CONCEPTOS BÁSICOS DE ESTADÍSTICA DESCRIPTIVA E INFERENCIAL

EVALUACIÓN DIAGNÓSTICA

Lea cuidadosamente cada una de las siguientes interrogantes y conteste según sus conocimientos, no importa si se equivoca. La presente evaluación puede ser resuelta de manera grupal o individual. Siempre trabaje con disciplina, honradez y buena voluntad. Recuerde que el éxito se refleja en nuestro trabajo y hay que lograrlo, tarea tras tarea, y merecer ese logro. Los Autores

Según la naturaleza de los siguientes enunciados, escriba en el paréntesis la letra V si es verdadero o la F si es falso. Si su respuesta es F escriba el ¿por qué? de su respuesta.

1) La Estadística se encarga del estudio de las características cualitativas del fenómeno. ( )

2) A la Estadística le interesan los fenómenos de tipo cuantitativo. ( )

3) A la Estadística solamente le interesa la recopilación de datos. ( )

4) Los fines de la estadística son conocer las características de los fenómenos, analizarlos y predecir lo que sucederá en el futuro. ( )

5) Los objetivos de la Estadística son recopilar, organizar, tabular y presentar gráficamente los datos, proporcionando una visión cuantitativa de los fenómenos observados. ( )

6) Los métodos de la Estadística son recopilar, clasificar, tabular y presentar datos para la toma de decisiones y solución de problemas. ( )

7) La estadística descriptiva busca obtener información sobre la población basándose en el estudio de los datos de una muestra tomada a partir de ella. ( )

8) La estadística inferencial se preocupa de llegar a conclusiones basados en la muestra y luego hacerlos válidos para toda la población. ( )

9) La muestra es el conjunto de todos los elementos que tienen una característica común ( )

10) Las partes de una tabla o cuadro estadístico son: título, conceptos o columna Matriz y cuerpo del cuadro. ( )

Conteste a las siguientes preguntas

11) Sugiera 5 referentes de información que usted suponga son de tipo estadístico.
12) ¿Qué piensa usted que es la Estadística?
13) ¿Para qué sirven los censos poblacionales o de alguna otra índole?
14) Redacte un pensamiento que indique la importancia de la Estadística.
15) ¿Para qué sirven los gráficos estadísticos?. Enumere los que usted conoce.
16) ¿Qué son las medidas de tendencia central?. Enumere las que usted conoce.
17) Defina con sus propias palabras lo que entiende por medidas de dispersión. Enumere las que usted conoce.
18) ¿Qué entiende por medidas de forma?
19) ¿En qué se diferencian la correlación y la regresión?
20) ¿Cuál es la aplicación principal de las series cronológicas?

¿QUÉ ES LA ESTADÍSTICA?

HISTORIA

Establecer con absoluta claridad y precisión el proceso de desarrollo de esta ciencia que actualmente se llama Estadística, es una tarea difícil ya que la información que se dispone es fragmentada, parcial y aislada.

Es seguro que desde la antigüedad se realizaron inventarios de habitantes, bienes, productos, etc. Estos inventarios o censos (palabra derivada del latín *cencere* que significa valuar o tasar) se realizaron con fines catastrales, tributarios y militares.

En Egipto ya en el año 3050 a. c se tiene noticias de estadísticas destinadas a fines semejantes a los señalados y especialmente en la construcción de las pirámides.

En China en el año 2000 a. c. se conocen estudios similares. El nacimiento de Cristo coincide con la realización de un censo poblacional en gran escala en el Imperio Romano. Durante mucho tiempo se entendía por "estadística" la información relacionada con el gobierno, la palabra misma se deriva del latín *statisticus* o *estatus* que significa "del estado".

Ya en nuestra era, en el año 727, los árabes realizaron estadísticas similares en lo que hoy es España. En Inglaterra en el año 1083 y 1662 y el Alemania en 1741, se llevaron a cabo censos referentes a defunciones, nacimientos, enfermedades, posesión de bienes, migraciones y otros problemas y los datos obtenidos se utilizaron en la previsión y planificación. En América se realizaron encuestas mediante el sistema de "quipus".

El desarrollo científico de la estadística comienza recién en el siglo XVII, con la introducción en el pensum de estudio de las universidades en Alemania.

A comienzos del siglo XX, una nueva aportación de la escuela inglesa, preocupada por problemas de índole agropecuaria y biométrica coloca a la estadística en el tramo final de su establecimiento como ciencia.

En general las primeras aplicaciones de la estadística tuvieron que ver directamente con las actividades del estado. Se cree que la primera persona que hizo uso de la palabra estadística fue Godofredo Achenwall (1719-1772), profesor y economista alemán, escribió sobre el descubrimiento de una nueva ciencia que llamó estadística (palabra derivada de *Staat* que significa gobierno) y que definió como "el conocimiento profundo de la situación respectiva y comparativa de cada estado".

DEFINICIÓN

Existen muchas definiciones de Estadística, pero en síntesis la podemos definir como la ciencia rama de la Matemática que se ocupa de recolectar, organizar, presentar, analizar e interpretar información cuantitativa para obtener conclusiones válidas, solucionar problemas, predecir fenómenos y ayudar a una toma de decisiones más efectivas.

APLICACIONES

La Estadística anteriormente sólo se aplicaba a los asuntos del Estado, pero en la actualidad la utilizan las compañías de seguros, empresarios, comerciantes, educadores, etc. No hay campo de la actividad humana que no requiera del auxilio de esta ciencia, así por ejemplo:

- El educador mediante la estadística podrá conocer si un estudiante lee muy bien o regular, si la asistencia es normal o irregular, si la estatura está en relación con la edad, media aritmética de rendimiento escolar en un período determinado, etc.

- El hombre de negocios realiza encuestas estadísticas para determinar la reacción de los consumidores frente a los actuales productos de la empresa y en el lanzamiento de los nuevos.

- El economista emplea una amplia gama de estadísticas para estudiar los planes de los consumidores y efectuar pronósticos sobre las tendencias de las actividades económicas

- El gerente de una empresa eléctrica proporciona un buen servicio a la comunidad mediante la variación estacional de las necesidades de carga

- El sociólogo trata de auscultar la opinión pública mediante encuestas, para determinar su preferencia por un candidato presidencial, o su posición frente a determinados problemas económicos, políticos o sociales

- El geólogo utiliza métodos estadísticos para determinar las edades de las rocas

- El Genetista determina las semejanzas entre los resultados observados y esperados en una experiencia genética se determina estadísticamente

FINES

- **Conocer** las características de un grupo de casos de estudio.

- **Comparar** entre los resultados actuales y los obtenidos en experiencias pasadas para determinar las causas que han influenciado en los cambios.

- **Predecir** lo que pude ocurrir en el futuro de un fenómeno.

OBJETIVOS

- **Describir** numéricamente las características de los conjuntos de observaciones. Esta etapa consiste en recopilar, organizar, tabular y presentar gráficamente los datos, proporcionando una visión cuantitativa de los fenómenos observados.

- **Analizar** los datos de manera objetiva con el fin de disponer de un concepto claro de universo o población y adoptar decisiones basadas en la información proporcionada por los datos de la muestra.

- **Estimar** o predecir lo que sucederá en el futuro con un fenómeno de una manera relativamente aceptable, así por ejemplo, podemos estimar cuál será la población del país dentro de un determinado número de años conociendo la actual.

MÉTODOS

- Recopilación.- Consiste en la obtención de datos relacionados con el problema motivo de estudio, utilizando instrumentos, tales como: cuestionarios, entrevistas, informes, memorias, etc.

- Organización.- Consiste en realizar una crítica, corrección, clasificación y tabulación de los datos obtenidos en el paso anterior.

- Presentación.- Consiste en mostrar datos de manera significativa y descriptiva. Los datos deben colocarse en un orden lógico que revele rápida y fácilmente el mensaje que contienen. La presentación se la puede hacer a través de gráficos estadísticos.

- Análisis.- Consiste en descomponer el fenómeno en partes y luego examinar cada una de ellas con el objetivo de lograr una explicación, haciendo uso, en su mayoría, de los cálculos matemáticos.

- Interpretación.- Consiste en un proceso mental, mediante el cual se encuentra un significado más amplio de los datos estadísticos con el objetivo de llegar a conclusiones para la toma de decisiones y solución de problemas.

CLASIFICACIÓN DE LA ESTADÍSTICA

Estadística Descriptiva o Deductiva

Es un proceso mediante el cual se recopila, organiza, presenta, analiza e interpreta datos de manera tal que describa fácil y rápidamente las características esenciales de dichos datos mediante el empleo de métodos gráficos, tabulares o numéricos, así por ejemplo:

Supóngase que un docente de Matemática calcula la calificación promedio de uno de sus cursos a su cargo. Como solo se está describiendo el desempeño del curso pero no hace ninguna generalización acerca de los diferentes cursos, en este caso el maestro está haciendo uso de la Estadística Descriptiva.

Estadística Inferencial o Inductiva

Llamada también inferencia estadística, la cual consiste en llegar a obtener conclusiones o generalizaciones que sobrepasan los límites de los conocimientos aportados por un conjunto de datos. Busca obtener información sobre la población basándose en el estudio de los datos de una muestra tomada a partir de ella, así por ejemplo:

Supóngase ahora que el docente de Matemática utiliza el promedio de calificaciones obtenidas por uno de sus cursos para estimar la calificación promedio de los 5 cursos a su cargo. Como se está realizando una generalización acerca los diferentes cursos, en este caso el maestro usa la Estadística Inferencial.

CONCEPTOS Y DEFINICIONES BÁSICAS

POBLACIÓN

Llamado también universo o colectivo es el conjunto de todos los elementos que tienen una característica común.

Una población puede ser finita o infinita. Es ***población finita*** cuando está delimitada y conocemos el número que la integran, así por ejemplo: Estudiantes de la Universidad UTN. Es ***población infinita*** cuando a pesar de estar delimitada en el espacio, no se conoce el número de elementos que la integran, así por ejemplo: Todos los profesionales universitarios que están ejerciendo su carrera.

MUESTRA

Es un subconjunto de la población. Ejemplo: Estudiantes de 2do Semestre de la Universidad UTN.

Sus principales características son:

Representativa.- Se refiere a que todos y cada uno de los elementos de la población tengan la misma oportunidad de ser tomados en cuenta para formar dicha muestra.

Adecuada y válida.- Se refiere a que la muestra debe ser obtenida de tal manera que permita establecer un mínimo de error posible respecto de la población.

Para que una muestra sea fiable, es necesario que su tamaño sea obtenido mediante procesos matemáticos que eliminen la incidencia del error.

Para calcular el tamaño de la muestra suele utilizarse la siguiente fórmula:

$$n = \frac{N\sigma^2 Z^2}{(N-1)e^2 + \sigma^2 Z^2}$$

Donde:

n = el tamaño de la muestra.

N = tamaño de la población.

$\sigma =$ Desviación estándar de la población que, generalmente cuando no se tiene su valor, suele utilizarse un valor constante de 0,5.

Z = Valor obtenido mediante niveles de confianza. Es un valor constante que, si no se tiene su valor, se lo toma en relación al 95% de confianza equivale a 1,96 (como más usual) o en relación al 99% de confianza equivale 2,58, valor que queda a criterio del encuestador.

e = Límite aceptable de error muestral que, generalmente cuando no se tiene su valor, suele utilizarse un valor que varía entre el 1% (0,01) y 9% (0,09), valor que queda a criterio del encuestador.

Ejemplo ilustrativo: Calcular el tamaño de la muestra de una población de 1000 elementos.

Solución:

Se tiene N=1000, y como no se tiene los demás valores se tomará $\sigma = 0{,}5$, Z = 1,96 y e = 0,05.

Reemplazando valores en la fórmula se obtiene:

$$n = \frac{N\sigma^2 Z^2}{(N-1)e^2 + \sigma^2 Z^2} = \frac{1000 \cdot 0{,}5^2 \cdot 1{,}96^2}{(1000-1) \cdot 0{,}05^2 + 0{,}5^2 \cdot 1{,}96^2} = \frac{1000 \cdot 0{,}25 \cdot 3{,}8416}{(999) \cdot 0{,}0025 + 0{,}25 \cdot 3{,}8416}$$

$$n = \frac{960{,}4}{2{,}4975 + 0{,}9604} = \frac{960{,}4}{3{,}4579} = 277{,}74 = 278$$

ELEMENTO O INDIVIDUO

Unidad mínima que compone una población. El elemento puede ser una entidad simple (una persona) o una entidad compleja (una familia), y se denomina unidad investigativa.

DATOS ESTADÍSTICOS

Son medidas, valores o características susceptibles de ser observados y contados. Como por ejemplo, la edad de los estudiantes de la Universidad UTN.

Los datos estadísticos pueden ser clasificados en ***cualitativos*** (la diferencia entre ellos es de clase y no de cantidad), ***cuantitativos*** (representan magnitudes), ***cronológicos*** (difieren en instantes o períodos de tiempo) y ***geográficos*** (referidos a una localidad).
Los datos estadísticos se obtienen de ***fuentes primarias*** (obtenidos directamente sin intermediarios valiéndose de observaciones, encuestas, entrevistas y sondeos de opinión) y ***fuentes secundarias*** (obtenidos a través de intermediarios valiéndose de textos, revistas, documentos, publicaciones de prensa, y demás trabajos hechos por personas o entidades).

CENSO

Es una técnica de recolección de datos estadísticos que se realiza a toda la población

ENCUESTA

Es la técnica que nos permite recolectar datos estadísticos que se realiza una muestra de la población.

Se clasifica en:

- *Descriptiva.-* Cuando registra datos referentes a las características de los elementos o individuos.

- *Explicativa.-* Cuando averigua las causas o razones que originan los fenómenos.

