Prefacio

La presente investigación se refiere al tema de la Estadística, que se puede definir como la ciencia cuyo objetivo es reunir una información para facilitar al hombre el estudio de datos masivos de individuos, grupos, series de hechos, etc. y deducir de ello gracias al análisis de estos datos unos significados precisos o unas previsiones para el futuro.

También se refiere a la importancia, métodos e importancia de la estadística ya que está relacionada con el estudio de proceso cuyo resultado es más o menos imprescindible y con la finalidad de obtener conclusiones para tomar decisiones razonables de acuerdo con tales observaciones

La Estadística se ocupa de los métodos científicos para recolectar, organizar, resumir, presentar y analizar datos, así como de sacar conclusiones válidas y tomar decisiones con base en este análisis, así también realizar predicciones a cerca del conjunto del cual se han seleccionado dichos datos. El empleo cuidadoso de los métodos estadísticos permite obtener información precisa de los datos.

Objetivos:

- Comprenderla importancia del estudio de la historia de la estadística, para lo cual es necesario un recorrido por sus conceptos, métodos e importancia y más definiciones, con el fin de acercarnos un poco más al tema de la Estadística.
- Conocer sobre el tema con el cual se trabajara a lo largo del semestre en aplicable a la Administración, a la Electrónica y al Análisis de sistemas.
- Aplicar apropiadamente los métodos estadísticos en la recolección de inf. y procesos matemáticos básicos en cálculos estadísticos.
- Adquirir los conocimientos y habilidades sobre el tema, ser capaz de reconocer los elementos habituales de la estadística
- Aplicar los fundamentos básicos para realizar un buen trabajo en clase.

Contenido del programa de estudios:

- Introducción a la estadística
- Variables y representaciones
- Distribuciones de frecuencias
- Medidas de tendencia central
- Medidas de dispersión
- Curtosis y simetría
- Probabilidades

Introducción a la Estadística

Historia De La Estadística

Desde 3.000 años antes de Cristo, se tienen noticias de los primeros censos hechos a la población, en la antigua Babilonia, Persia, Egipto y China, se elaboraban censos de las propiedades de los habitantes con fines impositivos.

El mismo Moisés, que existió en los siglos XV - XIV antes de Cristo, y que era profeta y legislador hebreo, levantó un censo de su pueblo en el desierto, según lo señala la Biblia.

Y en Grecia, el censo era algo muy usual en sus principales ciudades democráticas.

También Servio Tulio, que se supone vivió entre 578 y 534 antes de Cristo, y fue el sexto Rey de Roma, ordenó que se llevara a cabo un censo cada 5 años, y el fin era el de planificar los impuestos, preparar elecciones y la conscripción militar. Como ha de recordarse, San José y la Virgen María iban a Belén a inscribirse en el segundo de estos censos, cuando nació Jesús, según sus discípulos Lucas, y Mateo, ya en la época del Emperador Augusto.

El primer censo en América fue llevado a cabo por los Incas, y lo más probable es que haya sido en la época de Pachacútec Yupanqui, Inca que fue llamado "El Reformador del Mundo" quien organizó el Imperio Incaico económica y socialmente.

El matemático y filósofo italiano Girolano Cardano, que vivió entre los años 1510 y 1576, realizó los primeros estudios sobre *probabilidades*, y fueron publicados en su trabajo "Iber de Ludo Alea" que quiere decir "Manual para tirar los dados".

Felipe II (1575 - 1578) fue el Rey de España, e hizo levantar un censo en el Nuevo Mundo de sus dominios, en el año de 1576.

Gottfried Achenwall (Desde 1719 hasta 1772), un reconocido economista y profesor universitario, de origen alemán, profundizó en estudios que dieron origen a la Estadística Inductiva.

Juan Pedro Sussmilchi, que vivió desde 1707 hasta 1767, y fue un brillante matemático, estadístico y teólogo alemán, perfeccionó los estudios demográficos, al mismo tiempo que Antonio Deparcioux, que vivió entre 1703 y 1768 y fue un gran matemático francés, aplicó la Estadística para obtener las primeras "Tablas de Mortalidad", con lo cual se dio inicio el próspero negocio del seguro de vida.

Jacques Bernouilli (1654-1705) matemático suizo, escribió "Ars Cojetandi" que quiere decir en español, el Arte de Conjeturar, publicado póstumamente en 1713 y formula la Ley de los Grandes Números, primer paso hacia la Estadística Matemática.

El Marqués Pedro Simón de Laplace que vivió desde 1749 hasta 1827, matemático y astrónomo francés, anuncia su Teoría Analítica de las Probabilidades en 1812, y este fue otro gran impulso a la Estadística Matemática.

Lambert Jacques Quetelet (1796-1874), gran astrónomo y matemático de origen belga, aplicó el método estadístico al estudio de la Economía Social (Características físicas, intelectuales y morales de los humanos); creando así la Sociometría.

Pafnuti Lvovich Chevyshev (1821-1884) crea la Desigualdad de Chevyshev, que es de gran utilidad como herramienta teórica, aplicable a las distribuciones de medias y varianzas finitas.

Gregor Johann Mendel, (1822-1884), conocido botánico austríaco, que experimentó con 34 variedades de arvejas, durante un lapso de 2 años, descubre y enuncia, en el año de 1865, las leyes de Mendel; leyes estadísticas que rigen la herencia y la hibridación de los vegetales, lo cual es considerado el punto de partida de la biometría.

Instituto Universitario de Tecnología para la Informática –IutepiInstituto Universitario de Tecnología para la Informática - Iutepi

El científico inglés, Francis Galton (1822-1911), primo de Darwin y creador de la Eugenesia, de nuevos métodos antropométricos, de la moderna teoría de la Estadística y su aplicación a la Sociometría y a la Biometría. Ideó los deciles y centiles.

Karl Pearson (1857-1936), matemático inglés, crea el método de los momentos, la Prueba de chi cuadrada, los conceptos de Curva normal, y de Desviación normal. Publica sus trabajos bajo el epígrafe de Contribución a la teoría matemática de la evolución, y en total, da un gran impulso a las técnicas usadas en estudios de fenómenos sociales (Sociometría) y biológicos (Biometría).

