

Se llama también esperanza matemática. Se trata de un operador matemático que al ser aplicado a la función probabilidad permite el cálculo de ese valor en el caso discreto, mientras que en el caso continuo se aplica a la función frecuencia.

Media, varianza y desviación estándar de una distribución de probabilidad
MEDIA (μ).- Es un valor particular que sirve para representar una distribución probabilística. También es el valor promedio a largo plazo de la variable aleatoria. Esto también se denomina valor esperado, E(x). Es decir es un promedio ponderado para el que los valores posibles que se consideran, son afectados por las probabilidades correspondientes de ocurrencia.

La varianza tiene como formula $S^2=x^2*P(x)-E^2(x)$ La desviación estándar es: S

Ejemplo 1: Un juego consiste en lo siguiente: se lanza un dado, si se obtiene un número menor o igual que 1, se pierde $2,00, si se obtiene un número mayor que 3 se gana $5,00 y si se obtiene un 2 o un 3, se gana $ 1,00, ¿cuál es la esperanza del juego y la varianza?}

Tabla N°97: *Probabilidad de Esperanza Matemática y Varianza de un juego*

	x	P(x)	x*P(x)	x^2	x^2*P(x)
1	-2	0.17	0.33333333	4	0.666666667
2	1	0.17	0.16666667	1	0.166666667
3	1	0.17	0.16666667	1	0.166666667
4	5	0.17	0.83333333	25	4.166666667
5	5	0.17	0.83333333	25	4.166666667
6	5	0.17	0.83333333	25	4.166666667
			2.5		13.5

E(x)=	2.5

$S^2=x^2*P(x)-E^2=$	7.25

La esperanza matemática es 2.5
La varianza es 7.25

Ejemplo 2 : se tiene los siguientes datos

x	3	5	9	12	21	23	27
P(x)	0,1	0,02	0,13	0,34	0,22	0,15	0,04

¿cuál es la esperanza matemática y la varianza?

Tabla N°98: *Tabla de Probabilidad de Esperanza Matemática y Varianza*

x	P(x)	x*P(x)	x^2	x^2*Px)	(x-E)	$(x-E)^2$	$(x-P)^2$*P(x)
3	0.1	0.3	9	0.9	-11.8	139.24	13.924
5	0.02	0.1	25	0.5	-9.8	96.04	1.9208
9	0.13	1.17	81	10.53	-5.8	33.64	4.3732
12	0.34	4.08	144	48.96	-2.8	7.84	2.6656
21	0.22	4.62	441	97.02	6.2	38.44	8.4568
23	0.15	3.45	529	79.35	8.2	67.24	10.086
27	0.04	1.08	729	29.16	12.2	148.84	5.9536
	1	14.8		266.42			47.38

Media=	E(x)=	14.8

$S^2=x^2*P(X)-E^2=$	47.38

La esperanza matemática es 14.8 . La varianza es 47.38

Ejemplo 3: Un juego consiste en lo siguiente: se lanza un dado, si se obtiene un número menor o igual que 1, se pierde $2,00, si se obtiene un número mayor que 3 se gana $5,00 y si se obtiene un 2 o un 3, se gana $ 1,00, ¿cuál es la esperanza del juego y la varianza y la desviación estándar?

Tabla N°99: *Probabilidad de Esperanza Matemática y Varianza de un juego*

	x	P(x)	x*P(x)	x^2	x^2*P(x)
1	-14	0.17	-2.33333333	196	32.6666667
2	-11	0.17	-1.83333333	121	20.1666667
3	19	0.17	3.16666667	361	60.1666667
4	19	0.17	3.16666667	361	60.1666667
5	21	0.17	3.5	441	73.5
6	21	0.17	3.5	441	73.5
Suma			9.16666667		320.166667
	E(x)=	9.16666667			
	S^2=	236.138889			
	S=	15.3668113			

La esperanza matemática es 9.16666667. La varianza es 236.138889

La desviación estándar es S = 15.3668113.

Ejercicios de Aplicación

1. Sea el evento: “Lanzamiento de un dado”, escriba el conjunto de resultados posibles llamado espacio muestral.

1. 40 alumnos están matriculados en Lógica, 30 en Matemática y 20 en los dos cursos. Si se selecciona al azar uno de estos alumnos, ¿cuál es la probabilidad que esté matriculado en los dos cursos

2. Las probabilidades que Juan y Juana resuelvan este problema, son respectivamente, 1/4 y 1/5 , ¿ cuál es la probabilidad que los dos lo resuelvan?, ¿Qué ninguno de los dos lo resuelva?, ¿Qué solo lo resuelva Juan?

3. Calcule la esperanza matemática y la varianza:

x	3	5	9	12	21	23	27
P(x)	0,1	0,02	0,13	0,34	0,22	0,15	0,04

4. Un juego consiste en lo siguiente: Se lanza un dado, si se obtiene un número menor o igual que 1, se pierde S/. 11,00, si se obtiene un 2 se pierde S/. 4,00, si se obtiene un 3 o 4 se gana S/. 16,00 si se obtiene mayor que 4 se gana S/. 17,00. ¿Cuál es la esperanza matemática y la varianza?.

5. Si se lanza una moneda imparcial 6 veces, cuál es la probabilidad de obtener:

   a) 3 caras?
   b) 3 sellos?
   c) como mínimo 1 cara?
   d) como máximo 5 caras?
   e) 2 ó 3 caras?

6. Se lanzan 6 dados imparciales. Halle la probabilidad de obtener:
   a) 2 cincos.
   b) 3 cincos.
   c) 2 o 3 cincos.
   d) 4 cuatros.
   e) como mínimo 5 cuatros.
   f) como máximo un tres.