- *Mixtas.-* Cuando es descriptiva y explicativa.

- *Por muestreo.-* Cuando recolecta información de grupos representativos de la población.

Su estructura es:

- Nombre de la institución que auspicia la encuesta.

- Tema de la encuesta.

- Objetivos de la encuesta.

- Datos informativos: Lugar, fecha, y otros datos que se considere necesario según la naturaleza de la información estadística a encuestarse.

- Instrucciones para el encuestado para que sepa la forma de llenar la encuesta.

- Cuestionario o listado de preguntas (cerradas, abiertas, o ambas a la vez) sobre los diferentes aspectos motivo de estudio.

- Frase de agradecimiento al encuestado, como por ejemplo, ¡Gracias por su colaboración!

Las diferentes tipos de preguntas pueden ser:

- ***Abiertas.-*** Son aquellas en la cual el encuestado construye la respuesta de manera libre según su opinión y de la manera que él desea. Ejemplo: ¿Qué piensa usted sobre la política educativa del actual gobierno?

- ***Cerradas o dicotómicas.-*** Sólo pueden ser contestadas por un "sí" o por un "no". Ejemplo: ¿Está usted de acuerdo con la política educativa del actual gobierno?

Si ( )
No ( )

Como es obvio, la respuesta será forzosamente una de las alternativas planteadas: Las preguntas cerradas son fáciles de tabular y facilitan la cuantificación mediante la asignación de puntuaciones.

- ***Preguntas de elección múltiple o categorizada:*** Se trata en cierto modo de preguntas cerradas que, dentro de los extremos de una escala permiten una serie de alternativas de respuestas cuyos matices son fijados de antemano. Presentan dos formas: En abanico y de estimación

- ***Preguntas con respuesta en abanico:*** Estas preguntas permiten contestar señalando una o varias respuestas presentadas junto con la pregunta. Por ejemplo: Indique otras alternativas que considere importantes para mejorar la educación en nuestro país.

- ***Preguntas de Estimación:*** Son preguntas cuantitativas que introducen diversos grados de intensidad creciente o decreciente para un mismo ítem. Ejemplos:

-¿Cómo calificaría la política educativa del gobierno actual?

Excelente ( ) Muy Buena ( ) Regular ( ) Deficiente ( )

-¿En qué porcentaje está de acuerdo con la política educativa del gobierno actual?

100% ( ) 75% ( ) 50% ( ) 25% ( ) 0% ( )

- ¿Le interesa conocer el modelo educativo vigente?

Nada ( ) Poco ( ) Algo ( ) Mucho ( )

¿Piensa culminar sus estudios superiores?

Sí ( ) Probablemente Sí ( ) No ( ) Aún no decido ( )

VARIABLE

Son caracteres susceptibles a cambio y pueden tener diferentes valores en cada elemento o individuo.

Clasificación

- *Variable Cualitativa*
Son atributos que se expresan mediante palabras no numéricas. Como por ejemplo, profesión, religión, marca de automóvil, estado civil, sexo, raza, etc.

- *Variable Cuantitativa*
Es toda magnitud representada por números. Como por ejemplo, peso, estatura, número de habitantes, etc.

- *Variable Discreta*
Es una característica cuantitativa representada por números enteros o exactos, que generalmente resultan del proceso de conteo, como por ejemplo: número de estudiantes de la promoción del año anterior.

- *Variable Continua*
Es una característica cuantitativa que puede tomar cualquier valor representado por un número racional, que generalmente resultan del proceso de medición, como por ejemplo, tiempo destinado a estudiar Estadística

Niveles de medición

- *Nivel Nominal*
Cuando los datos sólo pueden contarse y clasificados en categorías, no existe un orden específico entre las clases. Como por ejemplo, se cuentan cuántos hombres y cuántas mujeres asisten a determinado evento.

- *Nivel Ordinal*
Cuando se ordenan los datos por jerarquías, una categoría es mayor que otra. Como por ejemplo, excelente es mejor que bueno o bueno es mejor que regular. Otro ejemplo: Una persona puede tener mucho o poco dinero.

- *Nivel de Intervalos*
Cuando se incluye todas las características del nivel ordinal, pero la diferencia entre los valores tiene un significado medido en unidades iguales que son comunes y constantes, que permiten asignar números reales a todos los miembros de la clase ordenada, facilitando el establecimiento de diferencias en grados de propiedad y entre objetos sobre la base de una medida. Como por ejemplo: La diferencia entre 70 kilogramos y 60 kilogramos, es de 10 kilogramos. Otro ejemplo: Si la temperatura de hoy es de 20 grados centígrados y la de ayer fue de 25 grados centígrados, se sabe que la de hoy es 5 grados centígrados más baja que la de ayer.

- Nivel de Razón o Cociente

Este es el nivel de medición “más alto”, tiene todas las características del nivel de intervalos y además en este nivel de medición el cero tiene significado (así si se tiene 0 dólares, entonces no se poseen fondos), y la razón (o cociente) entre dos números también es significativa (Un estudiante obtiene una calificación de 3/10 y otro 6/10, el segundo estudiante obtiene el doble que el primero).

FRECUENCIAS

- Frecuencia Absoluta (f)

Es el número de veces que se repite el valor de cada variable. La suma de frecuencias absolutas es siempre al total de datos observados.

- Frecuencia Relativa (fr)

Indica la proporción con que se repite un valor. Es el cociente entre la frecuencia absoluta y el número total de datos. La suma de las frecuencias relativas es siempre 1

$$fr = \frac{f}{n}$$

- Frecuencia Acumulada (fa)

Indica el número de valores que son menores o iguales que el valor dado. Es la suma de la frecuencia absoluta primera con la segunda, este valor con la tercera, y así sucesivamente.

- Frecuencia Porcentual ($f\%$)

Llamada también frecuencia relativa porcentual. Se obtiene multiplicando la frecuencia relativa por 100. La suma de las frecuencias porcentuales es siempre 100%. Se calcula así:

$$f\% = fr \cdot 100$$

- Frecuencia Relativa Acumulada (fra)

Es la suma de la frecuencia relativa primera con la segunda, este valor con la tercera, y así sucesivamente.

- Frecuencia Relativa Acumulada Porcentual ($fra\%$)

Indica el número de valores que son menores o iguales que el valor dado. Se obtiene multiplicando la frecuencia relativa acumulada por 100. Se calcula así:

$$fra\% = fra \cdot 100$$

Ejemplo ilustrativo:

Calcular las diferentes frecuencias de las siguientes calificaciones evaluadas sobre 10 obtenidas de 40 estudiantes en la asignatura de Estadística sin agrupar en clases:

10	8	9	8	7	8	9	10
6	7	10	9	8	8	10	8
6	5	6	8	10	5	9	9
8	10	9	7	6	7	7	6
8	10	7	8	5	9	8	5

Solución:

El ejercicio resuelto se muestra en la tabla:

Calificación	f	fr	fa	$f\%$	fra	$fra\%$
5	4	4/40 = 0,1	4	0,1·100 =10	0,1	0,1·100 =10
6	5	5/40 = 0,125	4+5 = 9	0,125·100 = 12,5	0,1+0,125 = 0,225	0,225·100 = 22,5
7	6	6/40 = 0,15	9+6 = 15	0,15·100 = 15	0,225+0,15 = 0,375	0,375·100 = 37,5
8	11	11/40 = 0,275	15+ 11 = 26	0,275·100 = 27,5	0,375+0,275 = 0,65	0,65·100 = 65
9	7	7/40 = 0,175	26+7 = 33	0,175·100 = 17,5	0,65+0,175 = 0,825	0,825·100 = 82,5
10	7	7/40 = 0,175	33+7 = 40	0,175·100 = 17,5	0,825+0,175 = 1	1·100 = 100
Total	40	1		100		

- *Calcule el Rango (R).-* También se llama recorrido o amplitud total. Es la diferencia entre el valor mayor y el menor de los datos.

$$R = x_{máx} - x_{mín}$$

- *Seleccione el Número de Intervalos de Clase (n_i).-* No debe ser menor de 5 y mayor de 12, ya que un número mayor o menor de clases podría oscurecer el comportamiento de los datos. Para calcular el número de intervalos se aplica la regla de Sturges, propuesta por Herberth Sturges en 1926:

$$n_i = 1 + 3{,}32 \cdot log(n)$$

Siendo n el tamaño de la muestra.

- *Calcule el Ancho del Intervalo (i).-* Se obtiene dividiendo el Rango para el número de intervalos

$$i = \frac{R}{ni}$$

Cuando el valor de i no es exacto, se debe redondear al valor superior más cercano. Esto altera el valor de rango por lo que es necesario efectuar un ajuste así:

$$Nuevo\ R = ni \cdot i$$

Por ejemplo:

Si una distribución de 40 datos el valor mayor es 41 y el menor es 20 se tiene:

Calculando el Rango se obtiene:
$R = x_{máx} - x_{mín} = 41 - 20 = 21$

Calculando el número de intervalos se obtiene:
$n_i = 1 + 3{,}32 \cdot log(n) = 1 + 3{,}32 \cdot log40 = 6{,}32 = 6$

Calculando el ancho se obtiene:
$i = \frac{R}{ni} = \frac{21}{6} = 3{,}5$
Redondeando se obtiene: i = 4

Calculando el nuevo rango se obtiene:
$Nuevo\ R = ni \cdot i = 6 \cdot 4 = 24$

El exceso de 3 que se tiene en este caso se distribuye entre $x_{máx}$ y $x_{mín}$. Por lo general se agrega al mayor y se quita al menor. Como por ejemplo, se podría agregar 2 al valor mayor y quitar 1 al valor menor, obteniéndose los siguientes nuevos valores:

$x_{máx} = 41 + 2 = 43$

$x_{min} = 20 - 1 = 19$

O también se podría agregar 1 al valor mayor y quitar 2 al valor menor, obteniéndose los siguientes nuevos valores:

$x_{máx} = 41 + 1 = 42$

$x_{min} = 20 - 2 = 18$

- ***Marca de Clase*** (x_m)**.-** Es el valor medio de cada clase, se obtiene sumando los límites superior (L_s) e inferior (L_i) del intervalo y dividiendo ésta suma entre 2

$$x_m = \frac{L_s + L_i}{2}$$

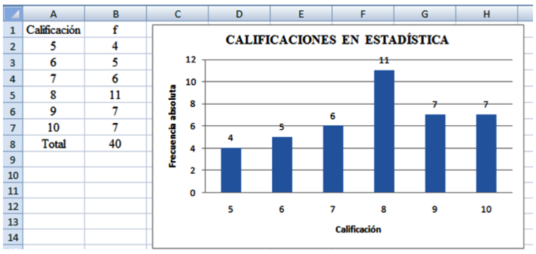
GRÁFICOS ESTADÍSTICOS BÁSICOS

Las empresas, industrias, instituciones, etc. emplean diversos gráficos estadísticas para presentar informaciones sobre diversos asuntos relativos a ellas.

Las representaciones gráficas deben conseguir que un simple análisis visual ofrezca la mayor información posible. Según el tipo del carácter que estemos estudiando, usaremos una representación gráfica u otra.

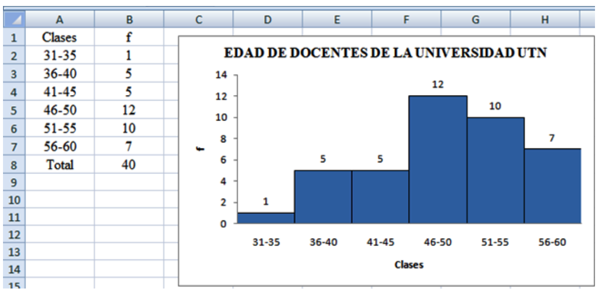
A continuación se presenta los diagramas más empleados:

DIAGRAMAS DE BARRAS.- Es un gráfico bidimensional en el que los objetos gráficos elementales son rectángulos de igual base cuya altura sea proporcional a sus frecuencias. Si en el eje horizontal se ubican las etiquetas con los nombres de las categorías, y en el eje vertical la frecuencia absoluta, la relativa o la frecuencia porcentual, toma el nombre de diagrama de barras vertical, y si se intercambian las ubicaciones de las categorías y las frecuencias, toma el nombre de diagrama de barras horizontal.



Calificación	f
5	4
6	5
7	6
8	11
9	7
10	7
Total	40

HISTOGRAMAS.- Se utiliza para datos agrupados en intervalos de clase, representando en el eje horizontal los intervalos de clase o la marca de clase, y en el eje vertical se elabora rectángulos contiguos de base el ancho del intervalo y de altura proporcional a las frecuencias representadas.

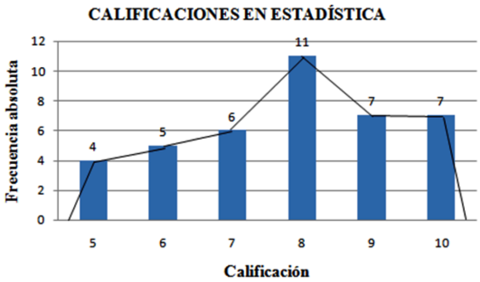


Clases	f
31-35	1
36-40	5
41-45	5
46-50	12
51-55	10
56-60	7
Total	40

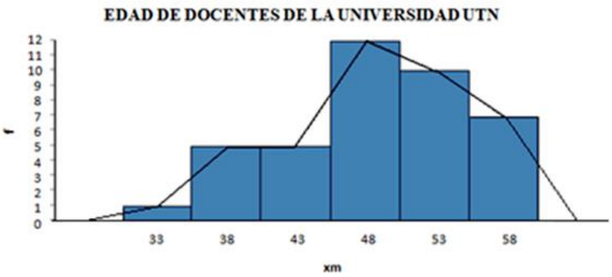
POLÍGONO DE FRECUENCIAS

Son gráficos lineales que se realizan uniendo:

a) Los puntos medios de las bases superiores de los rectángulos en un diagrama de barras.