Hoy en día la Estadística ha llegado a tal grado de perfeccionamiento y especialización, que casi no existe disciplina científica, o técnica, de investigación, control o planificación, en la cual no se apliquen los métodos estadísticos como una herramienta de trabajo valiosísima e insustituible.

Concepto de Estadística

La estadística es la ciencia que le facilita al hombre el estudio de datos masivos, pasa de esa manera sacar conclusiones valederas y efectuar predicciones razonables de ellos; y así mostrar una visión de conjunto clara y de más fácil apreciación, así como para describirlos y compararlos.

En una forma práctica, la ESTADÍSTICA nos proporciona los métodos científicos para la recopilación, organización, resumen, representación y análisis de datos, o análisis de hechos, que se presenten a una valuación numérica; tales como son: Características biológicas o sociológicas, fenómenos físicos, producción, calidad, población riqueza, impuestos, cosechas, etc., ya sea para ayudar en la toma de decisiones o para explicar condiciones regulares o irregulares de algún fenómeno o estudio aplicado, de ocurrencia en forma aleatoria o condicional. Sin embargo estadística es más que eso, en otras palabras es el vehículo que permite llevar a cabo el proceso relacionado con la investigación científica.

Uso de la estadística

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

Campos de aplicación

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

- **En las ciencias naturales**: se emplea con profusión en la descripción de modelos termodinámicos complejos (mecánica estadística), en física cuántica, en mecánica de fluidos o en la teoría cinética de los gases, entre otros muchos campos.
- **En las ciencias sociales y económicas**: es un pilar básico del desarrollo de la demografía y la sociología aplicada.
- **En economía**: suministra los valores que ayudan a descubrir interrelaciones entre múltiples parámetros macro y microeconómicos.
- **En las ciencias médicas**: permite establecer pautas sobre la evolución de las enfermedades y los enfermos, los índices de mortalidad asociados a procesos morbosos, el grado de eficacia de un medicamento, etcétera.

Instituto Universitario de Tecnología para la Informática –IutepiInstituto Universitario de Tecnología para la Informática - Iutepi

Variables y Representaciones

Población:

El concepto de población en estadística va más allá de lo que comúnmente se conoce como tal. Una población se precisa como un conjunto finito o infinito de personas u objetos que presentan características comunes. Una población es un conjunto de todos los elementos que estamos estudiando, acerca de los cuales intentamos sacar conclusiones.

El tamaño que tiene una población es un factor de suma importancia en el proceso de investigación estadística, y este tamaño vienen dado por el número de elementos que constituyen la población, según el número de elementos la población.

Cuando el número de elementos que integra la población es muy grande, se puede considerar a esta como una población infinita, por ejemplo; el conjunto de todos los elementos positivos. Una población finita es aquella que está formada por un limitado número de elementos, por ejemplo; el número de estudiantes del IUTEPI de la Isabelica.

Cuando la población es muy grande, es obvio que la observación de todos los elementos se dificulta en cuanto al trabajo, tiempo y costos necesarios para hacerlo. Para solucionar este inconveniente se utiliza una muestra estadística.

Es a menudo imposible o poco práctico observar la totalidad de los individuos, sobre todo si estos son muchos, en lugar de examinar el grupo entero llamado población o universo, se examinan una pequeña parte del grupo llamada muestra.

Ejemplo:

Miembros del Colegio de Ingenieros del Estado Carabobo.

Muestra:

Se llama muestra a una parte o subconjunto de la población a estudiar que sirve para representarla.

Una muestra es una colección de algunos elementos de la población, debe ser definida en base a la población determinada, y las conclusiones que se obtengan de dicha muestra solo podrán referirse a la población en referencia.

Ejemplo;

El estudio realizado a 50 miembros del Colegio de Ingenieros del Estado Carabobo.

El estudio de muestras es más sencillo que el estudio de la población completa; cuesta menos y lleva menos tiempo.

Instituto Universitario de Tecnología para la Informática –IutepiInstituto Universitario de Tecnología para la Informática - Iutepi

Estadística descriptiva e inferencial: la estadística se divide en dos grandes áreas:

- **La estadística descriptiva,** se dedica a la descripción, visualización y resumen de datos originados a partir de los fenómenos de estudio. Los datos pueden ser resumidos numérica o gráficamente. Ejemplos básicos de parámetros estadísticos son: la media y la desviación estándar. Algunos ejemplos gráficos son: histograma, pirámide poblacional, clúster, entre otros.
- **La estadística inferencial,** se dedica a la generación de los modelos, inferencia y predicciones asociadas a los fenómenos en cuestión teniendo en cuenta la aleatoriedad de las observaciones. Se usa para modelar patrones en los datos y extraer inferencias acerca de la población bajo estudio.

Variable:

Es una característica observable que varía entre los diferentes individuos de una población.

Existen dos categorías o tipo de variables:

Variable cualitativa: Es aquella que expresa un atributo o característica, ejemplo: Rubio, moreno, etc.

Variable cuantitativa: Es aquella que podemos expresar numéricamente: edad, peso, nº de hijos de una familia, etc. Esta a su vez la podemos subdividir en:

- ✓ **Variable discreta**, aquella que entre dos valores próximos puede tomar a lo sumo un número finito de valores. Ejemplos: el número de hijos de una familia, el de obreros de una fábrica, el de alumnos de la universidad, etc.
- ✓ **Variable continúa**, la que puede tomar los infinitos valores de un intervalo. En muchas ocasiones la diferencia es más teórica que práctica, ya que los aparatos de medida dificultan que puedan existir todos los valores del intervalo. Ejemplos, peso, estatura, distancias, etc.

La variable se denota por las mayúsculas de letras finales del alfabeto castellano. A su vez cada una de las variables puede tomar distintos valores, colocando un subíndice, que indique orden.

Distribuciones de Frecuencias

Se llama frecuencia a la cantidad de veces que se repite (aparece) el mismo dato estadístico en un conjunto de observaciones de una investigación determinada.