V. REFERENCIALES

1. Aguilar, J. Altamira, O. García. México (2014) *INTRODUCCIÓN A LA INFERENCIA ESTADÍSTICA* Ed. Pearson. Recuperado el 18 de agosto del 2018

   https://www.freelibros.org/estadistica/introduccion-a-la-inferencia-estadistica-armando-aguilar-marquez-jorge-altamira-ibarra-y-omar-garcia-leon.html

2. Box G., Hunter S., Hunter W. (2008) *ESTADÍSTICA PARA INVESTIGADORES Diseño, innovación y descubrimiento* Ed Reverte 2da Ed. EEUU
3. Bello R., Almonte L. (2001) *GLOSARIO DE EDUCACIÓN SUPERIOR, CIENCIA T TECNOLOGÍA* Impresión: Impresora la Trinitaria, República Dominicana. Recuperado 12 de marzo del 2019 file:///C:/2010/Temas%20Para%20CD/TIC/glosario%25202.pdf
4. Gómez M., Danglot C., Vega L. "Revista Mexicana de Pediatría" México (2003)- Recuperado el 18 de agosto del 2018 http://www.medigraphic.com/pdfs/pediat/sp-2003/sp032i.pdf
5. Córdova I (2019) : ESTADÍSTICA APLICAC A LA INVESTIGACIÓN Ed Sam Marcos Lima Perú.
6. Hernández-Sampieri R. Mendoza C. (2018) *METODOLOGÍA DE LA INVESTIGACIÓN Las rutas cuantitativa, cualitativa y mixta* Ed. Mc Graw Hill México.
7. Kerlinger F. (1992) *INVESTIGACIÓN DEL COMPORTAMIENTO* Ed. Mc Graw-Hill, México
8. Lapeyere J. (2017) Glosario para la competencia TIC definiciones breves explicaciones. Recuperado el 01 de febreo del 2018 http://docentesinnovadores.perueduca.pe/?get_group_doc=106/1500591219-GLOSARIOdelacompetenciaTIC.pdf
9. Mejía E. (2005). *Técnicas e instrumentos de investigación.* Recuperado el 18 de agosto del 2018

https://www.clubensayos.com/Ciencia/Tecnicas-De-Investigacion/36299.html

10. MORAN, A. (21 de 07 de 2015). RED DE *HUMANIDADES DIGITALES*. Recuperado el 29 de ABRIL de 2017, de ¿qué es filosofía de la INFORMACIÓN?: Recuperado el 11 de agosto del 2017

    http://humanidadesdigitales.net/blog/2015/07/21/que-es-la-filosofia-de-la-INFORMACIÓN/

11. Nel L. (2017) *ESTADÍSTICA* con SPSS 24 Ed MACRO Lima Perú
12. Nel L. (2019) *ESTADÍSTICA PARA INGENIEROS* 24 Ed MACRO Lima Perú
13. Pagano R. (1999). *ESTADÍSTICA PARA LAS CIENCIAS DEL COMPORTAMIENTO*. México D.F.: International Thomson Editores.
14. López O. (2011) Ensayo: Técnicas e Instrumentos de Investigación Recuperado el 18 de febrero del 2018 https://www.clubensayos.com/Biograf%C3%ADas/MEDICI%C3%93N-T%C3%89CNICAS-E-INSTRUMENTOS-DE-INVESTIGACI%C3%93N/2226596.html
15. Reyes P. (2007) Distribución normal, prueba de normalidad y transformación de datos. Recuperado el 19 de febreo del 2018

    https://www.google.com/search?q=Reyes+P.+(2007)+Distribuci%C3%B3n+normal%2C+prueba+de+normalidad+y+transformaci%C3%B3n+de+datos&rlz=1C1SQJL_esPE836PE836&oq=Reyes+P.+(2007)+Distribuci%C3%B3n+normal%2C+prueba+de+normalidad+y+transformaci%C3%B3n+de+datos&aqs=chrome..69i57.1808j0j7&sourceid=chrome&ie=UTF-8

16. Romero F. (2001) *APRENDIENDO ESTADÍSTICA* Vol I, Vol II Ed. Hozlo S.R.L, Lima-Perú
17. Signorell A. (2018) Plolychoric Correlación. Recuperado el 12 de abril del 2019.

https://www.rdocumentation.org/packages/DescTools/versions/0.99.19/topics/CorPolychor

18. Vinuesa P. (2016) *CORRELACIÓN: Teoría y Práctica* CCG-UNAM. Recuperado el 15 de Noviembre del 2017
http://www.ccg.unam.mx/~vinuesa/R4biosciences/docs/Tema8_correlacion.pdf
http:// www.ccg.unam.mx/ ~vinuesa/
19. Triola M. México (2009) "Estadística" Décima Edición Ed. Pearson Recuperado el 11 de octubre del 2018
https://www.uv.mx/rmipe/files/2015/09/Estadistica.pdf

VI APÉNDICES

UNIVERSIDAD NACIONAL DEL CALLAO

Facultad de Ingeniería Industrial y de Sistemas

Estadística

Sílabo

1. DATOS ADMINISTRATIVOS

Tipo de curso	*Obligatorio*
Código	
Semestre	2019-A
Créditos	04
No. de horas semanales:	Teoría: 3
	Práctica: 2
Requisito:	Calculo I
Profesor:	Dr. Guillermo Antonio Mas Azahuanche
	gmas@urp.edu.pe

2. SUMILLA.

En este curso teórico - practico se desarrollan los conceptos generales que se usan en estadística, así como las técnicas de presentación de la información psicológica y las correspondientes a la selección, obtención e interpretación de los estadísticos que describen y relacionan muestras, Medidas de tendencia central, Medidas de dispersión, Puntuaciones estándar, asimetría y kurtosis, correlación y regresión asimismo se desarrolla en forma introductoria el cálculo de probabilidades, las distribuciones de Probabilidades Binomial, y Distribución Normal.