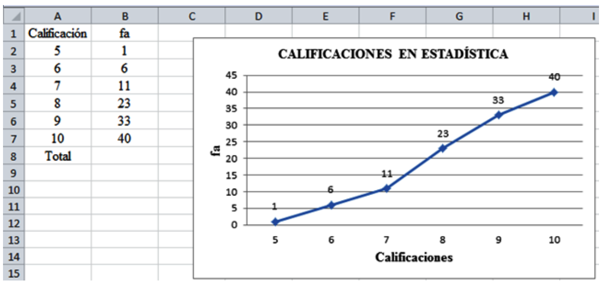


b) Los puntos medios (marcas de clase) de las bases superiores en el histograma.



Polígono de Frecuencias Acumuladas u Ojiva.- Un gráfico que recoja las frecuencias acumuladas por debajo de cualquiera de las fronteras de clase superiores respecto de dicha frontera se llama un polígono de frecuencias acumuladas u ojiva.

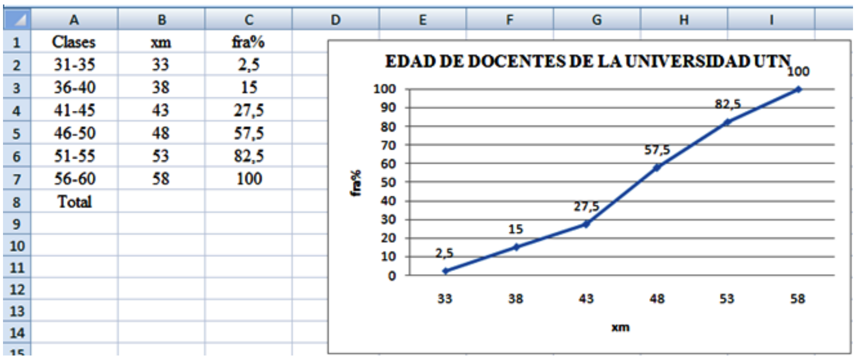
Empleando polígono de frecuencias en 2D anterior, borrando la columna de la frecuencia absoluta y escribiendo la columna de la frecuencia acumulada del ejemplo del cálculo de las frecuencias sobre las siguientes calificaciones obtenidas por 40 estudiantes en una evaluación de la asignatura de Estadística se obtiene la siguiente figura que representa a una Ojiva:



Calificación	fa
5	1
6	6
7	11
8	23
9	33
10	40
Total	

Polígono de Frecuencias Relativas Acumuladas Porcentuales.- Si se usan frecuencias fra% para realizar un polígono de frecuencias, este recibe el nombre de polígono de frecuencias relativas acumuladas porcentuales, o también llamado o*jiva de porcentajes*.

A continuación se presenta una ojiva de porcentajes elaborada en Excel empleando los datos del ejemplo de la Edad de 40 Docentes de la Universidad UTN:



Clases	xm	fra%
31-35	33	2,5
36-40	38	15
41-45	43	27,5
46-50	48	57,5
51-55	53	82,5
56-60	58	100
Total		

DIAGRAMA DE TALLO Y HOJAS

En el diagrama de tallo y hojas cada dato representa su valor y, a la vez, ocupa un espacio de forma que se obtiene simultáneamente la presentación de los datos y distribución gráfica.

En este diagrama cada valor se descompone en 2 partes: el primero o primeros dígitos (el tallo) y el dígito que sigue a los utilizados en el tallo (las hojas).Por ejemplo, el valor 32 puede descomponerse en un tallo de 3 y una hoja de 2; el valor 325 puede descomponerse en un tallo de 32 y una hoja de 5; el valor 3256 puede descomponerse en un tallo de 325 y una hoja de 6. Cada tallo puede ocupar una o más filas. Si un tallo ocupa una sola fila, sus hojas contendrán dígitos del 0 al 9; si ocupa dos filas, la primera fila contendrá dígitos del 0 al 4 y la segunda fila del 5 al 9.

La ventaja de este diagrama es que refleja a primera vista las mismas impresiones gráficas que el histograma sin necesidad de elaborar el gráfico. También tiene la ventaja de conservar los valores originales de los datos.
Ejemplo ilustrativo: A 40 estudiantes se les pidió que estimen el número de horas que habrían dedicado a estudiar la semana pasada (tanto en clase como fuera de ella), obteniéndose los siguientes resultados:

30	30	32	32	35	35	35	35
36	37	38	39	39	40	45	45
47	47	47	48	48	49	50	50
50	52	54	55	55	56	56	56
58	58	58	58	58	60	60	65

Solución:

A fin de elaborar el diagrama de tallo y hojas se ordena los datos con los dígitos iniciales de cada uno, las decenas (tallos) a la izquierda de una línea vertical, y a la derecha de esa recta el último dígito de cada dato, en este caso la unidad, conforme recorren los datos en el orden en que fueron anotados.

3	0022
3	555567899
4	0
4	55777889
5	00024
5	5566688888
6	00
6	5

Interpretaciones: Hay 4 estudiantes que dedican entre 30 y 32 horas semanales a estudiar, 10 estudiantes que dedican entre 55 y 58 horas semanales a estudiar, existe un solo estudiante que dedica 65 horas semanales a estudiar.

1.5.5) DIAGRAMA DE SECTORES

Llamado también diagrama circular o de pastel. Es un gráfico en el que a cada valor o modalidad se asigna un sector circular de área proporcional a la frecuencia que representan.

Ejemplo ilustrativo: Con los datos de la siguiente tabla sobre las calificaciones obtenidas por 40 estudiantes en una evaluación de Estadística, presentar la información a través de un diagrama de sectores:

Calificación	f
5	4
6	5
7	6
8	11
9	7
10	7
Total	40

Solución:

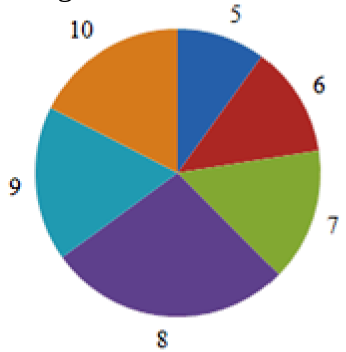
a) Se calcula la frecuencia relativa y el número de grados que representa cada calificación. El número de grados se calcula multiplicando la frecuencia relativa con 360^0, así:

$$número\ de\ grados = fr \cdot 360^0$$

Estos cálculos se muestran en la siguiente tabla:

Calificación	f	fr	Grados
5	4	0,100	36
6	5	0,125	45
7	6	0,150	54
8	11	0,275	99
9	7	0,175	63
10	7	0,175	63
Total	40	1	360

b) Se dibuja una circunferencia tomando para cada calificación tantos grados como indica la tabla anterior como se muestra en la siguiente figura:



PICTOGRAMAS

Son dibujos, figuras o signos llamativos alusivos al carácter que se está estudiando cuyo tamaño es proporcional a la frecuencia que representa los datos.

Ejemplo ilustrativo: Un equipo de fútbol en su trayectoria tiene 120 partidos ganados, 60 perdidos y 30 empatados. Al representar estos datos mediante pictogramas se obtiene:

Partidos Ganados  120

Partidos Perdidos  60

Partidos Empatados  30

Otra forma de representar los datos mediante pictogramas se muestra en la siguiente figura:

Partidos Ganados	Partidos Perdidos	Partidos Empatados
120	60	30

MEDIDAS DE TENDENCIA CENTRAL

MEDIA ARITMÉTICA SIMPLE.-Es la medida de tendencia central más utilizada por lo general se ubica hacia el centro de distribución estadística.
La principal propiedad de la media aritmética es: La suma algebraica de las desviaciones de un conjunto de datos respecto de su media aritmética es cero. Si x es un dato, su desviación respecto a $\bar{x}$ es la diferencia x-$\bar{x}$. La suma de estas diferencias es 0.

MEDIA ARITMÉTICA PONDERADA.- Cuando los números x_1, x_2, x_3,….x_k se les asocia ciertos factores peso (o pesos) w_1, w_2, w_3,….w_k, dependientes de la relevancia asignada a cada número, en tal caso se requiere calcular la media aritmética ponderada, la cual se calcula así:

MEDIA GEOMÉTRICA

- La media geométrica proporciona una medida precisa de un cambio porcentual promedio en una serie de números.
- Se utiliza con más frecuencia para calcular la tasa de crecimiento porcentual promedio de series de datos, a través del tiempo.
- Es una medida de tendencia central por lo general menor que la media aritmética salvo en el extraño caso en que todos los incrementos porcentuales sean iguales, entonces las dos medias serán iguales.
- Se le define como la raíz enésima del producto de “n” valores. Cuando los datos son bastantes o cantidades grandes, para facilitar el cálculo se lo debe simplificar pero sin alterar su naturaleza, para lo cual se puede utilizar los logaritmos de base 10.

LA MEDIANA.- La mediana, llamada algunas veces media posicional, es el valor del término medio que divide una distribución de datos ordenados en dos partes iguales, es decir, el 50% de los datos se ubican sobre la mediana o hacia los puntajes altos y el 50% restante hacia los puntajes bajos.

Propiedades:

-La Mediana no tiene propiedades que le permite intervenir en desarrollos algebraicos como la media aritmética, sin embargo, posee propiedades que ponen en evidencia ciertas cualidades de un conjunto de datos, lo cual no ocurre con la media aritmética que promedia todos los valores y suprime sus individualidades. En cambio, la mediana destaca los valores individuales.

- Tiene la ventaja de no estar afectada por las observaciones extremas, ya que no depende de los valores que toma la variable, sino del orden de las mismas.

-Para el cálculo de la mediana interesa que los valores estén ordenados de menor a mayor.

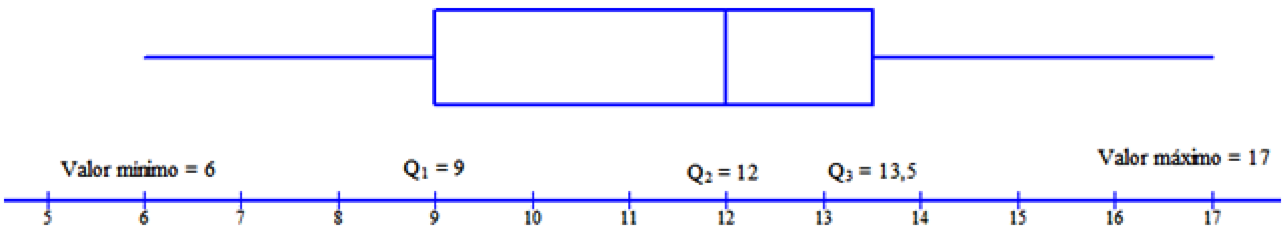
- Su aplicación se ve limitada, ya que solo considera el orden jerárquico de los datos y no alguna propiedad propia de los datos, como en el caso de la media aritmética.

MEDIDAS DE POSICIÓN

Son similares a la mediana en que también subdividen una distribución de mediciones de acuerdo con la proporción de frecuencias observadas. Mientas que la mediana divide a una distribución en mitades, los cuartiles (Q) la dividen en cuartos, los deciles (D) la dividen en décimos y los puntos percentiles (P) la dividen en centésimos.
Colectivamente, cuartiles, deciles y percentiles se denominan cuantiles. Puesto que sirven para ubicar datos particulares dentro de ciertas porciones de una distribución de datos, toman el nombre de medidas de posición.

CUARTILES.- Son cada uno de los 3 valores $\boldsymbol{Q_1, Q_2, Q_3}$ que dividen a la distribución de los datos en 4 partes iguales.
Los cuartiles son un caso particular de los percentiles. Hay 3 cuartiles:
Primer cuartil: $Q_1 = P_{25}$, segundo cuartil: $Q_2 = D_5 = P_{50}$ = Mediana, tercer cuartil: $Q_3 = P_{75}$

Diagrama de caja y bigotes es una representación gráfica que ayuda a visualizar una distribución de datos: caja desde Q_1 a Q_3 (50% de los datos), y bigotes el recorrido (distancia desde valor mínimo hasta el valor máximo).



DECILES
Son cada uno de los 9 valores $D_1, D_2, D_3, D_4, D_5, D_6, D_7, D_9$ que dividen a la distribución de los datos en 10 partes iguales.
El primer decil es igual al décimo percentil $(D_1 = P_1)$, el segundo decil es igual al veinteavo percentil $(D_2 = P_{20})$, y así sucesivamente.

PERCENTILES O CENTILES
Son cada uno de los 99 valores P_1, P_2, P_3,……..P_{99} que dividen atribución de los datos en 100 partes iguales.

MODA.- La moda de un conjunto de datos es el valor que aparece con mayor frecuencia.

Propiedades

- No es afectada por valores muy altos o muy bajos.

- La moda, al igual que la mediana, no se presta para tratamientos algebraicos como la media aritmética.

- La moda puede no existir, e incluso no ser única en caso de existir.

- Cuando en un conjunto de datos hay tres o más datos diferentes con la misma frecuencia mayor, esta información a menudo no resulta útil (demasiadas modas tienden a distorsionar el significado de moda). Por lo que en estos casos se considera que el conjunto de datos no tiene moda.

Para un conjunto de datos unimodales existe la siguiente relación empírica:

Media aritmética – moda = 3 (media aritmética – mediana)

MEDIDAS DE DISPERSIÓN

Las medias de tendencia central o posición nos indican donde se sitúa un dato dentro de una distribución de datos. Las medidas de dispersión, variabilidad o variación nos indican si esos datos están próximos entre sí o sí están dispersos, es decir, nos indican cuán esparcidos se encuentran los datos. Estas medidas de dispersión nos permiten apreciar la distancia que existe entre los datos a un cierto valor central e identificar la concentración de los mismos en un cierto sector de la distribución, es decir, permiten estimar cuán dispersas están dos o más distribuciones de datos.