Distribución de frecuencia

En estadística existe una relación con cantidades, números agrupados o no, los cuales poseen entre si características similares, donde se puede investigar como por ejemplo, los precios de los productos de la dieta diaria, la estatura y el peso de un grupo de individuos, las calificaciones de los estudiantes, etc., que pueden adquirir diferentes valores, que vienen siendo los datos estadísticos conocidos como variables.

La distribución de frecuencias o tabla de frecuencias es una ordenación en forma de tabla de los datos estadísticos, asignando a cada dato su frecuencia correspondiente. Las distribuciones de frecuencias pueden ser para datos directos o para datos agrupados o en intervalos de clase.

Instituto Universitario de Tecnología para la Informática –IutepiInstituto Universitario de Tecnología para la Informática - Iutepi

Distribución de frecuencia para datos directos o no agrupados

Es aquella distribución que indica las frecuencias con que aparecen los datos estadísticos, desde el menor de ellos hasta el mayor de ese conjunto.

Tipos de frecuencia

En estadística se pueden distinguir hasta cuatro tipos de frecuencias estas son:

Frecuencia absoluta (f_i) de una variable estadística X_i, es el número de veces que aparece en el estudio este valor. A mayor tamaño de la muestra, aumentará el tamaño de la frecuencia absoluta; es decir, la suma total de todas las frecuencias absolutas debe dar el total de la muestra estudiada (*N*).

Frecuencia relativa (h_i), es el cociente entre la frecuencia absoluta y el tamaño de la muestra (*N*). Es decir,

$$\boxed{hi=fi/N}$$

Si multiplicamos la frecuencia relativa por 100 obtendremos el porcentaje o tanto por ciento (*%*) que presentan esta característica respecto al total de *N*, es decir el 100% del conjunto.

Frecuencia absoluta acumulada (F_a), es el número de veces n_i en la muestra *N* con un valor igual o menor al de la variable. La última frecuencia absoluta acumulada deberá ser igual a *N*.

Frecuencia relativa acumulada (*Ha*), es el cociente entre la frecuencia absoluta acumulada y el número total de datos, *N*. Es decir,

$$\boxed{Ha=hi/N}$$

Ejemplo de distribución de frecuencia para datos Directos o no agrupados

Durante el mes de julio, en una ciudad se han registrado las siguientes temperaturas máximas:

32, 31, 28, 29, 33, 32, 31, 30, 31, 31, 27, 28, 29, 30, 32, 31, 31, 30, 30, 29, 29, 30, 30, 31, 30, 31, 34, 33, 33, 29, 29.

En la primera columna de la tabla colocamos la variable ordenada de menor a mayor, en la segunda hacemos el recuento y en la tercera anotamos la frecuencia absoluta.

Instituto Universitario de Tecnología para la Informática –IutepiInstituto Universitario de Tecnología para la Informática - Iutepi

Tabla N° 1. Distribución de Frecuencias para datos directos o no agrupados.

X_i	Repeticiones	f_i	F_a	h_i	H_a	$\%h_i$	$\%H_a$
27	/	1	1	0,032	0,032	3,2	3,2
28	//	2	3	0,065	0,097	6,5	9,7
29	//////	6	9	0,194	0,290	19,4	29
30	///////	7	16	0,226	0,516	22,6	51,6
31	////////	8	24	0,258	0,774	25,8	77,4
32	///	3	27	0,097	0,871	9,7	87,1
33	///	3	30	0,097	0,968	9,7	96,8
34	/	1	31	0,032	1	3,2	100
Total		31		1		100	

Fuente: García, D(2019)

Cálculos:

Para X = 27; h_i= 1/31 = 0,032 Ha = $\sum h_i$

$\%h_i = h_i*100 = 0{,}032*100 = 3{,}2\%$

$\%H_a = H_a*100 = 0{,}032*100 = 3{,}2\%$

Para cada ejercicio propuesto determine, distribución de frecuencias y sus porcentajes:

1) Al preguntar a 20 individuos por el número de personas que viven en su casa, hemos obtenido las siguientes respuestas:

5- 3- 4- 4- 1- 2- 4- 4- 5- 3- 4- 4- 3- 5- 4- 3- 2- 4- 5- 3

2) En una empresa de telefonía están interesados en saber cuál es el número de aparatos telefónicos (incluidos teléfonos móviles) que se tiene en las viviendas. Se hace una encuesta y, hasta ahora, han recibido las siguientes respuestas:

3-2- 1- 2- 2- 3- 4- 2- 3- 4- 2-1- 3- 3- 4- 3-2- 3- 2- 3

Instituto Universitario de Tecnología para la Informática –IutepiInstituto Universitario de Tecnología para la Informática - Iutepi

3) Hemos preguntado a 20 personas por el número medio de días que practican deporte a la semana y hemos obtenido las siguientes respuestas:

3- 1- 2- 3- 3- 6-2- 3- 5- 3- 6- 2- 0- 1- 6- 4-3- 2- 3- 7

4) Hemos lanzado un dado 20 veces y hemos ido anotando los resultados que obteníamos:

6- 3- 5- 3- 2- 3- 2- 4- 5- 1- 1- 1- 4- 4- 5- 1- 2- 6- 3- 5

5) En una clase se ha realizado un examen tipo test de 40 preguntas. El número de respuestas correctas conseguidas por cada uno de los alumnos de esa clase ha sido:

30- 25- 5- 10- 20- 20- 15- 10- 20- 40- 40- 30- 10- 30- 25- 30- 5- 40- 10- 20

Distribución de frecuencia para datos agrupados

Cuando se dispone de gran número de datos, es útil el distribuirlos en clases o categorías y determinar el número de individuos pertenecientes a cada clase, que viene siendo la frecuencia de clase.

Los datos ordenados y resumidos se suelen llamar *datos agrupados.* Aunque con el proceso de agrupamiento generalmente se pierde parte del detalle original de los datos, tiene la importante ventaja de presentarlos «todos» en un sencillo cuadro que facilita el hallazgo de las relaciones que pueda haber entre ellos, puestas así de manifiesto.

Ejemplo de distribución de frecuencia para datos agrupados

La siguiente tabla señala la distribución de frecuencias de alturas (registradas con aproximación de pulgada) de 100 estudiantes de una Universidad.