3.- ASPECTOS DEL PERFIL PROFESIONAL QUE APOYA EL CURSO

3.1. Formula y desarrolla proyectos de investigación en las diversas áreas de la psicología científica, aplicando diferentes métodos y diseños de investigación y recolectando los datos mediante la elaboración y/o estandarización de instrumentos válidos, confiables y con normas adecuadas de interpretación.

3.2. Aplica técnicas apropiadas para determinar muestras representativas de poblaciones y hace uso pertinente de las pruebas estadística paramétricas, recurriendo a programas computarizados para el análisis de datos.

4. OBJETIVOS DEL CURSO

Al finalizar el curso el alumno será capaz de:

4.1. Conocer el concepto de Estadística Descriptiva, los conceptos de Población y Muestra. Saber los distintos tipos de variables estadísticas

4.2. Conocer el concepto de Frecuencia, concepto de Intervalo de clase y marca de clase. Entender las Tablas de Frecuencias.

4.3. Comprender las distintas representaciones gráficas: Diagramas de barras, Histogramas etc.

4.2. Obtener, presentar e interpretar las medidas de tendencia central (La Media aritmética, Mediana, Moda los Deciles, Percentiles).

4.3. Obtener, procesar, presentar e interpretar las medidas de dispersión (Rango, Desviación Cuartil, Desviación media, Varianza, Desviación estándar, Coeficiente de variación)

4.5. Obtener, procesar, presentar e interpretar las puntuaciones estándar, asimetría y kurtosis

4.6. Obtener, procesar, presentar e interpretar la correlación y conocer la recta de regresión lineal

4.7. Procesar, analizar e interpretar la relación existente entre los datos y analiza su probabilidad.

4.8. Emplear el análisis combinatorio y probabilidad en los recursos que se necesitan en la estadística inferencial

4.9 Conocer, seleccionar y aplicar las técnicas para describir muestras de sujetos e interpreta correctamente los resultados obtenidos, valorando la importancia de la estadística en la evaluación e investigación psicológica.

Destrezas tecnológicas:

- Manejar con fluidez el programa de software de hoja de cálculo EXCEL que servirá para hacer todo tipo de cálculo estadístico y gráficos y la resolución de problemas de estadística.
- Manejar con fluidez el paquete de software SPSS que servirá para la resolución de problemas de estadística.
- Habilidades básicas de consulta de la red informática para la obtención y manejo de información relacionada con la estadística.

Destrezas lingüísticas:

- Adquirir y fomentar el rigor en el uso del lenguaje matemático.
- Conocer y utilizar la terminología usual de la estadística

5. PROGRAMACIÓN DE LOS CONTENIDOS

Primera Unidad

Nociones Básicas, Recolección y representación de los Datos

Logros de aprendizaje: Obtiene, procesa, presenta e interpreta datos estadísticos a fin de obtener conclusiones válidas para un grupo específico de datos, valorando su importancia para captar la realidad, comunicar y sustentar sus apreciaciones y conclusiones con precisión.

Clasificar las variables estadísticas. Agrupar correctamente los datos numéricos continuos en intervalos de clase. Calcular con soltura las Tablas de Frecuencias para los distintos tipos de variables. Representar gráficamente las distribuciones de frecuencias.

Interpretar las representaciones gráficas de las distribuciones de frecuencias.

N° de Horas: 15 horas

Semana	Contenidos	Actividades
1	- Revisión de conceptos matemáticos. Definición y Funciones de la Estadística.	- Solución de ejercicios y problemas por grupos. - Lectura comentada
2	- Términos utilizados en Estadística. - Niveles de Medición.	- Lectura analizada por grupos.
3	- Clasificación y presentación de la información psicológica	- Solución de ejercicios y problemas por grupos. - Tareas grupales

Lecturas selectas:

AUTOR	TITULO	Año	Lugar	Editorial	N° pág.
Arthur, Aron.	Estadística para Psicología (Segunda Edición),	2001	Brasil	Prentice Hall y Pearson Education S.A.- Sao Paulo-Brasil	123 – 217
Martínez Bencardino,	Estadística y Muestreo	2016	Colombia	Edición ECOE	De la 22 a 91

Segunda Unidad

Medidas de Tendencia Central

Logros de aprendizaje: Obtiene, procesa, presenta e interpreta datos estadísticos a fin de obtener conclusiones válidas para un grupo específico de datos, valorando su importancia para captar la realidad, comunicar y sustentar sus apreciaciones y conclusiones con precisión. Calcular satisfactoriamente las medidas de posición centrales y no centrales.

N° de Horas: 5 horas

SEMANA	CONTENIDOS	ACTIVIDADES
4	- Media aritmética, Mediana, Moda - Deciles , Percentiles	- Lecturas individuales - Ejercicios y Problemas Primera Práctica Calificada (PC1)

Lecturas selectas:

AUTOR	TITULO	Año	Lugar	Editorial	N° pág.
Romero Revilla, Felix	Aprendiendo estadística Vol I,	2001	Perú	Hozlo S.R.L, Lima-Perú	98 - 130

Tercera Unidad

Medidas de Dispersión

Logros de aprendizaje: Obtiene, procesa, presenta e interpreta datos estadísticos a fin de obtener conclusiones válidas para un grupo específico de datos, valorando su importancia para captar la realidad, comunicar y sustentar sus apreciaciones y conclusiones con precisión.