Estas medidas permiten evaluar la confiabilidad del valor del dato central de un conjunto de datos, siendo la media aritmética el dato central más utilizado. Cuando existe una dispersión pequeña se dice que los datos están dispersos o acumulados cercanamente respecto a un valor central, en este caso el dato central es un valor muy representativo. En el caso que la dispersión sea grande el valor central no es muy confiable. Cuando una distribución de datos tiene poca dispersión toma el nombre de distribución homogénea y si su dispersión es alta se llama heterogénea.

DESVIACIÓN MEDIA O DESVIACIÓN PROMEDIO

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

Propiedades

Guarda las mismas dimensiones que las observaciones. La suma de valores absolutos es relativamente sencilla de calcular, pero esta simplicidad tiene un inconveniente: Desde el punto de vista geométrico, la distancia que induce la desviación media en el espacio de observaciones no es la *natural* (no permite definir ángulos entre dos conjuntos de observaciones). Esto hace que sea muy engorroso trabajar con ella a la hora de hacer inferencia a la población.

Cuando mayor sea el valor de la desviación media, mayor es la dispersión de los datos. Sin embargo, no proporciona una relación matemática precisa entre su magnitud y la posición de un dato dentro de una distribución.

La desviación media al tomar los valores absolutos mide una observación sin mostrar si la misma está por encima o por debajo de la media aritmética.

VARIANZA Y DESVIACIÓN ESTÁNDAR.- La varianza es la media aritmética de los cuadrados de las desviaciones respecto a la media aritmética, es decir, es el promedio de las desviaciones de la media elevadas al cuadrado. La desviación estándar o desviación típica es la raíz de la varianza.

La varianza y la desviación estándar proporcionan una medida sobre el punto hasta el cual se dispersan las observaciones alrededor de su media aritmética.

Propiedades

- La varianza y desviación estándar (o cualquier otra medida de dispersión) indican el grado en que están dispersos los datos en una distribución. A mayor medida, mayor dispersión.

- La varianza es un número muy grande con respecto a las observaciones, por lo que con frecuencia se vuelve difícil para trabajar.

- Debido a que las desviaciones son elevadas al cuadrado y la varianza siempre se expresa en términos de los datos originales elevados al cuadrado, se obtiene unidades de medida de los datos que no tiene sentido o interpretación lógica. Por ejemplo, si se calcula la varianza de una distribución de datos medidos en metros, segundos, dólares, etc., se obtendrá una varianza mediada en metros cuadrados, segundos cuadrados, dólares cuadrados, respectivamente, unidades de medida que no tienen significado lógico respecto a los datos originales.

- Para solucionar las complicaciones que se tiene con la varianza, se halla la raíz cuadrada de la misma, es decir, se calcula la desviación estándar, la cual es un número pequeño expresado en unidades de los datos originales y que tiene un significado lógico respeto a los mismos.

A pesar de lo anterior, es difícil describir exactamente qué es lo que mide la desviación estándar. Sin embargo, hay un resultado útil, que lleva el nombre del matemático ruso Pafnuty Lvovich Chebyshev, y se aplica a todos los conjuntos de datos. Este teorema de Chebyshev establece que para todo conjunto de datos, por lo menos $1 - 1/k^2$ de las observaciones están dentro de k desviaciones estándar de la media, en donde k es cualquier número mayor que 1. Este teorema se expresa de la siguiente manera:

$$1 - \frac{1}{k^2}$$

Así por ejemplo, si se forma una distribución de datos con k =3 desviaciones estándar por debajo de la media hasta 3 desviaciones estándar por encima de la media, entonces por lo menos

$$1 - \frac{1}{3^2} = 1 - \frac{1}{9} = \frac{9-1}{9} = \frac{8}{9} = 0{,}8889 = 88{,}89\%$$

Interpretación: El 88,89% de todas las observaciones estarán dentro ± 3 desviaciones de la media.

OTRAS MEDIDAS DE DISPERSIÓN

RANGO, AMPLITUD TOTAL O RECORRIDO

Dada una serie de valores $x_1, x_2, \dots . x_n$, su recorrido es la diferencia aritmética entre el máximo y el mínimo de estos valores.

$$Re = xmáx - xmín$$

Es la medida de dispersión más sencilla y también, por tanto, la que proporciona menos información. Además, esta información puede ser errónea, pues el hecho de que no influyan más de dos valores del total de la serie puede provocar una deformación de la realidad

AMPLITUD INTERCUARTÍLICA

La amplitud intercuartílica es la distancia entre el tercer cuartil Q_3 y el primer cuartil Q_1.
Amplitud intercuartílica = tercer cuartil - primer cuartil = $Q_3 - Q_1$

RANGO SEMI-INTERCUARTIL O DESVIACIÓN CUARTÍLICA

La desviación cuartílica es la mitad de la distancia entre el tercer cuartil y el primero

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

RANGO PERCENTIL O AMPLITUD CUARTÍLICA
Cada conjunto de datos tiene 99 percentiles, que dividen el conjunto en 100 partes iguales.

La amplitud cuartílica es la distancia entre dos percentiles establecidos.

El rango percentil o amplitud cuartílica 10 a 90 es la distancia entre el 10º y 90º percentiles, definida por

$$Rango\ percentil\ 10 - 90 = P_{90} - P_{10}$$

DISPERSIÓN RELATIVA O COEFICIENTE DE VARIACIÓN

Las medidas de dispersión anteriores son todas medidas de variación absolutas. Una medida de dispersión relativa de los datos, que toma en cuenta su magnitud, está dada por el coeficiente de variación.

El Coeficiente de variación (*CV*) es una medida de la dispersión relativa de un conjunto de datos, que se obtiene dividiendo la desviación estándar del conjunto entre su media aritmética y se expresa generalmente en términos porcentuales.

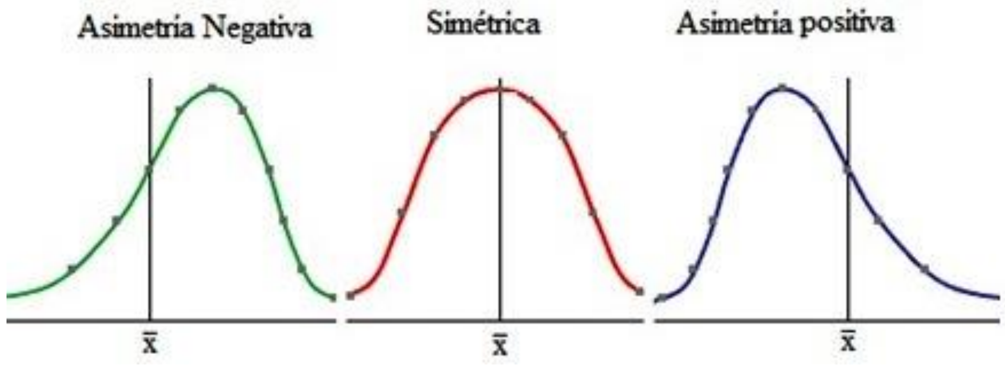
Propiedades

- Puesto que tanto la desviación estándar como la media se miden en las unidades originales, el *CV* es una medida independiente de las unidades de medición.

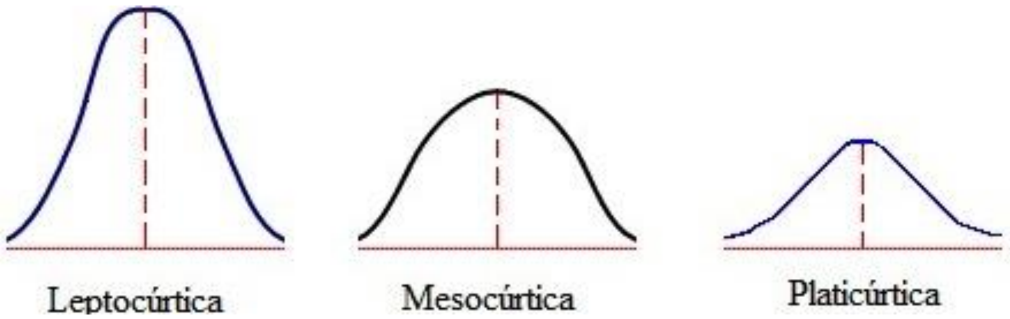
- Debido a la propiedad anterior el *CV* es la cantidad más adecuada para comparar la variabilidad de dos conjuntos de datos.

MEDIDAS DE FORMA

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



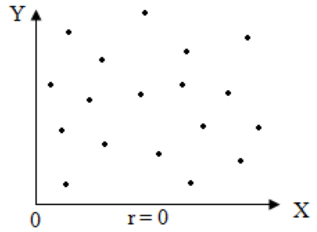
CURTOSIS O APUNTAMIENTO.- La curtosis mide el grado de agudeza o achatamiento de una distribución con relación a la distribución normal, es decir, mide cuán puntiaguda es una distribución.



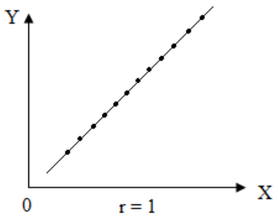
CORRELACIÓN Y REGRESIÓN
Cuando se estudian en forma conjunta dos características (variables estadísticas) de una población o muestra, se dice que estamos analizando una variable estadística bidimensional. La correlación es el grado de relación que existe entre ambas características, y la regresión es la forma de expresar matemáticamente dicha relación.

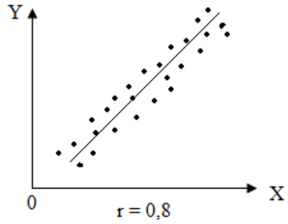
ANÁLISIS DE CORRELACIÓN.- Dado dos variables, la correlación permite hacer estimaciones del valor de una de ellas conociendo el valor de la otra variable.

Coeficientes de correlación.- Los coeficientes de correlación son medidas que indican la situación relativa de los mismos sucesos respecto a las dos variables (Correlación de Karl Pearson y por Rangos de Spearman), es decir, son la expresión numérica que nos indica el grado de relación existente entre las 2 variables y en qué medida se relacionan. Son números que varían entre los límites +1 y -1. Su magnitud indica el grado de asociación entre las variables; el valor r = 0 indica que no existe relación entre las variables; los valores ± 1 son indicadores de una correlación perfecta positiva (al crecer o decrecer X, crece o decrece Y) o negativa (Al crecer o decrecer X, decrece o crece Y).

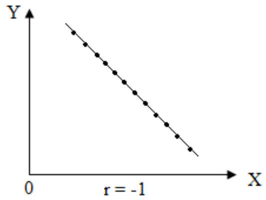


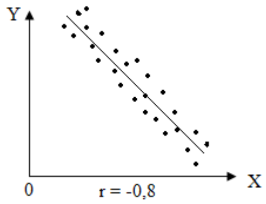
No hay correlación





Correlación Positiva





Correlación Negativa

Coeficiente de determinación

Revela qué porcentaje del cambio en Y se explica por un cambio en X. Se calcula elevando al cuadrado el coeficiente de correlación.

ANÁLISIS DE REGRESIÓN.- Los primeros y más importantes estudios al respecto se deben a los científicos Francis Galton (1822-1911) y Karl Pearson (1857-1936). Fue Galton quien utilizó por primera vez el término regresión para indicar que, aunque influida por la estatura de sus padres, la estatura de los hijos "regresaba" a la media general.

La regresión examina la relación entre dos variables, pero restringiendo una de ellas con el objeto de estudiar las variaciones de una variable cuando la otra permanece constante. En otras palabras, la regresión es un método que se emplea para predecir el valor de una variable en función de valores dados a la otra variable. En estadística la palabra predecir no se utiliza en el sentido empleado por los astrólogos, futurólogos y mentalistas, sino mas bien en un sentido lógico como es el de utilizar el conocimiento del comportamiento de una variable para obtener información sobre otra variable. Por ejemplo, puede predecirse el resultado que obtendrá un estudiante en su examen final, basados en el conocimiento de las calificaciones promedio de sus exámenes parciales, o predecir la preferencia de los estudiantes por profesiones científicas, conociendo los promedios de sus calificaciones en los estudios escolares.

En todos los casos de regresión existe una dependencia funcional entre las variables. En el caso de dos variables, siendo una de ellas (X) variable independiente y la otra (Y) la dependiente, se habla de regresión de Y sobre X; Por ejemplo, los ingenieros forestales utilizan la regresión de la altura de los árboles sobre su diámetro, lo cual significa que midiendo el diámetro (variable independiente) y reemplazando su valor en una relación definida según la clase de árbol se obtiene la altura, y aun sin necesidad de cálculos aprecian la altura utilizando gráficas de la función de dependencia, altura = función del diámetro. *Suele emplearse el principio de los mínimos cuadrados*

La recta de los mínimos cuadrados.- Se llama línea de mejor ajuste y se define como la línea que hace mínima la suma de los cuadrados de las desviaciones respecto a ella de todos los puntos que corresponden a la información recogida.
El punto de intersección entre las rectas $Y = a_0 + a_1X$ con $X = b_0 + b_1Y$ se simboliza $(\bar{X}, \bar{Y})$ y se llama **centroide** o centro de gravedad

La parábola de los mínimos cuadrados.- La parábola de mínimos cuadrados que aproxima el conjunto de puntos $(X_1, Y_1), (X_2, Y_2), (X_3, Y_3), \ldots (Y_N, Y_N)$ tiene ecuación dada por $Y = a_0 + a_1X + a_2X^2$, donde las constantes a_0, a_1 y a_2 se determinan al resolver simultáneamente el sistema de ecuaciones que se forma al multiplicar la ecuación $Y = a_0 + a_1X + a_2X^2$ por $1, X, Y$ sucesivamente, y sumando después.

$$\begin{cases} \Sigma Y = a_0N + a_1\Sigma X + a_2\Sigma X^2 \\ \Sigma XY = a_0\Sigma X + a_1\Sigma X^2 + a_2\Sigma X^3 \\ \Sigma X^2Y = a_0\Sigma X^2 + a_1\Sigma X^3 + a_2\Sigma X^4 \end{cases}$$

SERIES CRONOLÓGICAS

Las series de tiempo llamadas también series cronológicas o series históricas son un conjunto de datos numéricos que se obtienen en períodos regulares y específicos a través del tiempo, los tiempos pueden ser en años, meses, semanas, días u otra unidad adecuada al problema que se esté trabajando. Ejemplos de series de tiempo son: Ventas mensuales de un producto en una empresa, producción total anual de petróleo en Ecuador durante un cierto número años o las temperaturas anunciadas cada hora por el meteorólogo para un aeropuerto.