Tabla N° 2. Altura de 100 estudiantes de una Universidad.

Altura(pulgadas)	Número de estudiantes
60 – 62	5
63 – 65	18
66 – 68	42
69 – 71	27
72 – 74	8
Total	100

Fuente: García, D(2019)

La primera clase o categoría, por ejemplo, comprende las alturas de 60 – 62 pulgadas que corresponde a 5 estudiantes, así sucesivamente con cada categoría.

Instituto Universitario de Tecnología para la Informática –IutepiInstituto Universitario de Tecnología para la Informática - Iutepi

Rango o Amplitud total (recorrido)

Es el límite dentro del cual están comprendidos todos los valores de la serie de datos, en otras palabras, es el número de diferentes valores que toma la variable en un estudio o investigación dada. Es la diferencia entre el valor máximo de una variable y el valor mínimo que ésta toma en una investigación cualquiera. El rango es el tamaño del intervalo en el cual se ubican todos los valores que pueden tomar los diferentes datos de la serie de valores, desde el menor de ellos hasta el valor mayor estando incluidos ambos extremos. El rango de una distribución de frecuencia se designa con la letra R.

$$\boxed{A = L_s - L_i}$$

En tal caso, A = 62,5 – 59,5 = 3

Intervalos de clase y límites de clase

Se utiliza cuando n ≥ 30, los intervalos de clase permiten simplificar el manejo de los datos.

Un símbolo que define una clase, tal como 60 - 62 de la tabla anterior, se conoce como *intervalo de clase.* Los números extremos, 60 y 62, son los *límites de clase;* el número menor 60 es el *límite inferior de la clase* (X_i) y el mayor 62 es el *límite superior* (X_s). Los términos clase e intervalo de clase se utilizan a menudo indistintamente, aunque el intervalo de clase es realmente un símbolo para la clase.

Un intervalo de clase que, al menos teóricamente, no tiene límite superior o inferior, se conoce como *intervalo de clase abierto.* Por ejemplo, al referirse a la edad de grupos de individuos el intervalo de clase, «mayores de 65 años» es un intervalo de clase abierto.

Límites reales de clase

Si las alturas se registran con aproximación de pulgada, el intervalo de clase 60 – 62 teóricamente incluye todas las medidas desde 59,5000…..pulgadas. Estos números, representados brevemente por los números exactos 59,5 y 62,5; se conocen como *límites reales de clase o límites verdaderos de clase*; el menor de ellos, 59,5 es el límite real inferior (Li) y el mayor de ellos es 62,5 es el límite real superior (L_s).

A veces, los límites reales de clase se utilizan para simbolizar las clases. Por ejemplo, las diferentes clases de la primera columna de la tabla anterior podría indicarse por 59,5 – 62,5 – 65,5; etc., sin embargo, con total notación aparece una ambigüedad, pues los límites reales de clase no coincidirían con las observaciones reales. Así una observación fuese 62,5 no sería posible discernir si pertenece al intervalo de clase 59,5 – 62,5 – o al 62,5 – 65, 5.

Es decir; $L_i = 60 - 0,5 = 59,5$

$L_s = 62 + 0,5 = 62,5$

Marca de clase

La marca de clase es el punto medio del intervalo de clase y se obtiene sumando los límites inferior y superior de la clase y dividiendo por 2.

$$\boxed{X_m = (X_s + X_i) / 2 = (L_s + L_i) / 2}$$

Así la marca de clase del intervalo (60 – 62) es $\Rightarrow X_m = (60 + 62)/2 = 61$

Para los demás intervalos la marca de clase se calcula de la misma manera.

Instituto Universitario de Tecnología para la Informática –IutepiInstituto Universitario de Tecnología para la Informática - Iutepi

La tabla de distribución de frecuencias se determina de la misma forma que el ejemplo anterior de datos directos; obteniéndose la siguiente tabla:

Tabla N° 3. Distribución de Frecuencias para datos agrupados.

$X_i - X_s$	fi	F_a	L_i	L_s	Xm	h_i	H_a	$\%h_i$	$\%H_a$
60 – 62	5	5	59,5	62,5	61	0,05	0,05	5	5
63 – 65	18	23	62,5	65,5	64	0,18	0,23	18	23
66 – 68	42	65	65,5	68,5	67	0,42	0,65	42	65
69 – 71	27	92	68,5	71,5	70	0,27	0,92	27	92
72 – 74	8	100	71,5	74,5	73	0,08	1	8	100
Total	100					1		100	

Fuente: García, D(2019)

Ejercicio Propuesto:

1) Considere los siguientes datos:

14	21	23	21	16
19	22	25	16	16
24	24	25	19	16
19	18	19	21	12
16	17	18	23	25
20	23	16	20	19
24	26	15	22	24
20	22	24	22	20

a) Elabore una distribución de frecuencias usando las clases: 12 – 14, 15 – 17, 18 – 20, 21 – 23 y 24 – 26.
b) Elabore la distribución de frecuencias relativas y sus porcentajes con la misma amplitud de la parte a.

Histogramas y Polígonos de Frecuencia: Polígono de frecuencia es el nombre que recibe una clase de gráfico que se crea a partir de un histograma de frecuencia. Estos histogramas emplean columnas verticales para reflejar frecuencias, se forma a partir de la unión de los distintos puntos medios de las cimas de las columnas que configuran lo que es un histograma de frecuencias.

En las ciencias sociales, naturales y también en las económicas es donde con más frecuencia se hace uso de estos mencionados histogramas ya que se emplean para llevar a cabo lo que es la comparación de los resultados de un proceso determinado.

Polígono de frecuencia para datos agrupados: son aquellos que se desarrollan mediante la marca de clase que tiene coincidencia con el punto medio de las distintas columnas del histograma. En el momento de la representación de todas las frecuencias que forman parte de una tabla de datos agrupados.