Calcular correctamente las medidas de dispersión absolutas y relativas. Construir el gráfico caja de una variable estadística. Interpretar el gráfico caja de una variable estadística

N° de Horas: 5 horas

5	- Rango - Desviación Cuartil, Desviación media - Varianza - Desviación estándar - Coeficiente de variación	- Lecturas comentadas. - Ejercicios y Problemas

Lecturas selectas:

AUTOR	TITULO	Año	Lugar	Editorial	N° pág.
Romero Revilla, Felix	Aprendiendo estadística Vol I,	2001	Perú	Hozlo S.R.L, Lima-Perú	131 - 160

Cuarta Unidad

Puntuaciones Estándar, Asimetría y Kurtosis

Logros de aprendizaje: Obtiene, procesa, presenta e interpreta datos estadísticos a fin de obtener conclusiones válidas para un grupo específico de datos, valorando su importancia para captar la realidad, comunicar y sustentar sus apreciaciones y conclusiones con precisión.

N° de Horas: 10 horas

SEMANA	CONTENIDOS	ACTIVIDADES
6	- Puntuaciones estándar y la curva normal.	- Estudio de investigaciones realizadas en Psicología. - Lecturas comentadas
7	- Asimetría y Kurtosis. - Aplicaciones	- Presentación y discusión de ejemplos. - Solución de ejercicios y problemas por grupos. - Segunda Práctica Calificada(PC2)
8	EXAMEN PARCIAL.	

Lecturas selectas:

AUTOR	TITULO	Año	Lugar	Editorial	N° pág.
Romero Revilla, Felix	Aprendiendo estadística Vol I,	2001	Perú	Hozlo S.R.L, Lima-Perú	161 - 186

Quinta Unidad

Correlación y Regresión

Logros de aprendizaje: Procesa, analiza e interpreta la relación existente entre 2 variables. Formula y aplica los modelos de tendencia más adecuados, con rigurosidad y precisión.

Nº de Horas: 10 horas

SEMANA	CONTENIDOS	ACTIVIDADES
9	- Introducción al Análisis Multivariado - Correlación. Aplicaciones	- Ejemplificación con variables psicológicas. - Solución de ejercicios por grupo.
10	- Regresión - Aplicaciones	- Solución de ejercicios y problemas por grupos.

Lecturas selectas:

AUTOR	TITULO	Año	Lugar	Editorial	Nº pág.
Romero Revilla, Felix	Aprendiendo estadística Vol I,	2001	Perú	Hozlo S.R.L, Lima-Perú	187 - 226

Sexta Unidad

Probabilidades

Logros de aprendizaje:

- Entender los principios básicos de conteo y los de variaciones, permutaciones y combinaciones ya sea con o sin repetición.
- Entender el concepto de suceso aleatorio.
- Conocer la axiomática de la teoría de la probabilidad y comprender la interpretación frecuencialista de los axiomas.
- Entender el concepto de probabilidad condicional e independencia de sucesos.
- Ser capaz de interpretar los teoremas de la probabilidad total y de Bayes.
- Comprender el concepto de variable aleatoria unidimensional y bidimensional.
- Comprender el concepto de función de distribución.
- Saber los distintos tipos de variables aleatorias, discretas y continuas.

- Entender el concepto de función de probabilidad y su relación con la función de distribución.
- Entender el concepto de distribuciones condicionales.
- Saber las distintas medidas de una variable aleatoria.
- Entender los diversos modelos de variables aleatorias.
- Saber identificar, a partir de un fenómeno aleatorio real, cuál de los diversos modelos de variables aleatorias vistos es el que mejor se ajusta.

N° de Horas: 15 horas

SEMANA	CONTENIDOS	ACTIVIDADES
11	- Análisis Combinatorio	- Solución de ejercicios y problemas de la Guía de Prácticas.
12	- Probabilidades: nociones preliminares, reglas. Aplicaciones - Sucesos independientes, probabilidad total y teorema de Bayes. Aplicaciones	- Solución de ejercicios y problemas de la Guía de Prácticas. - Lectura.
13	- Distribución de Probabilidades: variable aleatoria, esperanza matemática, varianza y desviación estándar de la población. - Distribución binomial. Aplicaciones	- Solución de ejercicios y problemas de la Guía de Prácticas, por grupos. - Lectura
14	- Distribución normal. Propiedades. Aplicaciones.	- Solución de ejercicios y problemas de la Guía de Prácticas.
15	- Distribución normal estándar. Uso de la curva normal Estándar. Aplicaciones.	- Cuarta práctica calificada (PC4).

Lecturas selectas:

AUTOR	TITULO	Año	Lugar	Editorial	N° pág.
Romero Revilla, Felix	Aprendiendo estadística Vol II,	2001	Perú	Hozlo S.R.L, Lima-Perú	1 – 122
Martínez Bencardino, Ciro	Estadística y Muestreo	2016	Colombia	Edición ECOE	

16	- **Examen Final**

17	- **Examen Sustitutorio**

Técnicas didácticas utilizadas:

La asignatura se desarrolla en tres modalidades didácticas:

- **Clases teóricas:** Se desarrollan mediante exposición del profesor cumpliendo el calendario establecido. En estas clases se estimula la participación activa del estudiante, mediante preguntas, solución de problemas, discusión de casos, búsqueda de información bibliográfica y por Internet.
- **Clases prácticas:** Se desarrollan las habilidades y actitudes descritas en las competencias. Se plantean ejercicios y casos a ser resueltos con los conocimientos adquiridos en las clases teóricas.
- **Clases de laboratorio:** Se realizarán con el software adecuado, tanto el Excel y el SPSS, que permita al alumno hacer cálculos y/o ejercicios y problemas de, Medidas de tendencia central, Medidas de dispersión, Puntuaciones estándar, asimetría y kurtosis, correlación y regresión y visualizar los aspectos más importantes de uso de la estadística en la Psicología. Los casos a resolver se entregarán con anticipación para que los informes incluyan investigación, actualización y conocimiento profundo del mismo.