Matemáticamente, una serie de tiempo se define por los valores $Y_1, Y_2, Y_{3,}$ .... de una variable Y (ventas mensuales, producción total, etc.) en tiempos $t_1, t_2, t_{3,}$ .... Si se reemplaza a X por la variable tiempo, estas series se definen como distribuciones de pares ordenados (X,Y) en el plano cartesiano, siendo Y una función de X; esto se denota por:

$$Y = f(t) \rightarrow Y = f(X)$$

El principal objetivo de las series de tiempo es hacer proyecciones o pronósticos sobre una actividad futura, suponiendo estables las condiciones y variaciones registradas hasta la fecha, lo cual permite planear y tomar decisiones a corto o largo plazo. Después, con base en esa situación ideal, que supone que los factores que influyeron en la serie en el pasado lo continuarán haciendo en el futuro, se analizan las tendencias pasadas y el comportamiento de las actividades bajo la influencia de ellas; por ejemplo, en la proyección de ventas de un producto o de un servicio de una empresa se calculan los posibles precios, la reacción del consumidor, la influencia de la competencia, etc.

MOVIMIENTOS O COMPONENTES DE SERIES DE TIEMPO

El modelo clásico o de descomposición, considera que los datos de series de tiempo están compuestas de los siguientes cuatro patrones básicos:

Tendencia secular .- La tendencia secular o simplemente tendencia, son movimientos o variaciones continuas de la variable de modo uniforme y suave, por encima o por debajo, que se observan en el largo plazo durante un período de longitud prolongada. Representan el comportamiento predominante o dirección general de la serie de tiempo como ascendente o descendente. La gráfica de la tendencia suele ser una curva suave y aun una línea recta que muestra la tendencia de las variaciones. Ejemplos de tendencia secular son las ventas, exportaciones, producción y el empleo.

La siguiente gráfica muestra la tendencia de exportaciones de la Empresa D & M en período 2000-2009. Aunque los datos muestran ciertas variaciones están por encima y por debajo de la recta de tendencia, la tendencia secular es ascendente.

Movimientos estacionales.- Representa un movimiento periódico que se producen en forma similar cada año por la misma época, en correlación con los meses o con las estaciones del año y aun con determinadas fechas. Si los sucesos no se repiten anualmente, los datos deben recolectarse trimestral, mensual o incluso semanalmente. Ejemplos de movimientos estacionales son la variación de precios de ciertos productos, incremento de ventas de juguetes y disminución de ventas de útiles Navidad, incremento de ventas de flores por el día del amor y la amistad, etc.

Movimientos cíclicos.- Son variaciones hacia arriba y hacia abajo de la tendencia que se presentan cada cierto número de intervalos, en forma periódica de manera ondular a modo de oscilaciones más o menos regulares durante un período relativamente prolongado, que por lo general abarca tres o más años de duración. La producción, empleo, promedio industrial, etc. son ejemplos de este tipo de movimientos.

Movimientos irregulares o aleatorios.- Son aquellas variaciones producidas por sucesos de ocurrencia imprevisible o accidental que producen movimientos sin un patrón discernible; así por ejemplo, las exportaciones de una empresa pueden ser afectadas por sucesos inusuales no previsibles tales como huelgas, guerras, terremotos, inundaciones, etc. Estas variaciones irregulares son de corta duración y de magnitud muy variable.

MODELOS DE SERIES DE TIEMPO

Son expresiones matemáticas de relación entre los movimientos de tendencia secular (T), movimientos cíclicos (C), movimientos estacionales (E) y movimientos irregulares (I) que generan la variable Y. Hay dos modelos para la definición de Y, los cuales son:

Modelo multiplicativo.- En el que Y queda definida por el producto de las variaciones.

$$Y = T \cdot C \cdot E \cdot I$$

Modelo aditivo.- En el que Y queda definida por la suma de las variaciones.

$$Y = T + C + E + I$$

En el modelo multiplicativo, las variaciones se expresan en términos relativos o porcentuales de la tendencia, en tanto que en el modelo aditivo las variaciones se expresan como residuos en las mismas unidades originales. El modelo aditivo sufre el supuesto irreal de que los movimientos o componentes son independientes uno de otro, algo que difícilmente se da en el caso de la vida real. El modelo multiplicativo supone que los movimientos o componentes interactúan entre sí y no se mueven independientemente, por lo que este modelo es más utilizado que el aditivo. Sin embargo, el criterio fundamental que se debe seguir en el caso de una situación dada es emplear el modelo que mejor se ajuste a los datos.

MÉTODOS DE SUAVIZAMIENTO Y PRONÓSTICO.- Estos métodos eliminan las fluctuaciones aleatorias de la serie de tiempo, proporcionando datos menos distorsionados del comportamiento real de misma *(método de los promedios móviles y suavización exponencial)*

ANÁLISIS DE TENDENCIA.- Es necesario describir la tendencia ascendente o descendente a largo plazo de una serie cronológica por medio de alguna línea, y la más adecuada será la que mejor represente los datos y sea útil para desarrollar pronósticos. Para lograr la estimación de la tendencia se utilizan con más frecuencia los siguientes métodos: *Método de los mínimos cuadrados y método de los semipromedios*

REFERENCIAS BIBLIOGRÁFICAS

SUÁREZ, Mario y TAPIA, Fausto (2012), Interaprendizaje de Estadística Básica, Universidad Técnica de Norte Ibarra, Ecuador.

PROBABILIDADES

ANÁLISIS COMBINATORIO

A) FACTORIAL.- La factorial está relacionada con el cálculo del número de maneras en las que un conjunto de cosas puede arreglarse en orden.
El número de maneras en el que las n cosas pueden arreglarse en orden es:

$$n! = n(n-1)(n-2)\ldots\ldots..1$$

Donde n! se llama el factorial de n y 0! se define como 1

B) PERMUTACIONES.-En muchos casos se necesita saber el número de formas en las que un subconjunto de un grupo completo de cosas puede arreglarse en orden. Cada posible arreglo es llamado permutación. Si un orden es suficiente para construir otro subconjunto, entonces se trata de permutaciones.

El número de maneras para arreglar r objetos seleccionados a la vez de n objetos en orden, es decir, el número de permutaciones de n elementos tomados r a la vez es:

$${}_nP_r = n(n-1)(n-2)\ldots(n-r+1) = \frac{n!}{(n-r)!}$$

C) COMBINACIONES

En muchos situaciones no interesa el orden de los resultados, sino sólo el número de maneras en las que r objetos pueden seleccionarse a partir de n cosas, sin consideración de orden. Si dos subconjntos se consideran iguales debido a que simplemente se han reordenado los mismos elementos, entonces se trata de combinaciones.

El número de maneras para arreglar r objetos seleccionados a la vez de n objetos, sin considerar el orden, es decir, el número de combinaciones de n elementos tomados r a la vez es:

$$_nC_r = \binom{n}{r} = \frac{n(n-1)(n-2)\ldots(n-r+1)}{r!} = \frac{n!}{r!\,(n-r)!}$$

CONCEPTOS BÁSICOS

A) EXPERIMENTO.- Es toda acción sobre la cual vamos a realizar una medición u observación, es decir cualquier proceso que genera un resultado definido.

B) EXPERIMENTO ALEATORIO.- Es toda actividad cuyos resultados no se determinan con certeza. Ejemplo: lanzar una moneda al aire. No podemos determinar con toda certeza ¿cuál será el resultado al lanzar una moneda al aire?, por lo tanto constituye un experimento aleatorio.

C) ESPACIO MUESTRAL (S).- Es un conjunto de *todos los resultados posibles* que se pueden obtener al realizar un experimento aleatorio. Ejemplo: sea el experimento E: lanzar un dado y el espacio muestral correspondiente a este experimento es: S = $\{1, 2, 3, 4, 5, 6\}$.

D) PUNTO MUESTRAL.- Es un elemento del espacio muestral de cualquier experimento dado.

E) EVENTO O SUCESO.- Es todo subconjunto de un espacio muestral. Se denotan con letras mayúsculas: A, B, etc. Los resultados que forman parte de este evento generalmente se conocen como *"resultados favorables"*. Cada vez que se observa un resultado favorable, se dice que "ocurrió" un evento. Ejemplo: Sea el experimento E: lanzar un dado. Un posible evento podría ser que salga número par. Definimos el evento de la siguiente manera: A = sale número par = $\{2, 4, 6\}$, resultados favorables n(E) = 3

Los eventos pueden ser:

i) Evento cierto.- Un evento es cierto o seguro si se realiza siempre. Ejemplo: Al introducirnos en el mar, en condiciones normales, es seguro que nos mojaremos.

ii) Evento imposible.- Un evento es imposible si nunca se realiza. Al lanzar un dado una sola vez, es imposible que salga un 10

iii) Evento probable o aleatorio.- Un evento es aleatorio si no se puede precisar de antemano el resultado. Ejemplo: ¿Al lanzar un dado, saldrá el número 3?

F) PROBABILIDAD.- Es el conjunto de posibilidades de que un evento ocurra o no en un momento y tiempo determinado. Dichos eventos pueden ser medibles a través de una escala de 0 a 1, donde el evento que no pueda ocurrir tiene una probabilidad de 0 (evento imposible) y un evento que ocurra con certeza es de 1 (evento cierto).

La probabilidad de que ocurra un evento, siendo ésta una medida de la posibilidad de que un suceso ocurra favorablemente, se determina principalmente de dos formas: empíricamente (de manera experimental) o teóricamente (de forma matemática).

i) Probabilidad empírica.- Si E es un evento que puede ocurrir cuando se realiza un experimento, entonces la probabilidad empírica del evento E, que a veces se le denomina *definición de frecuencia relativa de la probabilidad*, está dada por la siguiente fórmula:

$$P(E) = \frac{número\ de\ veces\ que\ ocurre\ el\ evento\ E}{número\ de\ veces\ que\ se\ realizó\ el\ experimento}$$

Nota: P(E), se lee probabilidad del evento E

ii) Probabilidad teórica.- Si todos los resultados en un espacio muestral S finito son igualmente probables, y E es un evento en ese espacio muestral, entonces la probabilidad teórica del evento E está dada por la siguiente fórmula, que a veces se le denomina la *definición clásica de la probabilidad*, expuesta por Pierre Laplace en su famosa Teoría analítica de la probabilidad publicada en 1812:

$$P(E) = \frac{número\ de\ resultados\ favorables}{número\ total\ de\ posibles\ resultados} = \frac{n(E)}{n(S)}$$

G) POSIBILIDADES.- Las posibilidades comparan el número de resultados favorables con el número de resultados desfavorables. Si todos los resultados de un espacio muestral son igualmente probables, y un número n de ellos son favorables al evento E, y los restantes m son desfavorables a E, entonces las *posibilidades a favor* de E sonde de n(E) a m(E), y las *posibilidades en contra* de E son de m(E) a n(E)

Ejemplos ilustrativos: Mathías se le prometió comprar 6 libros, tres de los cuales son de Matemática. Si tiene las mismas oportunidades de obtener cualquiera de los 6 libros, determinar las posibilidades de que le compren uno de Matemática.

Solución:

Número de resultados favorables = n(E) = 3
Número de resultados desfavorables = m(E) = 3
Posibilidades a favor son n(E) a m(E), entonces,
Posibilidades a favor = 3 a 3, y simplificando 1 a 1.