Instituto Universitario de Tecnología para la Informática –IutepiInstituto Universitario de Tecnología para la Informática - Iutepi

Un polígono de frecuencia, por ejemplo, permite reflejar las temperaturas máximas promedio de una ciudad en un determinado período temporal. En el eje X (horizontal), deben indicarse los meses del año (enero, febrero, marzo, etc.). En el eje Y (vertical), se registran las temperaturas más altas promedio de cada mes (28°, 26°, 22°….). Los polígonos de frecuencia se suelen usar cuando se pretende retratar varias distribuciones distintas o la clasificación cruzada de una variable cuantitativa continua con una cualitativa discreta en el mismo dibujo. En el siguiente link pueden consultar ejemplos:
https://www.google.com/search?q=pol%C3%ADgonos+de+frecuencia&tbm=isch&source=univ&sa=X&ved=2ahUKEwjh4LiKosjgAhUix1kKHeImCTMQsAR6BAgAEAE&biw=1024&bih=710&dpr=1.25#imgdii=fBzKTtSk2uS5EM:&imgrc=Ro-VzfNIm6G6qM:

Curva de frecuencias (Ojiva de Galton): En estadística, la ojiva es un polígono frecuencial acumulado, es decir, que permite ver cuántas observaciones se encuentran por encima o debajo de ciertos valores, en lugar de solo exhibir los números asignados a cada intervalo. La ojiva apropiada para información que presenta frecuencias mayores que el dato que se está comparando tendrá pendientes negativa (hacia abajo y a la derecha) y en cambio la que se asigna a valores menores, tendrá una pendiente positiva.

Una gráfica similar al polígono de frecuencias es la ojiva, pero ésta se obtiene de aplicar parcialmente la misma técnica a una distribución acumulada y de igual manera que éstas.

Medidas de Tendencia Central

Son conocidas también como valores medios o medidas representativas, son de gran importancia en la estadística.

La media Aritmética: ($\bar{X}$)

Conocida también como promedio aritmético o media aritmética, es la medida de tendencia central más conocida y utilizada. Se representa por la letra griega µ cuando se trata del promedio del Universo o Población, y por $\bar{X}$ (léase X barra) cuando se trata del promedio de la muestra. Es importante destacar que µ es una cantidad fija mientras que el promedio de la muestra es variable puesto que diferentes muestras extraídas de la misma población tienden a tener diferentes medias. La media se expresa en la misma unidad que los datos originales: centímetros, horas, gramos, etc.

Se puede calcular para datos directos como para datos agrupados por intervalos, a continuación la fórmula para cada caso.

Media aritmética para datos directos sin repetición: Se suman todos los datos o variables y se divide entre el número total de ellos.

$$\boxed{\bar{X} = \sum X_i/n}$$

∑: Sumatoria

X_i: Variable en estudio

n: total de los datos

Ejemplo: Consideré los siguientes datos 3, 8, 4, 10, 6, 2. Determine la media aritmética

$$\bar{X} = (3 + 8 + 4 + 10 + 6 + 2) / 6 = 33/6$$

$$\bar{X} = 5{,}5$$

Instituto Universitario de Tecnología para la Informática –IutepiInstituto Universitario de Tecnología para la Informática - Iutepi

Media aritmética para datos directos con repetición: A la tabla de distribución de frecuencia para datos agrupados se le agrega una columna que indica la multiplicación de la variable (X_i) por la frecuencia absoluta (f_i), se totaliza la columna para obtener la sumatoria y se divide entre el total de la población.

$$\boxed{\bar{X} = \sum X_i.f_i/N}$$

f_i: Frecuencia (Número de veces que se repite el dato)

N: total de los datos de la población o muestra

Ejemplo: Calcular la media de una distribución estadística que viene dada por los siguientes datos;

X_i	61	64	67	70	73
f_i	5	18	42	27	8

Se procede a realizar la tabla de distribución de frecuencias, para obtener la media aritmética.

Tabla N°4. Distribución estadística

X_i	f_i	$X_i.f_i$
61	5	305
64	18	1152
67	42	2814
70	27	1890
73	8	584
Total	100	6745

Fuente: García, D(2019)

$\bar{X} = 6745/100 \Rightarrow \bar{X} = 67,45$

Instituto Universitario de Tecnología para la Informática –IutepiInstituto Universitario de Tecnología para la Informática - Iutepi

Media aritmética para datos agrupados en intervalos: Para éste caso hay que calcular la marca de clase para cada intervalo, como se explicó anteriormente en la distribución de frecuencia para datos agrupados.

$\boxed{\bar{X} = \sum X_m.f_i/N}$ X_m: marca de clase del intervalo

N: total de la población o muestra

Ejemplo: El peso de 50 trabajadores de una empresa se representa en la siguiente tabla de distribución de frecuencias.

Tabla N°5. Distribución de frecuencias

Intervalos de clase	53 – 57	58 – 62	63 – 67	68 – 72	73 – 77	78 – 82	83 – 87
f_i	2	7	10	12	9	6	4

Tabla N°6. Distribución de frecuencias para los pesos de 50 trabajadores

Intervalos de clase (kg)	**f_i**	**F_a**	**X_m**	**$X_m.f_i$**
53 – 57	2	2	55	110
58 – 62	7	9	60	420
63 – 67	10	19	65	650
68 – 72	12	31	70	840
73 – 77	9	40	75	675
78 – 82	6	46	80	480
83 – 87	4	50	85	340
Total	50			3515

Fuente: García, D(2019)

La marca de clase para el intervalo (53 – 57) se calcula; $X_m = (53 + 57) / 2 = 55$

Instituto Universitario de Tecnología para la Informática –IutepiInstituto Universitario de Tecnología para la Informática - Iutepi

Para el intervalo (58 – 62) ➡ $X_m = (58 + 62) / 2 = 60$, igual para los demás intervalos.

Calculando $X_m.f_i = 55x2 = 110$

$\bar{X} = 3515/50$ ➡ $\bar{X} =$ 70,3 ≈ 70 Kg.

La mediana:

En estadística se denomina mediana al valor que se encuentran en el lugar central de todos los datos de un estudio o una muestra de números. Podremos obtener la mediana para variables únicamente cuantitativas. El símbolo de la mediana se representa por Me, la mediana es por tanto el numero central de un grupo de números ordenados por su tamaño.