Equipos y materiales:

Los equipos como computador y proyector multimedia y los materiales como el texto, separatas, software SPSS y Excel 2010 y el aula virtual permitirán la mejor comprensión de los temas tratados en donde se colocará los PTT y los PDF de contenidos del curso semana a semana.

6.- EVALUACIÓN

La asistencia mínima del 70% a las clases realizadas es condición para la aprobar la asignatura. El alumno o alumna no estará calificada para rendir el examen final si es que no cumple esta condición.

6.1 Criterios

El sistema de evaluación es permanente. Comprende evaluaciones de los conocimientos, habilidades y actitudes.

Para evaluar los conocimientos se utilizan las prácticas calificadas y exámenes. Para evaluar las habilidades se utilizan adicionalmente a las anteriores las intervenciones orales y exposiciones. Para evaluar las actitudes, se utiliza la observación del alumno, su comportamiento, responsabilidad, respeto, iniciativa y relaciones con el profesor y alumnos.

La redacción, orden y ortografía influyen en la calificación de las pruebas escritas.

6.2. TEORÍA.

		Peso
EXAMEN PARCIAL	= EP	1
EXAMEN FINAL	= EF	1

EXAMEN SUSTITUTORIO = ES (Reemplaza al examen más bajo)

6.3. PRÁCTICA

Se tomarán 4 prácticas calificadas: PC1, PC2, PC3, PC4, SEGÚN CALENDARIO DE ACTIVIDADES **y se eliminará la practica de menor calificación.**

6.4. PROMEDIO DE PRÁCTICAS: PP (Con 2 decimales)

$\boxed{PP=(PC1+PC2+PC3+PC4)/3}$

6.5. PROMEDIO FINAL: PF

La nota final del curso se calculará utilizando la siguiente fórmula: Promedio de las tres prácticas calificadas + nota del primer examen parcial + nota del examen final. La sumatoria dividida entre tres.

$\boxed{PF= (PP+EP+EF) / 3}$, Sólo se redondeará la fracción 0.5 o mayor en el promedio final. (Promedio de prácticas + Parcial + Final) / 3 = **Nota final**

Sólo se redondeará la fracción 0.5 o mayor en el promedio final

7. BIBLIOGRAFÍA GENERAL

1) Pagano, Robert. (1999). Estadística Para las Ciencias del Comportamiento. México D.F.: International Thomson Editores.
2) Box G., Hunter S., Hunter W. (2008) *Estadística para Investigadores. Diseño, innovación y descubrimiento*. Editorial Reverté Segunda Edición 2008

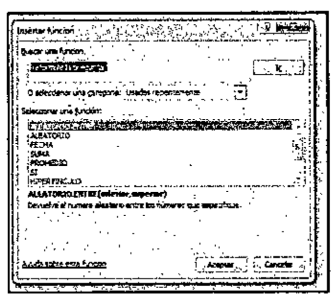
Bibliografía Básica

AUTOR	TITULO	Año	Lugar	Editorial	Nº pág.
Pagano, Robert	Estadística Para las Ciencias del Comportamiento	1999	México	International Thomson	
Johnson, Robert	Estadística Elemental	1991	México	Trillas, México	542
Kerlinger, Fred	Investigación del Comportamiento	1992	México	Mc Graw-Hill, México	754
Dalen, Van y Meyet	Manual de Técnicas de la Investigación Educacional	1994	Educador	Paidós Educador, México	464
Romero Revilla, Felix	Aprendiendo estadística Vol I,	2001	Perú	Hozlo S.R.L, Lima-Perú	294
Arthur, Aron.	Estadística para Psicología (Segunda Edición),	2001	Brasil	Prentice Hall y Pearson Education S.A.-Sao Paulo-Brasil	439
Box G., Hunter S., Hunter W.	Estadística para Investigadores	2008	EEUU	Reverte 2da Ed.	

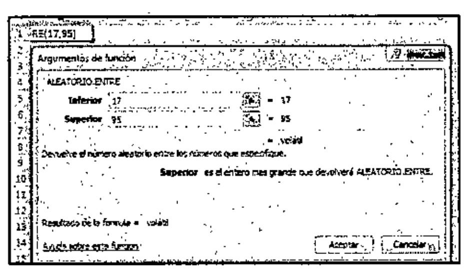
VI. ANEXOS

CLASE DE LABORATORIO

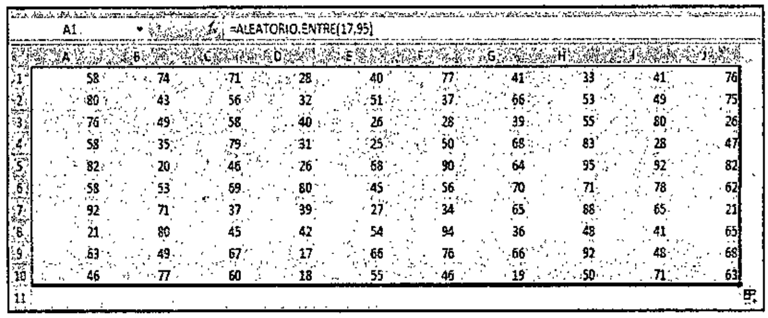
1) Crear un grupo de 100 números de forma aleatoria y elabore una clasificación usando la regla se Stuker para clasificar y agrupar estos datos en una tabla de datos agrupados, llene la tabla y elabore un histograma de frecuencia absoluta simple. Suponga que son asistentes a una película de extreno según su edad. Esto se realiza de la siguiente manera.
   - Damos click en insertar función y buscamos la función de ALEATORIO.ENTRE.