Nota: A las posibilidades de 1 a 1 se les conoce como "igualdad de posibilidades" o "posibilidades de 50-50"

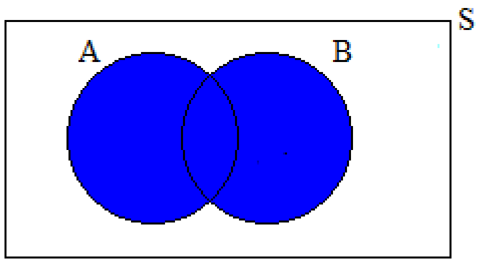
REGLAS DE LA PROBABILIDAD

A) REGLA DE LA ADICIÓN DE PROBABILIDADES

i) REGLA GENERAL PARA EVENTOS NO MUTUAMENTE EXCLUYENTES

Si A y B son dos eventos no mutuamente excluyentes (eventos intersecantes), es decir, de modo que ocurra A o bien B o ambos a la vez (al mismo tiempo), entonces se aplica la siguiente regla para calcular dicha probabilidad:

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



A U B

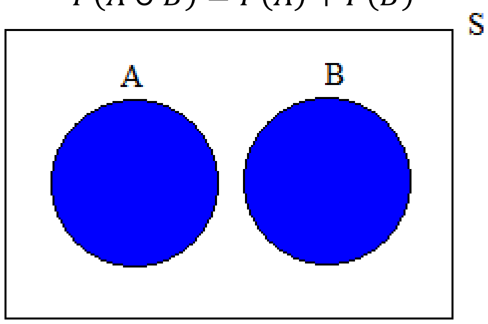
En donde:

El conectivo lógico "*o*" corresponde a la "*unión*" en la teoría de conjuntos (o =∪)
El conectivo "*y*" corresponde a la "*intersección*" en la teoría de conjuntos (y =∩)
El espacio muestral (S) corresponde al conjunto universo en la teoría de conjuntos

ii) REGLA PARTICULAR O ESPECIAL PARA EVENTOS MUTUAMENTE EXCLUYENTES

Si A y B son dos eventos mutuamente excluyentes (eventos no intersecantes), es decir, si la ocurrencia de cualquiera de ellos excluye la del otro, no pueden ocurrir a la vez, o cuando no tienen ningún punto muestral en común ($A \cap B = \emptyset$), entonces se aplica la siguiente regla para calcular dicha probabilidad:

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



A U B

En donde:

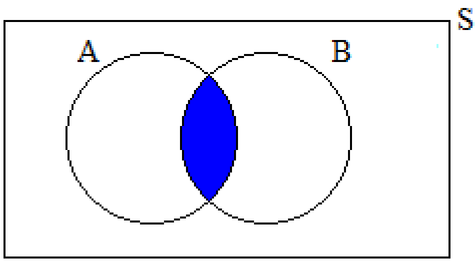
El conectivo lógico "*o*" corresponde a la "*unión*" en la teoría de conjuntos (o =∪)
El espacio muestral (S) corresponde al conjunto universo en la teoría de conjuntos

B) REGLA DE LA MULTIPLICACIÓN DE PROBABILIDADES

i) REGLA GENERAL PARA EVENTOS DEPENDIENTES

Si A y B son dos eventos dependientes, es decir, si la ocurrencia de A afecta la probabilidad de ocurrencia de B, entonces, dicha probabilidad de calcula empleando la siguiente regla:

$$P(AyB) = P(A) \cdot P(B/A)$$
$$o$$
$$P(A \cap B) = P(A) \cdot P(B/A)$$



A ∩ B

En donde:
El conectivo "*y*" corresponde a la "*intersección*" en la teoría de conjuntos (y =∩)
El espacio muestral (S) corresponde al conjunto universo en la teoría de conjuntos
$P(B/A) = Probabilidad\ condicional\ de\ B, dado\ A$

Nota:
La probabilidad del evento B, calculada bajo la suposición de que el evento A ha ocurrido, se denomina *probabilidad condicional de B, dado A*, y se denota por P (B/A).
Si se desea obtener una fórmula para calcular la probabilidad condicional se despeja de la fórmula general de la multiplicación$P(A \cap B) = P(A) \cdot P(B/A)$, obteniéndose:

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

La probabilidad condicional de A dado B se denota por P(A/B) y se calcula empleando la siguiente fórmula:

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

ii) REGLA PARTICULAR O ESPECIAL PARA EVENTOS INDEPENDIENTES

Si A y B son dos eventos independientes, es decir, si el conocimiento de la incidencia de uno de ellos no tiene efecto en la probabilidad de ocurrencia del otro, entonces, para calcular la probabilidad de dichos eventos se aplica la siguiente regla:

$$P(AyB) = P(A) \cdot P(B)$$
$$o$$
$$P(A \cap B) = P(A) \cdot P(B)$$

Nota: Dos eventos A y B son independientes si la ocurrencia de uno de ellos no afecta la probabilidad de ocurrencia del otro, esto es, si $P(B/A) = P(B)$

1.4) PROBABILIDAD TOTAL Y TEOREMA DE BAYES

A) PROBABILIDAD TOTAL
Sea $A_1, A_2, \ldots\ldots A_n$ un sistema completo de eventos tales que la probabilidad de cada uno de ellos es distinto de cero, y sea B un evento cualquiera del que se conocen las probabilidades condicionales $P(B/A_i)$, entonces, la probabilidad del evento B, llamada probabilidad total, se calcula empleando la siguiente fórmula:

$$P(B) = P(A_1) \cdot P(B/A_1) + P(A_2) \cdot P(B/A_2) + \cdots P(A_n) \cdot P(B/A_n)$$

B) TEOREMA DE BAYES

El teorema de Bayes se utiliza para revisar probabilidades previamente calculadas cuando se posee nueva información. Desarrollado por el reverendo Thomas Bayes en el siglo XVII, el teorema de Bayes es una extensión de lo que ha aprendido hasta ahora acerca de la probabilidad condicional.

Comúnmente se inicia un análisis de probabilidades con una asignación inicial, probabilidad a priori. Cuando se tiene alguna información adicional se procede a calcular las probabilidades revisadas o a posteriori. El teorema de Bayes permite calcular las probabilidades a posteriori y es:

$$P(A_i/B) = \frac{P(A_i) \cdot P(B/A_i)}{P(B)}$$

Donde:
$P(A_i)$ = Probabilidad a priori
$P(B/A_i)$ = Probabilidad condicional
$P(B)$ = Probabilidad Total
$P(A_i/B)$ = Probabilidad aposteriori

2.1) DISTRIBUCIONES DISCRETAS

A) INTRODUCCIÓN

Una distribución de probabilidad es una representación de todos los resultados posibles de algún experimento y de la probabilidad relacionada con cada uno.
Una distribución de probabilidad es discreta cuando los resultados posibles del experimento son obtenidos de variables aleatorias discretas, es decir, de variables que sólo puede tomar ciertos valores, con frecuencia números enteros, y que resultan principalmente del proceso de conteo.

Ejemplos de variables aleatorias discretas son:
Número de caras al lanzar una moneda
El resultado del lanzamiento de un dado
Número de hijos de una familia
Número de estudiantes de una universidad

Ejemplo ilustrativo
Sea el experimento aleatorio de lanzar 2 monedas al aire. Determinar la distribución de probabilidades del número de caras.

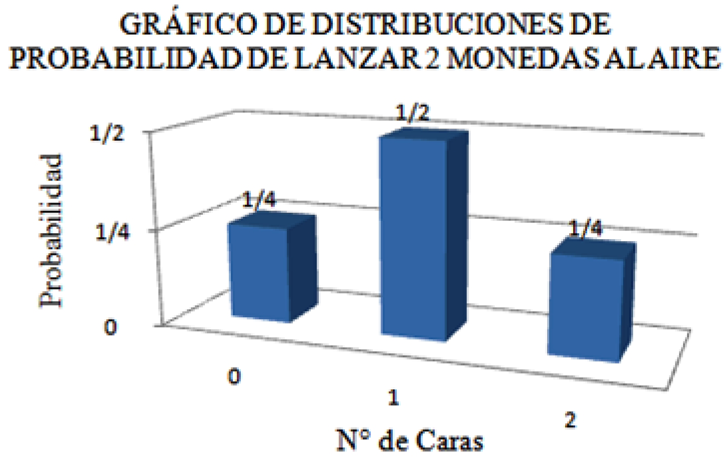
Solución:

El espacio muestral es S = {CC, CS, SC, SS}
La probabilidad de cada punto muestral es de 1/4, es decir, P(CC) = P(CS) = P(SC) = P(SS) = 1/4

La distribución de probabilidades del número de caras se presenta en la siguiente tabla:

Resultados (N° de Caras)	Probabilidad
0	1/4 = 0,25 = 25%
1	2/4 = 0,50 = 50%
2	1/4 = 0,25 = 25%

El gráfico de distribuciones de probabilidad en 3D elaborado en Excel se muestra en la siguiente figura:



Interpretación:
La probabilidad de obtener 0 caras al lanzar 2 monedas al aire es de 1/4 = 0,25 = 25%
La probabilidad de obtener una cara al lanzar 2 monedas al aire es de 2/4 = 0,5 = 50%
La probabilidad de obtener 2 caras al lanzar 2 monedas al aire es de 1/4 = 0,25 = 25%

B) LA MEDIA Y LA VARIANZA DE LAS DISTRIBUCIONES DISCRETAS

i) Media

La media llamada también valor esperado, esperanza matemática o simplemente esperanza de una distribución de probabilidad discreta es la media aritmética ponderada de todos los resultados posibles en los cuales los pesos son las probabilidades respectivas de tales resultados. Se halla multiplicando cada resultado posible por su probabilidad y sumando los resultados. Se expresa mediante la siguiente fórmula:

$$\mu = \mathrm{E(X)} = \Sigma(x_i \cdot P(x_i))$$

Donde:
$\mu = \mathrm{E(X)}$ = Media, Valor Esperado, Esperanza Matemática o simplemente Esperanza
x_i = Posible resultado
$P(x_i)$ = Probabilidad del posible resultado

ii) Varianza

La varianza es el promedio de las desviaciones al cuadrado con respecto a la media. La varianza mide la dispersión de los resultados alrededor de la media y se halla calculando las diferencias entre cada uno de los resultados y su media, luego tales diferencias se elevan al cuadrado y se multiplican por sus respectivas probabilidades, y finalmente se suman los resultados. Se expresa mediante la siguiente fórmula:

$$\sigma^2 = \Sigma[(x_i - \mu)^2 \cdot P(x_i)]$$

Nota: La varianza se expresa en unidades al cuadrado, por lo que es necesario calcular la desviación estándar que se expresa en las mismas unidades que la variable aleatoria y que por lo tanto tiene una interpretación más lógica de la dispersión de los resultados alrededor de la media. La desviación estándar se calcula así: $\sigma = \sqrt{\sigma^2}$

Ejemplo ilustrativo:
Hallar la esperanza matemática, la varianza y la desviación estándar del número de caras al lanzar tres monedas al aire.

Solución:

El espacio muestral es S = {CCC, CCS, CSC, SCC, CSS, SCS, SSC, SSS}
La probabilidad de cada punto muestral es de 1/8

Se elabora las distribuciones de probabilidad y se realiza los cálculos respectivos. Estos resultados se presentan en la siguiente tabla:

x_i	$P(x_i)$	$x_i \cdot P(x_i)$	$(x_i - \mu)^2 \cdot P(x_i)$
0	1/8	0·1/8 = 0	$(0\text{-}1{,}5)^2 \cdot 1/8 = 0{,}281$
1	3/8	1·3/8 = 3/8	$(1\text{-}1{,}5)^2 \cdot 3/8 = 0{,}094$
2	3/8	2·3/8 = 3/4	$(2\text{-}1{,}5)^2 \cdot 3/8 = 0{,}094$
3	1/8	3·1/8 = 3/8	$(3\text{-}1{,}5)^2 \cdot 1/8 = 0{,}281$
Total	1	1,5	0,750

Observando la tabla se tiene:
$\mu = E(X) = 1{,}5$; $\sigma^2 = 0{,}75$

Y calculando la desviación estándar se obtiene:
$\sigma = \sqrt{\sigma^2} = \sqrt{0{,}75} = 0{,}866$

Interpretación:
El valor de $\mu = E(X) = 1{,}5$ significa que si se promedian los resultados del lanzamiento de las tres monedas (teóricamente, un número infinito de lanzamientos), se obtendrá 1,5.
Los valores de $\sigma^2 = 0{,}75$ y $\sigma = 0{,}866$ miden la dispersión de los resultados de lanzar las tres monedas alrededor de su media.

C) DISTRIBUCIÓN BINOMIAL

i) Definición:
Cuando se dispone de una expresión matemática, es factible calcular la probabilidad de ocurrencia exacta correspondiente a cualquier resultado específico para la variable aleatoria.

La *distribución de probabilidad binomial* es uno de los modelos matemáticos (expresión matemática para representar una variable) que se utiliza cuando la variable aleatoria discreta es el número de éxitos en una muestra compuesta por n observaciones.

ii) Propiedades:

- La muestra se compone de un número fijo de observaciones n

- Cada observación se clasifica en una de dos categorías, *mutuamente excluyentes* (los eventos no pueden ocurrir de manera simultánea. Ejemplo: Una persona no puede ser de ambos sexos) y *colectivamente exhaustivos* (uno de los eventos debe ocurrir. Ejemplo: Al lanzar una moneda, si no ocurre cruz, entonces ocurre cara). A estas categorías se las denomina éxito y fracaso.

- La probabilidad de que una observación se clasifique como *éxito, p,* es constante de una observación o otra. De la misma forma, la probabilidad de que una observación se clasifique como *fracaso, 1-p,* es constante en todas las observaciones.

- La variable aleatoria binomial tiene un rango de 0 a n

iii) Ecuación:

$$P(X) = \frac{n!}{X!\,(n-X)!} \cdot p^X \cdot (1-p)^{n-X}$$

Donde:
$P(X)$ =Probabilidad de X éxitos, dadas n y p
n = Número de observaciones
p = Probabilidad de éxitos
$1-p$ = Probabilidad de fracasos
X = Número de éxitos en la muestra (X = 0, 1, 2, 3, 4,……… n)

iv) Media de la distribución binomial
La media μ de la distribución binomial es igual a la multiplicación del tamaño n de la muestra por la probabilidad de éxito p

$$\mu = np$$

v) Desviación estándar de la distribución binomial

$$\sigma = \sqrt{\sigma^2} = \sqrt{np(1-p)}$$

D) DISTRIBUCIÓN DE POISSON

i) Introducción.- Muchos estudios se basan en el conteo de las veces que se presenta un evento dentro de *un* área de oportunidad dada. El *área de oportunidad* es una unidad continua o intervalo de tiempo o espacio (volumen o área) en donde se puede presentar más de un evento. Algunos ejemplos serían los defectos en la superficie de un refrigerador, el número fallas de la red en un día, o el número de pulgas que tiene un perro. Cuando se tiene un área de oportunidad como éstas, se utiliza la *distribución de Poisson* para calcular las probabilidades si:

- Le interesa contar las veces que se presenta un evento en particular dentro de un área de oportunidad determinada. El área de oportunidad se define por tiempo, extensión, área, volumen, etc.
- La probabilidad de que un evento se presente en un área de oportunidad dada es igual para todas las áreas de oportunidad.
- El número de eventos que ocurren en un área de oportunidad es independiente del número de eventos que se presentan en cualquier otra área de oportunidad.
- La probabilidad de que dos o más eventos se presenten en un área de oportunidad tiende a cero conforme esa área se vuelve menor.

ii) Fórmula.- La distribución de Poisson tiene un parámetro, llamado λ (letra griega lambda minúscula), que es la media o el número esperado de eventos por unidad. La varianza de la distribución de Poisson también es igual a λ, y su desviación estándar es igual a $\sqrt{\lambda}$. El número de eventos X de la variable aleatoria de Poisson fluctúa desde 0 hasta infinito.

$$P(X) = \frac{e^{-\lambda} \cdot \lambda^X}{X!}$$

Donde:
$P(X) =$ Probabilidad de X eventos en un área de oportunidad
$\lambda =$ Número de eventos esperados
$X =$ Número de eventos
$e =$ Constante matemática base de los logaritmos naturales aproximadamente igual a 2718281828….