Para hallar la mediana en estadística, se ordenan los números de una muestra según su valor y se determina el que queda en el medio. Si la cantidad de términos es impar, la mediana es el valor central. Si la cantidad de términos es par, se suma los dos términos del medio y se divide entre 2.

Ejemplo: para los siguientes datos 8, 5, 2, 3,6 determine la mediana

1. Se ordenan de menor a mayor: 2,3,5,6,8
2. como se puede observar el valor central es 5, por lo tanto la Me = 5

La mediana para datos agrupados: La mediana para datos agrupados es necesario determinar el intervalo de clase donde ésta se localizará. Para ello, basta con observar en la tabla de frecuencias, la clase que tiene mayor frecuencia absoluta; la fórmula a utilizar es la siguiente:

$$\boxed{Me = L_i + (N/2 - F_a) * A/f_i}$$

L_i: Límite inferior correspondiente al intervalo donde se encuentra la mediana

N: total de la población o muestra

F_a: Frecuencia acumulada correspondiente al intervalo anterior a la mediana

f_i: frecuencia absoluta del intervalo donde se encuentra la mediana

A: amplitud del intervalo

Ejercicio: Tomando en cuenta los datos de la tabla N° 5, determine la Mediana

1) Calculamos N/2 = 50/2 = 25, se busca en la columna de F_a igual a 25 o mayor que 25, para el ejercicio se puede observar en el intervalo 68 – 72, la ubicación de la Mediana.

Instituto Universitario de Tecnología para la Informática –IutepiInstituto Universitario de Tecnología para la Informática - Iutepi

Tabla N°7. Distribución de frecuencias

Intervalos de clase (kg)	f_i	F_a
53 – 57	2	2
58 – 62	7	9
63 – 67	10	**19**
68 – 72	**12**	31
73 – 77	9	40
78 – 82	6	46
83 – 87	4	50
Total	50	

Intervalo de la Mediana →

← Intervalo de la Moda

Fuente: García, D(2019)

2) Se toman los datos del intervalo para calcular la mediana, L_I = 68 – 0,5= 67,5
F_a= 19, f_i = 12, A= 72,5 – 67,5= 5, sustituyendo los datos en la fórmula se obtiene

$$Me= 67,5 + (25 – 19)*5/12 \Rightarrow Me= 70\ Kg.$$

La Moda:

La moda es el valor de un conjunto de datos que aparece con mayor frecuencia. No depende de valores extremos, pero es más variable que la media y la mediana. Aquí se va determinar la moda para datos agrupados en intervalos, cuya fórmula es la siguiente:

$$\boxed{Mo = L_i + \Delta_1 * A/(\Delta_1 + \Delta_2)}$$

L_i: Límite inferior correspondiente al intervalo modal, es decir el intervalo con mayor frecuencia

Δ_1: Es la diferencia de la frecuencia modal y la frecuencia anterior al intervalo modal

Δ_2: Es la diferencia de la frecuencia modal y la frecuencia posterior al intervalo modal

A: amplitud del intervalo modal

Ejercicio: continuando con el ejercicio anterior, determine la moda

Instituto Universitario de Tecnología para la Informática –IutepiInstituto Universitario de Tecnología para la Informática - Iutepi

1) Para conseguir el intervalo correspondiente a la moda, se busca en la columna f_i el mayor valor de todos, en la tabla coincide con el intervalo de la mediana, ya que el mayor valor es 12
2) Se determina $L_i = 67{,}5$ A = 5
   $\Delta_1 = 12 - 10 = 2$ $\Delta_2 = 12 - 9 = 3$ Sustituyendo los valores en la fórmula

$$Mo = 67{,}5 + 2*5/(2 + 3) \Rightarrow Mo = 69{,}5 \approx 70 \text{ Kg.}$$

Ejercicio Propuesto:

La distribución de frecuencias siguiente muestra los precios de las 30 acciones del Promedio Industrial Dow Jones (*The Wall Street Joumal,* 16 de enero de 2006).

Tabla N°8. Distribución de frecuencias de los precios de 30 acciones

Precio por acción ($)	Frecuencia
20 – 29	7
30 – 39	6
40 – 49	6
50 – 59	3
60 – 69	4
70 – 79	3
80 – 89	1

Fuente: García, D(2019)

Calcule el precio medio por acción y la desviación estándar de los precios por acción en el Promedio Industrial Dow Jones.

Medidas de Orden o de Posición

Son indicadores usados para señalar que porcentaje de datos dentro de una distribución de frecuencias superan estas expresiones, cuyo valor representa el valor del dato que se encuentra en el centro de la distribución de frecuencia, por lo que también se les llama “medidas de tendencia central.

Las medidas de posición dividen un conjunto de datos en grupos con el mismo número de individuos. Para calcular las medidas de posición es necesario que los datos estén ordenados de menor a mayor.

Instituto Universitario de Tecnología para la Informática –IutepiInstituto Universitario de Tecnología para la Informática - Iutepi

Se determina como la mediana, a continuación se describen las medidas de posición más comunes utilizadas en estadística, como lo son:

- ✓ **Cuartiles:** Hay 3 cuartiles que dividen a una distribución en 4 partes iguales (primero, segundo y tercer cuartil).
- ✓ **Deciles:** Hay 9 deciles que la dividen en 10 partes iguales (primero al noveno decil).
- ✓ **Percentiles:** Hay 99 percentiles que dividen a una serie en 100 partes iguales (primero al noventa y nueve percentil).