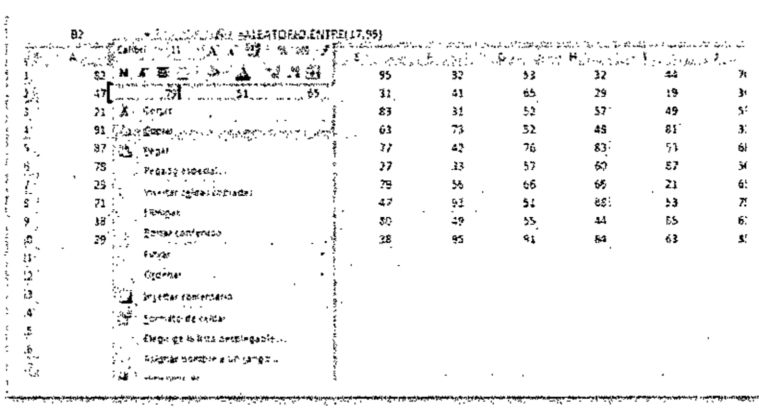
   - Luego insertamos nuestros límites de números.



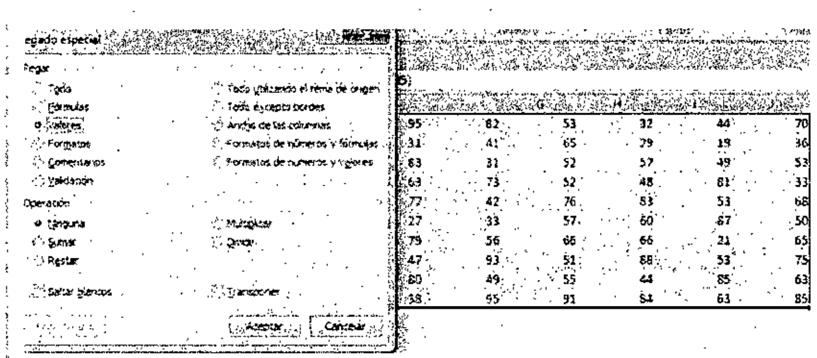
   - Luego seleccionamos 100 datos arrastramos número y hacemos un cuadrado de 10x10.



- Ahora se genera un problema que al presionar suprimir nuestros números irán variando. Esto se soluciona dando un pegado especial.



- Una vez que copiamos los números daremos click en pegado especial o pegado1,2,3.



2) Ahora hallaremos nuestro valor máximo, valor mínimo, R, K, TIC y D. Con la teoría hecha en clase.

54	79	77	41	92	91	18	25	89	3
55	24	73	36	49	74	63	94	67	4
54	66	60	65	48	55	30	55	58	7
15	76	28	93	25	73	28	18	47	9
20	50	94	66	31	49	78	76	66	9
41	17	43	45	26	48	91	61	67	8
54	44	78	35	42	63	66	40	90	7
33	60	18	58	65	51	34	33	91	3
43	49	21	64	30	36	37	82	84	6

Vmax=	94	
Vmín=	17	
R=	77	
K=	7.60	8
TIC=	10.13	11
D=	11	

3) Construiremos nuestros intervalos.
   - Dividimos nuestro D en 2, en nuestro caso fue 11, entonces será 11/2=5.5, nuestro D podría tomar dos valores que son el 5 y 6 y utilizaremos la que mejor nos convenga.

Vmín-5=12

y comenzamos a correr de 11 en 11 que fue nuestro valor del TIC

Vmax=	94	
Vmín=	17	
R=	77	
K=	7.60	8
TIC=	10.13	11
D=	11	
	12	23
	23	34
	34	45
	45	56
	56	67
	67	78
	78	89
	89	100

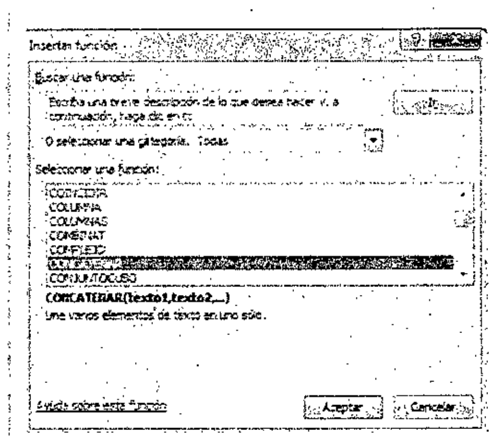
   - **Llenamos nuestra parte superior del cuadro.**

		Clase	Int de Clase	xi	fi	Fi	Fi*	hi	Hi	Hi*
12	23									
23	34									
34	45									
45	56									
56	67									
67	78									
78	89									
89	100									

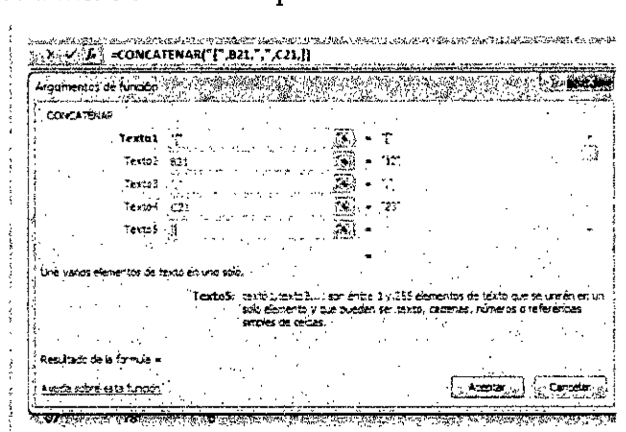
   - Nuestro K hallado debe de coincidir con nuestra cantidad de clases.