E) DISTRIBUCIÓN HIPERGEOMÉTRICA

i) Definición

La distribución binomial es apropiada sólo si la probabilidad de un éxito permanece constante. Esto ocurre si el muestreo se realiza con reemplazo en una población grande. Sin embrago, si la población es pequeña y ocurre sin reemplazo, la probabilidad de éxito variará, y la distribución hipergeométrica es que se utiliza.

ii) Fórmula
Se calcula empleando la siguiente fórmula:

$$P(X) = \frac{C_X^r \cdot C_{n-X}^{N-r}}{C_n^N}$$

Donde:
C = combinación
N = tamaño de la población
r = número de éxitos en la población
n = tamaño de la muestra
X = número de éxitos en la muestra

Notas:
- Si se selecciona una muestra sin reemplazo de una población grande conocida y contiene una proporción relativamente grande de la población, de manera que la probabilidad de éxito varía de una selección a la siguiente, debe utilizarse la distribución hipergeométrica.

- Cuando tamaño de la población (N) es muy grande, la distribución hipergeométrica tiende aproximarse a la binomial.

Ejemplo ilustrativo
Si se extraen juntas al azar 3 bolas de una urna que contiene 6 bolas rojas y 4 blancas. ¿Cuál es la probabilidad de que sean extraídas 2 bolas rojas?.

Solución:

Los datos son: N =10; r = 6; n = 3 y X= 2

Aplicando la fórmula se obtiene:

$$P(X) = \frac{C_X^r \cdot C_{n-X}^{N-r}}{C_X^N}$$

$$P(X=2) = \frac{C_2^6 \cdot C_{3-2}^{10-6}}{C_3^{10}} = \frac{C_2^6 \cdot C_1^4}{C_3^{10}} = \frac{\frac{6!}{2!\,(6-2)!} \cdot \frac{4!}{1!\,(4-1)!}}{\frac{10!}{3!\,(10-3)!}} = \frac{\frac{6!}{2!\,4!} \cdot \frac{4!}{1!\,3!}}{\frac{10!}{3!\,7!}} = \frac{15 \cdot 4}{120} = 0{,}5$$

2.2 DISTRIBUCIONES CONTINUAS

A) INTRODUCCIÓN
Una distribución de probabilidad es continua cuando los resultados posibles del experimento son obtenidos de variables aleatorias continuas, es decir, de variables cuantitativas que pueden tomar cualquier valor, y que resultan principalmente del proceso de medición.
Ejemplos de variables aleatorias continuas son:
La estatura de un grupo de personas
El tiempo dedicado a estudiar
La temperatura en una ciudad

B) DISTRIBUCIÓN EXPONENCIAL

i) Definición
La distribución de Poisson calcula el número de eventos sobre alguna área de oportunidad (intervalo de tiempo o espacio), la distribución exponencial mide el paso del tiempo entre tales eventos. Si el número de eventos tiene una distribución de Poisson, el lapso entre los eventos estará distribuido exponencialmente.

ii) Fórmula
La probabilidad de que el lapso de tiempo sea menor que o igual a cierta cantidad x es:

$$P(X \leq x) = 1 - e^{-\lambda \cdot t}$$

Donde:
$t =$Lapso de tiempo
$e =$ Base del logaritmo natural aproximadamente igual a 2,718281828
$\lambda =$Tasa promedio de ocurrencia

C) DISTRIBUCIÓN UNIFORME

i) Definición
Es una distribución en el intervalo $[a, b]$ en la cual las probabilidades son las mismas para todos los posibles resultados, desde el mínimo de **a** hasta el máximo de **b**. El experimento de lanzar un dado es un ejemplo que cumple la distribución uniforme, ya que todos los 6 resultados posibles tienen 1/6 de probabilidad de ocurrencia.

ii) Función de densidad de una distribución uniforme (altura de cada rectángulo en la gráfica anterior) es:

$$\mathrm{f(X) = Altura} = \frac{1}{\mathrm{b - a}}$$

Donde:
a = mínimo valor de la distribución
b = máximo valor de la distribución
b – a = Rango de la distribución

iii) La media, valor medio esperado o esperanza matemática de una distribución uniforme se calcula empleando la siguiente fórmula:

$$\mathrm{E(X)} = \mu = \frac{\mathrm{a + b}}{2}$$

iv) La varianza de una distribución uniforme se calcula empleando la siguiente fórmula:

$$\sigma^2 = \frac{(\mathrm{b - a})^2}{12}$$

De donde la desviación estándar es $\sigma = \sqrt{\sigma^2}$

v) La probabilidad de que una observación caiga entre dos valores se calcula de la siguiente manera:

$$P(X_1 \leq X \leq X_2) = \frac{X_2 - X_1}{b - a}$$

D) DISTRIBUCIÓN NORMAL

i) Reseña histórica

Abrahan De Moivre (1733) fue el primero en obtener la ecuación matemática de la curva normal. Kart Friedrich Gauss y Márquez De Laplece (principios del siglo diecinueve) desarrollaron más ampliamente los conceptos de la curva. La curva normal también es llamada curva de error, curva de campana, curva de Gauss, distribución gaussiana o curva de De Moivre.
Su altura máxima se encuentra en la media aritmética, es decir su ordenada máxima corresponde a una abscisa igual a la media aritmética. La asimetría de la curva normal es nula y por su grado de apuntamiento o curtosis se clasifica en mesocúrtica.

ii) Ecuación

Su ecuación matemática de la función de densidad es:

$$Y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$$

Donde:
$\sigma = desviación\ estándar$
$\sigma^2 = varianza$
$\pi = 3{,}141592654 \ldots\ldots constante\ matemática$
$e = 2{,}7182818 \ldots\ldots constante\ matemática$
X = valor en el eje horizontal
Y = altura de la curva para cualquier valor de x
$\mu = media\ aritmética$

Cuando se expresa la variable x en unidades estándar (fórmula de estandarización)

$$Z = \frac{X - \mu}{\sigma}$$

La ecuación anterior es reemplazada por la llamada forma canónica, la cual es

$$Y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}Z^2}$$

Nota: No existe una única distribución normal, sino una familia de distribuciones con una forma común, diferenciadas por los valores de su media y su varianza. De entre todas ellas, la más utilizada es la **distribución normal estándar**, que corresponde a una distribución con una media aritmética de 0 y una desviación típica de 1.

Nota: Los ejemplos ilustrativos resueltos y el detalle de cada una de las fórmulas de cada tema se los puede obtener de las siguientes referencias bibliográficas

REFERENCIAS BIBLIOGRÁFICAS

SUÁREZ, Mario (2012), Interaprendizaje de Probabilidades y Estadística Inferencial con Excel, Winstats y Graph, Ibarra, Ecuador.

SUÁREZ, Mario, y TAPIA, Fausto (2012), Interaprendizaje de Estadística Básica, Universidad Técnica de Norte Ibarra, Ecuador.

SUÁREZ Mario, (2011), Análisis Combinatorio, http://repositorio.utn.edu.ec/handle/123456789/761

SUÁREZ, Mario, (2011), Distribución Binomial, http://repositorio.utn.edu.ec/handle/123456789/771

SUÁREZ Mario, (2011), Distribución de frecuencias para datos agrupados en intervalos, www.monografias.com/trabajos87/.

SUÁREZ, Mario, (2011), Cálculo del tamaño de la muestra,www.monografias.com/trabajos87/

SUÁREZ, Mario, (2011), Gráficos estadísticos básicos,www.monografias.com/trabajos88/

SUÁREZ, Mario, (2011), Guía didáctica para el interaprendizaje de medidas de tendencia central, www.monografias.com/trabajos85/

SUÁREZ, Mario, (2011), Media aritmética, www.monografias.com/trabajos85/

SUÁREZ, Mario, (2011), Ejemplos ilustrativos resueltos de la Moda, www.monografias.com/trabajos85/

SUÁREZ, Mario, (2011), La mediana para datos no agrupados y agrupados, www.monografias.com/trabajos87/

SUÁREZ, Mario, (2011), Medidas de posición, www.monografias.com/trabajos87/

SUÁREZ, Mario, (2011), Medidas de dispersión, www.monografias.com/trabajos89/

SUÁREZ, Mario, (2011), Medidas de forma: asimetría y curtosis, www.monografias.com/trabajos87/

SUÁREZ, Mario, (2011), Coeficiente de correlación de Pearson para datos agrupados en intervalos, www.monografias.com/trabajos86/

SUÁREZ, Mario, (2011), Coeficiente de Correlación por Rangos de Spearman, www.monografias.com/trabajos85/

SUÁREZ, Mario, (2011), Dispersión relativa o coeficiente de variación, www.monografias.com/trabajos88/

SUÁREZ, Mario, (2011), La recta de los mínimos cuadrados, www.monografias.com/trabajos85/

SUÁREZ, Mario (2011), Análisis de regresión mediante la parábola de los mínimos cuadrados, www.monografias.com/trabajos86/

SUÁREZ, Mario (2011), Regresión potencial mediante el método de los mínimos cuadrados, www.monografias.com/trabajos89/

SUÁREZ, Mario (2011), Regresión exponencial mediante el método de los mínimos cuadrados, www.monografias.com/trabajos89/

SUÁREZ, Mario, (2011), Análisis de tendencia para series de tiempo, www.monografias.com/trabajos87/

SUÁREZ Mario, (2011), Métodos de suavizamiento y pronóstico para series de tiempo, www.monografias.com/trabajos87/

SUÁREZ, Mario, (2012), Interaprendizaje de Probabilidades y Estadística Inferencial con Excel, Winstats y Graph. http://repositorio.utn.edu.ec/handle/123456789/940

SUÁREZ, Mario, (2011), Cálculo del tamaño de la muestra, http://repositorio.utn.edu.ec/handle/123456789/765

SUÁREZ, Mario, (2012), Gráficas de control de la calidad para variables, www.monografias.com/trabajos91

SUÁREZ, Mario, (2011), Distribución Binomial, www.monografias.com/trabajos85/

SUÁREZ, Mario, (2011), Introducción a las distribuciones de probabilidad discretas, www.monografias.com/trabajos85/

SUÁREZ, Mario, (2011), Distribución de probabilidad uniforme, www.monografias.com/trabajos86/

SUÁREZ, Mario, (2011), Cálculo del tamaño de la muestra, www.monografias.com/trabajos87/

SUÁREZ, Mario, (2011), Probabilidad Teórica, www.monografias.com/trabajos88/

SUÁREZ, Mario, (2011), Regla General de la Adición de Probabilidades para Eventos No Mutuamente Excluyentes, www.monografias.com/trabajos88/

SUÁREZ, Mario, (2011), Regla General y Particular de la Multiplicación de Probabilidades, www.monografias.com/trabajos89/

SUÁREZ, Mario (2011), Regla Particular de la Adición de Probabilidades para Eventos Mutuamente Excluyentes, www.monografias.com/trabajos89/

SUÁREZ Mario, (2011), Análisis Combinatorio, www.monografias.com/trabajos89/

SUÁREZ, Mario, (2011), Ejercicios de la Distribución Binomial resueltos con Excel, www.monografias.com/trabajos89/

SUÁREZ, Mario, (2011), Distribución Normal con Excel, www.monografias.com/trabajos89/

SUÁREZ, Mario, (2011), Ejercicios resueltos de prueba de hipótesis, www.monografias.com/trabajos89/

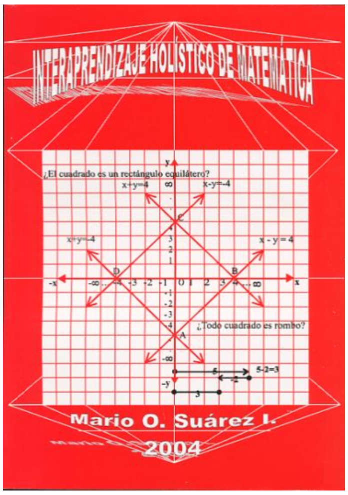
SUÁREZ, Mario, (2011), Probabilidad Total y Teorema de Bayes, www.monografias.com/trabajos89/

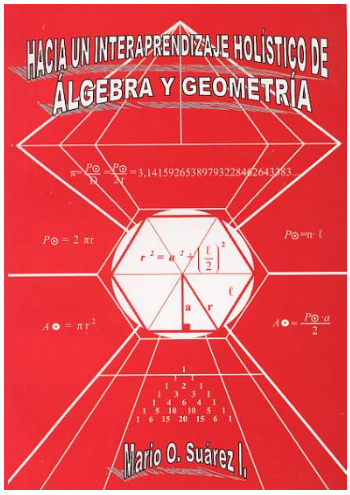
SUÁREZ, Mario, (2012), Gráficas de control de la calidad para variables, www.monografias.com/trabajos91/