Cuartiles: (Q_1, Q_2, Q_3)

Para datos agrupados:

$$Q_k = L_K + (K*N/4 - F_K)*A/f_k$$

L_k: Límite real inferior correspondiente a la clase del cuartil

K*N/4: Lugar donde está el cuartil para ubicarlo en la tabla de distribución

F_k: Frecuencia acumulada de la clase que antecede al cuartil

f_k: frecuencia absoluta de la clase del cuartil

Existen tres tipos de cuartiles
Q_1 representa el 25 % de todos los datos
Q_2 representa el 50% de todos los datos
Q_3 representa el 75% de todos los datos

Deciles: (D_1, D_2, ….D_9)

$$D_k = L_K + (K*N/10 - F_K)*A/f_k$$

L_k: Límite real inferior correspondiente a la clase del decil

K*N/4: Lugar donde está el decil para ubicarlo en la tabla de distribución

F_k: Frecuencia acumulada de la clase que antecede al decil

f_k: frecuencia absoluta de la clase del decil

Percentiles: se obtiene al dividir la distribución en 100 partes iguales

$$P_k = L_K + (K*N/100 - F_K)*A/f_k$$

L_k: Límite real inferior correspondiente a la clase del percentil

K*N/4: Lugar donde está el percentil para ubicarlo en la tabla de distribución

F_k: Frecuencia acumulada de la clase que antecede al percentil

f_k: frecuencia absoluta de la clase del percentil

Instituto Universitario de Tecnología para la Informática –IutepiInstituto Universitario de Tecnología para la Informática - Iutepi

Ejercicio: en el siguiente cuadro se cuenta con la distribución de edades del profesorado de una Escuela Superior de Enfermería:

Tabla N°9. Distribución de edades de la plantilla Docente de una Escuela de Enfermería

Edad (Años)	**Frecuencia absoluta**	**Frecuencia acumulada**	
30 – 35	3	3	
36 – 40	3	6	D_1
41 – 45	6	12	
46 – 50	14	26	Q_1
51 – 55	9	35	Q_2; D_5
56 – 60	8	43	P_{70}
61 – 65	7	50	Q_3
66 – 70	6	56	P_{90}
71 – 75	4	60	
TOTAL	60		

Fuente: García, D(2019)

Determine: a) Q_1 b) Q_2 c) Q_3 d) D_1 e) D_5 f) P_{70} g) P_{90}

Solución:

a) Primero se obtiene K*N/4, Para ubicar en la tabla la posición del cuartil;
K*N/4 = 1*60/4 = 15, se busca en la columna de Fa igual o mayor al valor, en éste caso mayor a 15, como se puede observar en la tabla anterior, se encuentra en el intervalo 46 – 50, se procede de igual forma cuando se calculó la mediana, es decir L_k= 46 – 0,5 = 45,5 F_k= 12 f_k=14

41 – 45	6	12	
46 – 50	**14**	**26**	Q_1

Al sustituir se obtiene; $Q_1 = 45,5 + (15 – 12)*5/14 = 45,5 + 15/14 = 46,57 \approx$ $Q_1 =$ 47 Años.

b) $K*N/4 = 2*60/4 = 30$; $L_i = 51 – 0,5 = 50,5$

46 – 50	14	26
51 – 55	**9**	**35**

Q_2

Sustituyendo se tiene; $Q_2 = 50,5 + (30 – 26)*5/9 = 52,72 \approx$ Q_2 = 53 Años.

c) $K*N/4 = 3*60/4 = 45$

56 – 60	8	43
61 – 65	**7**	**50**

Q_3

Sustituyendo en la fórmula se obtiene; $Q_3 = 60,5 + (45 – 43)*5/7 = 61,92$ $\approx$ Q_3 = 62 Años.

d) $K*N/10 = 1*60/10 = 6$;

30 – 35	3	3
36 – 40	**3**	**6**

D_1

Sustituyendo valores; $D_1 = 35,5 + (6 – 3)*5/3 = 40,5$ $\approx$ D_1 = 41 Años.

e) $K*N/10 = 5*60/10 = 30$; Como se puede observar en la tabla anterior coincide el Decil 5to con el 2do cuartil, por lo que el $D_5 = Q_2$ = 53 Años.

f) $K*N/100 = 70*60/100 = 42$

51 – 55	9	35
56 – 60	**8**	**43**

P_{70}

Calculando $P_{70} = 55,5 + (42 – 35)*5/8 = 59,875$ $\approx$ P_{70} = 60 Años.

g) $K*N/100 = 90*60/100 = 54$

61 – 65	7	50

Instituto Universitario de Tecnología para la Informática –IutepiInstituto Universitario de Tecnología para la Informática - Iutepi

66 – 70	**6**	**56**

$\}$ P_{90}

Calculando P_{90} = 65,5 + (54 – 50)*5/6 = 68,83 ≈ P_{90} = 69 Años.

Medidas de Dispersión

Variación o Dispersión

En las secciones anteriores (ordenamiento, agrupación de datos y medidas de tendencia central) se ha visto que es de utilidad ubicar el centro del conjunto de datos. Pero identificar una de las medidas de tendencia central rara vez es suficiente para describir de manera más completa los datos. Una descripción más completa del conjunto de datos, puede obtenerse si se mide qué tan dispersos están los datos alrededor de ese punto central, en otras palabras, que tan cerca o que tan lejos pueden estar los datos con relación al punto central. Y eso es lo que hacen las medidas de dispersión, que indican cuanto se desvían las observaciones alrededor de ese punto central.

La dispersión se relaciona con la concentración (mayor o menor) de los datos en torno al valor central, generalmente la media.

Las medidas de dispersión muestran la variabilidad de una distribución, indicándolo por medio de un número, si las diferentes puntuaciones de una variable están muy alejadas de la media. Cuanto mayor sea ese valor, mayor será la variabilidad, cuanto menor sea, más homogénea será la media. Así se sabe si todos los casos son parecidos o varían mucho entre ellos.