Vmáx=	94		
Vmín=	17		
R=	77		
K=	7.60	8	
TIC=	10.13	11	
D=	11		
			Clase
	12	23	1
	23	34	2
	34	45	3
	45	56	4
	56	67	5
	67	78	6
	78	89	7
	89	100	8

- Damos click en función y buscamos la función concatenar.



- Comenzamos a llenar los datos requeridos.



- Finalmente arrastramos para todas nuestras clases.

	Clase	Int de Clas
12	23	1 [12,23)
23	34	2 [23,34)
34	45	3 [34,45)
45	56	4 [45,56)
56	67	5 [56,67)
57	78	6 [67,78)
78	89	7 [78,89)
89	100	8 [89,100)

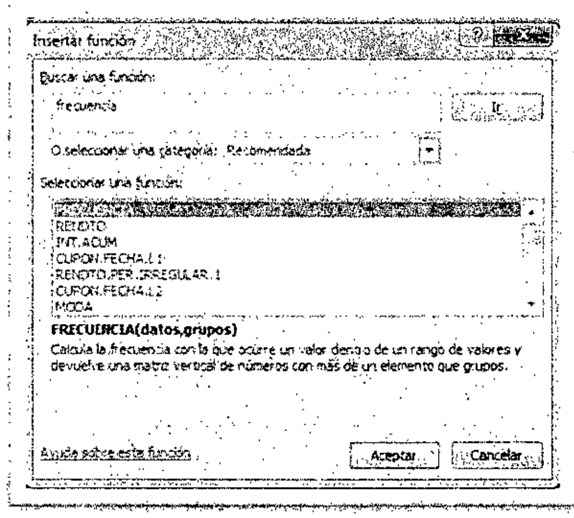
4) Aumentamos un intervalo inferior más y uno superior más.

	Clase	Int de Clase
1	12	[1,12)
12	23	1 [12,23)
23	34	2 [23,34)
34	45	3 [34,45)
45	56	4 [45,56)
56	67	5 [56,67)
67	78	6 [67,78)
78	89	7 [78,89)
89	100	8 [89,100)
100	111	[100,111)

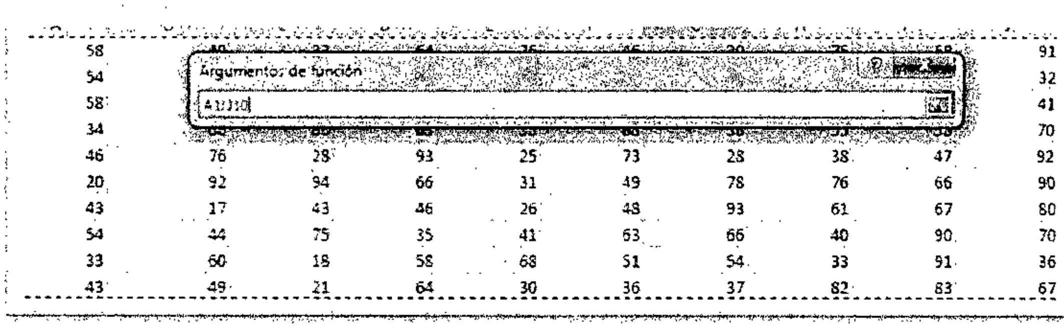
5) Llenamos nuestro "xi" o "mi" es la marca de clase que es igual a semisuma de nuestros intervalos de clase.

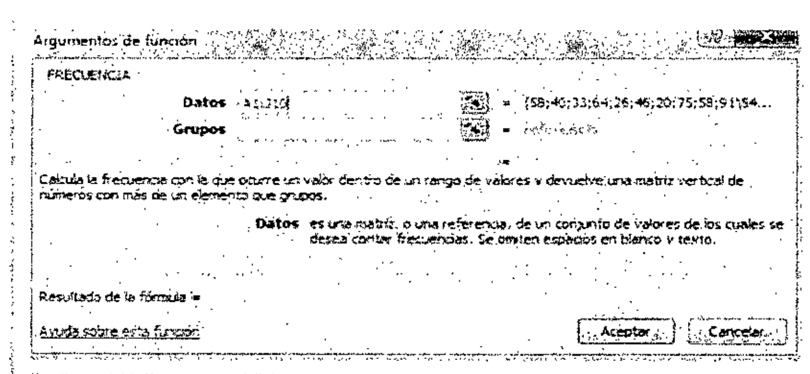
Int de Clase	xi
[1,12)	
[12,23)	17.5
[23,34)	28.5
[34,45)	39.5
[45,56)	50.5
[56,67)	61.5
[67,78)	72.5
[78,89)	83.5
[89,100)	94.5
[100,111)	

6) Llenamos la frecuencia.
   - Usamos la función frecuencia.