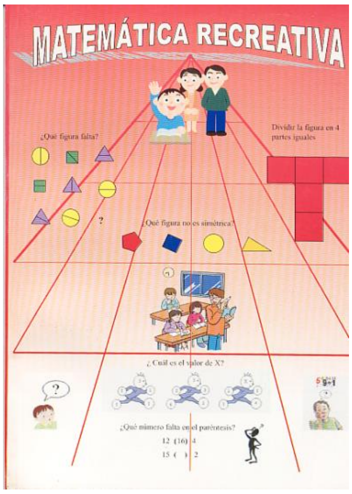
SUÁREZ, Mario, (2012), Análisis de varianza, www.monografias.com/trabajos91/

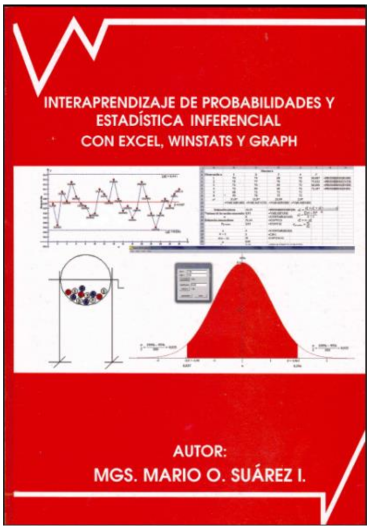
OTROS TRABAJOS

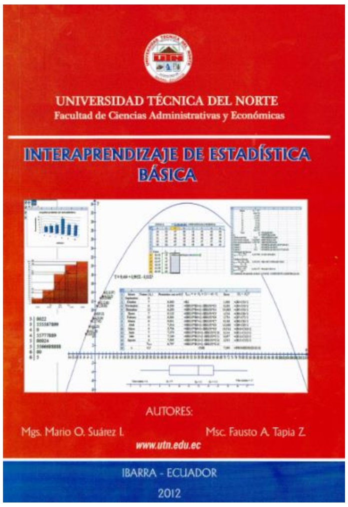
Libros











Obras artísticas inéditas

Instituto Ecuatoriano de la Propiedad Intelectual

IEPI

Dirección Nacional de Derechos de Autor
y Derechos Conexos

Certificado No. QUI-039247

Trámite N° 001501

La Dirección Nacional de Derecho de Autor y Derechos Conexos, en atención a la solicitud presentada el 08 de agosto del año 2012, **EXPIDE** el certificado de registro:

AUTOR(es):	SUÁREZ IBUJES, MARIO ORLANDO
TITULAR(es):	SUÁREZ IBUJES, MARIO ORLANDO
CLASE DE OBRA:	ARTÍSTICA (Inédita)
TÍTULO DE LA(s) OBRA(s):	POLIPRISMA 9.0 (Rompecabezas tridimensional bicolor de 9 partes).

Quito, a 08 de agosto del año 2012

Lcda. Pilar Freile Murillo
Experta en Registro 1

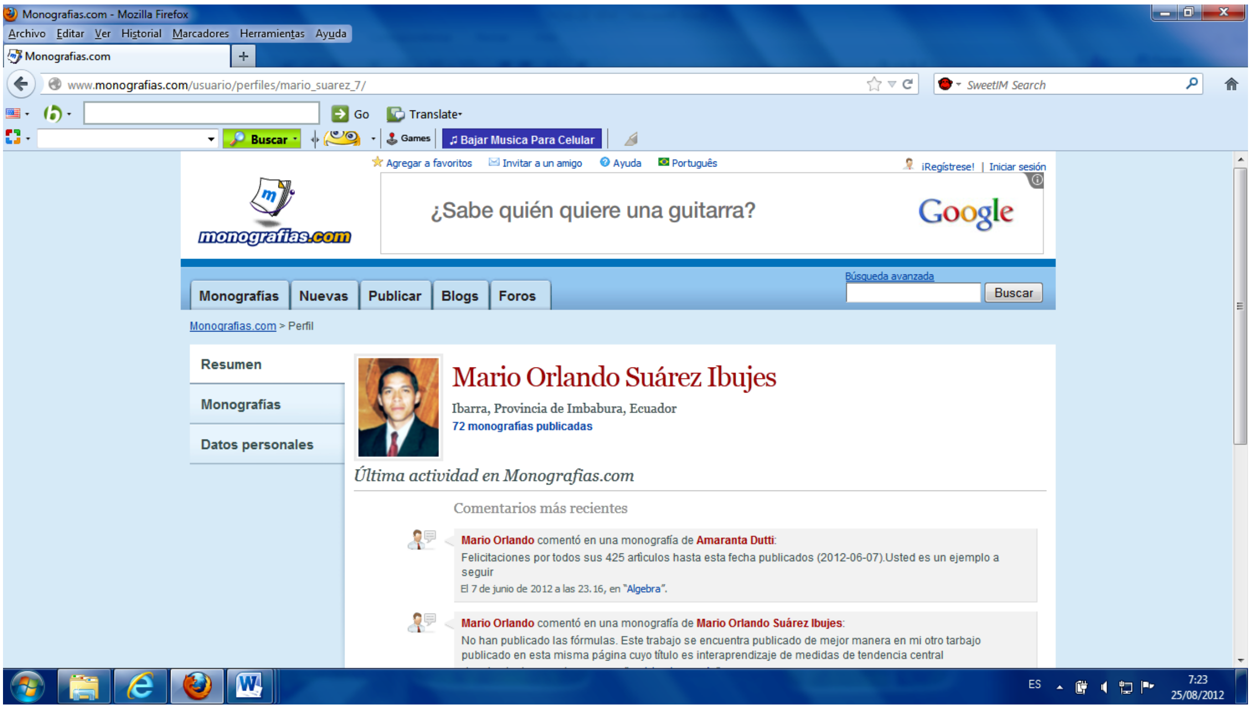
Delegada del Director Nacional de Derecho de Autor y Derechos Conexos,
mediante Resolución No.003-2012 DNDAyDC-IEPI.

El presente certificado no prejuzga sobre la originalidad de lo presentado para el registro, o su carácter literario, artístico o científico, ni acerca de la autoría o titularidad de los derechos por parte de quien solicita la inscripción. Solamente da fe del hecho de su declaración y de la identidad del solicitante.

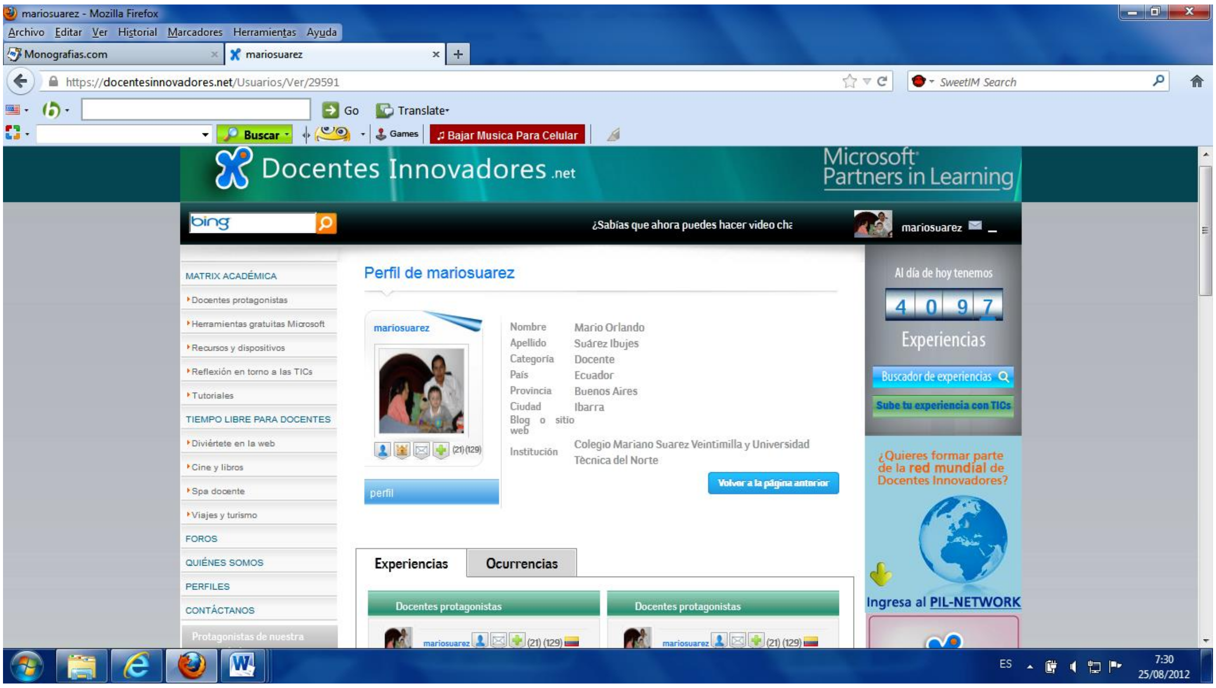
PFM.

Artículos en internet

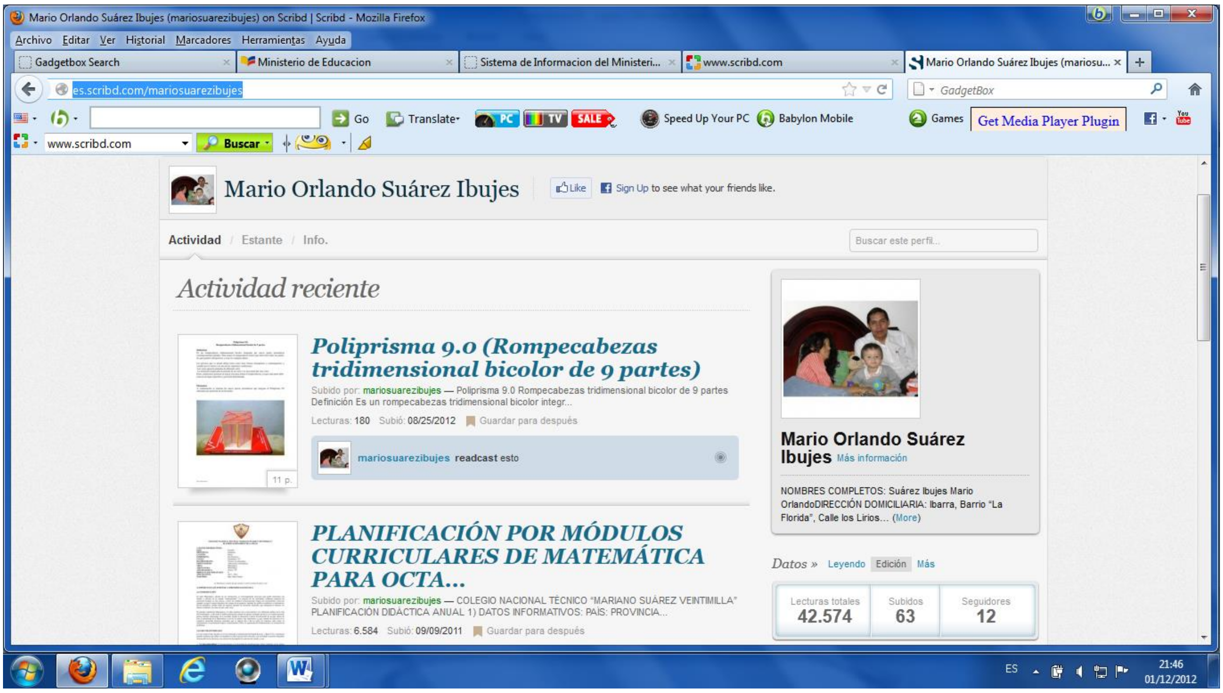
http://www.monografias.com/usuario/perfiles/mario_suarez_7/monografias



https://docentesinnovadores.net/Usuarios/Ver/29591



http://es.scribd.com/mariosuarezibujes



http://repositorio.utn.edu.ec/handle/123456789/24

Tema: Guía didáctica para el interaprendizaje de Trigonometría Básica empleando el Poliprisma.
Revista El Investigador Nº 4. Año 2012. Pág. 34-40. ISSN N: 1390-4833. www.utn.edu.ec/cuicyt.
Revista Científico Tecnológica de la Universidad Técnica del Norte. Indexada a LATIDEX folio 1806.

ARTICULOS CIENTIFICOS

Revista El Investigador No 4 Año 2012 Pags. 34-40
ISSN N: 1390-4833 www.utn.edu.ec/cuicyt

Guía didáctica para el interaprendizaje de Trigonometría Básica empleando el Poliprisma

Autor: Lic. Mario O. Suárez I. Mgs.
Tutor: Lic. Marco A. Benalcázar G.
Facultad Ciencias Administrativas y Económicas
Universidad Técnica del Norte
mgsmariosuarez@gmail.com

Receptado: 01/11/2011
Aprobado: 21/12/2011

Resumen

El presente trabajo de investigación previo a la obtención del Grado de Magíster en Gerencia de Proyectos Educativos y Sociales en la Universidad Técnica del Norte, tuvo como problema el investigar ¿cómo mejorar el interaprendizaje de Trigonometría Básica en el Primer Año de Bachillerato de la Especialidad de Física y Matemática del Colegio Nacional "Teodoro Gómez de la Torre"?. La presente investigación fue elaborada cumpliendo la modalidad de proyectos especiales, con enfoque cualitativo y diseño no experimental, que tuvo como objetivo general el elaborar una Guía Didáctica para mejorar el interaprendizaje de Trigonometría Básica en el Primer Año de Bachillerato de la Especialidad de Física y Matemática del Colegio Nacional "Teodoro Gómez de la Torre", por lo que en la presente investigación, como objetivos específicos, se diseñó y construyó un prototipo didáctico de estudio trigonométrico en forma de rompecabezas tridimensional bicolor de 9 partes prismáticas (Poliprisma) con sus respectivos ensayos experimentales, los que permiten