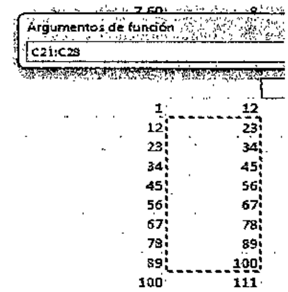


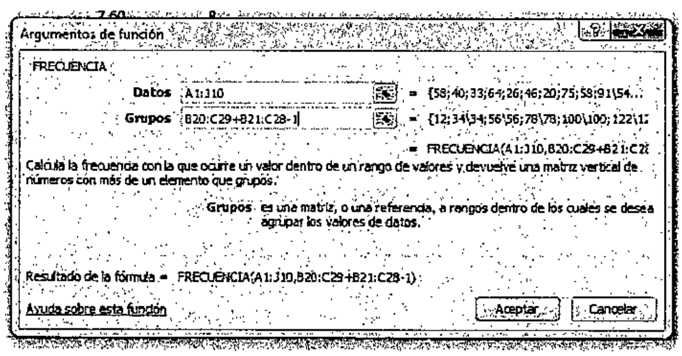
- En datos seleccionaremos todos nuestros datos.





- En grupos seleccionaremos nuestros límites superiores y le restaremos 1. OJO LE RESTAMOS 1 PORQUE ESTAMOS USANDO UNA LLAVE SEMIABIERTA.





- Seleccionamos toda nuestra columna de frecuencia.

		Clase	Int de Clase	xi	fi
1	12		[1,12)		
12	23	1	[12,23)	17.5	6
23	34	2	[23,34)	28.5	
34	45	3	[34,45)	39.5	
45	56	4	[45,56)	50.5	
56	67	5	[56,67)	61.5	
67	78	6	[67,78)	72.5	
78	89	7	[78,89)	83.5	
89	100	8	[89,100)	94.5	
100	111		[100,111)		

- Presionamos F2

		Clase	Int de Clase	xi	fi	Fi	Fi'
1	12		[1,12)				
12	23	1	[12,23)	17.5	=FRECUENCIA(A1:J10,C21:C28-1)		
23	34	2	[23,34)	28.5			
34	45	3	[34,45)	39.5			
45	56	4	[45,56)	50.5			
56	67	5	[56,67)	61.5			
67	78	6	[67,78)	72.5			
78	89	7	[78,89)	83.5			
89	100	8	[89,100)	94.5			
100	111		[100,111)				

- Presionamos **Ctrl + shift + enter**

		Clase	Int de Clase	xi	fi
1	12		[1,12)		
12	23	1	[12,23)	17.5	6
23	34	2	[23,34)	28.5	15
34	45	3	[34,45)	39.5	17
45	56	4	[45,56)	50.5	13
56	67	5	[56,67)	61.5	16
67	78	6	[67,78)	72.5	13
78	89	7	[78,89)	83.5	7
89	100	8	[89,100)	94.5	13
100	111		[100,111)		

- Esto se comprueba haciendo la sumatoria de las frecuencias y nos debe salir igual a la cantidad de datos.

		Clase	Int de Clase	xi	fi
1	12		[1,12)		
12	23	1	[12,23)	17.5	6
23	34	2	[23,34)	28.5	15
34	45	3	[34,45)	39.5	17
45	56	4	[45,56)	50.5	13
56	67	5	[56,67)	61.5	16
67	78	6	[67,78)	72.5	13
78	89	7	[78,89)	83.5	7
89	100	8	[89,100)	94.5	13
100	111		[100,111)		100

Nuestra tabla fue de 10x10 por ende hay 100 datos.

7) Llenamos toda nuestra tabla con la teoría explicada en clase.

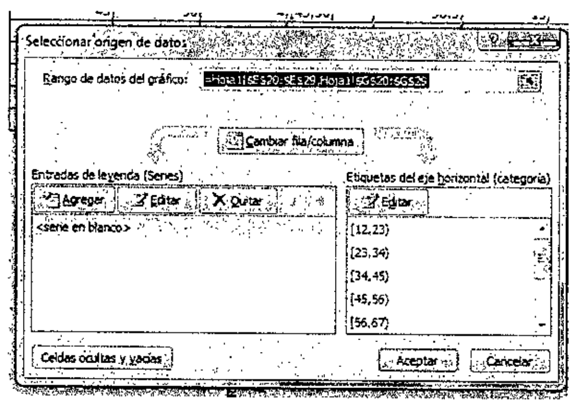
		Clase	Int de Clase	xi	fi	Fi	Fi*	hi	Hi	Hi*
1	12		[1,12)							
12	23	1	[12,23)	17.5	6	6	100	0.06	0.06	1
23	34	2	[23,34)	28.5	15	21	94	0.15	0.21	0.94
34	45	3	[34,45)	39.5	17	38	79	0.17	0.38	0.79
45	56	4	[45,56)	50.5	13	51	62	0.13	0.51	0.62
56	67	5	[56,67)	61.5	16	67	49	0.16	0.67	0.49
67	78	6	[67,78)	72.5	13	80	33	0.13	0.8	0.33
78	89	7	[78,89)	83.5	7	87	20	0.07	0.87	0.2
89	100	8	[89,100)	94.5	13	100	13	0.13	1	0.13
100	111		[100,111)		100			1		

8) Realizaremos el histograma.

- Seleccionamos nuestra fi VS los intervalos de clase.

Edad	
[12-23]	6
[23-34]	15
[34-45]	17
[45-56]	13
[56-67]	16
[67-78]	13
[78-89]	7
[89-100]	13

- Colocamos insertar Grafico
- Luego cambiamos los colores de las barras y el diseño para poder apreciarlo mejor.
- Damos click derecho en nuestro histograma y seleccionamos en seleccionar datos.



- Damos click en agregar y en nombre de serie seleccionamos todo nuestro "fi".
- En valores de serie, seleccionamos todo nuestro "fi".
- Luego aceptamos y luego seleccionamos celdas ocultas y vacías y seleccionamos cero (para que el polígono de frecuencias empiece en cero.
- Luego ponemos colores a las columnas y damos un título al grafico y ponemos las variables.