Las medidas de dispersión son:

1) **Desviación media para datos sin agrupar:** la desviación respecto a la media es la diferencia entre cada valor de la variable estadística y la media aritmética.

$$D_i = X_i - \overline{X}.$$

La desviación media es la media aritmética de los valores absolutos de las desviaciones respecto a la media. La desviación media se representa por D_x, como se señala a continuación

$$D_m = \frac{1}{N}\sum_{i=1}^{n}|x_i - \overline{x}|$$

2) **Desviación media para datos agrupados:** si los datos vienen agrupados en una tabla de frecuencias, la expresión de la desviación media es:

$$D_{\overline{x}} = \frac{\sum_{i=1}^{n}|x_i - \overline{x}|f_i}{N}$$

Donde $X_i = X_m$

Instituto Universitario de Tecnología para la Informática –IutepiInstituto Universitario de Tecnología para la Informática - Iutepi

Ejercicio: Determine la desviación media para la siguiente distribución

Tabla N°10. Distribución de frecuencias para los pesos de 50 trabajadores

Intervalos de clase (kg)	f_i	F_a	X_m	$X_m.f_i$	$\lvert X_m - \overline{X} \rvert$	$\lvert (X_m - \overline{X})*f_i \rvert$
53 – 57	2	2	55	110	15	30
58 – 62	7	9	60	420	10	70
63 – 67	10	19	65	650	5	50
68 – 72	12	31	70	840	0	0
73 – 77	9	40	75	675	5	45
78 – 82	6	46	80	480	10	60
83 – 87	4	50	85	340	15	60
Total	50			3515		315

Fuente: García, D(2019)

Calculando la media:

$\boxed{\overline{X} = \sum X_m.f_i/N}$

$\overline{X} = 3515/50 = 70$

$\lvert X_m - \overline{X} \rvert = \lvert 55 - 70 \rvert = 15$

$\lvert X_m - \overline{X} * f_i \rvert = 15*2 = 30$, igual procedimiento para los demás intervalos

$D_x = 315/50 \rightarrow D_x = 6,3$

3) **Varianza:** la varianza es la media aritmética del cuadrado de las desviaciones respecto a la media de una distribución estadística. La varianza se representa por σ^2

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2 f_i}{N}$$

Instituto Universitario de Tecnología para la Informática –IutepiInstituto Universitario de Tecnología para la Informática - Iutepi

Ejercicio: calcular la varianza de la distribución de la tabla anterior, donde la media es de 70 Kg

Tabla N°11. Distribución de frecuencias para los pesos de 50 trabajadores

Intervalos de clase (kg)	f_i	$X_m = X_i$	$(X_m - \bar{X})^2$	$(Xm - X)^2 * f_i$
53 – 57	2	55	225	450
58 – 62	7	60	100	700
63 – 67	10	65	25	250
68 – 72	12	70	0	0
73 – 77	9	75	25	225
78 – 82	6	80	100	600
83 – 87	4	85	225	900
Total	50			3125

Fuente: García, D(2019)

Para el intervalo 53 – 57

$(X_m - X)^2 = (55 - 70)^2 = 225$ $(Xm - X)^2 * f_i = 225*2 = 450$

Obteniendo la varianza: $\sigma^2 = 3125/50 = 62,5$ Kg.

Propiedades de la Varianza:

1. La varianza será siempre un valor positivo o cero, en el caso de que las puntuaciones sean iguales.
2. Si a todos los valores de la variable se les suma un número la varianza no varía.
3. Si todos los valores de la variable se multiplican por un número la varianza queda multiplicada por el cuadrado de dicho número.
4. Si tenemos varias distribuciones con la misma media y conocemos sus respectivas varianzas se puede calcular la varianza total.

4) Desviación típica: La desviación típica es la raíz cuadrada de la varianza, es decir, la raíz cuadrada de la media de los cuadrados de las puntuaciones de desviación.

Instituto Universitario de Tecnología para la Informática –IutepiInstituto Universitario de Tecnología para la Informática - Iutepi

La desviación típica se representa por σ.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 f_i}{N}}$$

O se determina calculando la raíz cuadrada al resultado de la varianza.

$\sigma^2 = \sqrt{62{,}5} = 7{,}91$ kg

Propiedades de la desviación típica:

1. La desviación típica será siempre un valor positivo o cero, en el caso de que los valores sean iguales.
2. Si a todos los valores de la variable se les suma un número la desviación típica no varía.
3. Si todos los valores de la variable se multiplican por un número la desviación típica queda multiplicada por dicho número.
4. Si tenemos varias distribuciones con la misma media y conocemos sus respectivas desviaciones típicas se puede calcular la desviación típica total.

Ejercicio Propuesto:

Un sistema de radar de la policía vigila los automóviles en una carretera que permite una velocidad máxima de 55 millas por hora. La siguiente es una distribución de frecuencias de las velocidades.

Tabla N°12. Distribución de frecuencias de las velocidades de los automóviles

Velocidad (millas por hora)	Frecuencia
45 – 49	10
50 – 54	40
55 – 59	150
60 – 64	175
65 – 69	75
70 – 74	15
75 – 79	10
Total	**475**

Fuente: García, D(2019)

a) ¿Cuál es la velocidad media de los automóviles en esta carrera?
b) Calcule la varianza y la desviación estándar.

Instituto Universitario de Tecnología para la Informática –IutepiInstituto Universitario de Tecnología para la Informática - Iutepi

Curtosis y simetría

Curtosis: también conocida como medida de apuntamiento, es una medida estadística, que determina el grado de concentración que presentan los valores de una variable alrededor de la zona central de la distribución de frecuencias.

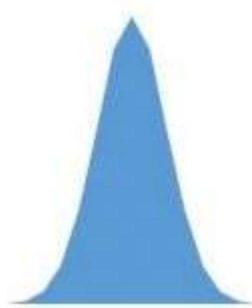
Cuando medimos una variable aleatoria, por lo general, los resultados que tienen una mayor frecuencia son los que se sitúan en torno a la media de la distribución.

Imaginemos la altura de los alumnos de una clase. Si la altura media de la clase es 1,72 lo más normal es que las alturas del resto de los alumnos estén en torno a este valor, se considera que la distribución de la variable aleatoria se distribuye con normalidad. Pero dada la infinidad de variables que se pueden medir, esto no siempre sucede así.

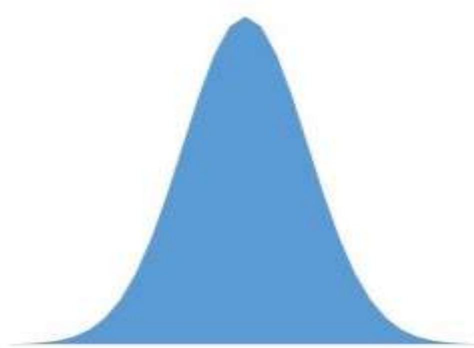
Existen algunas variables que presentan un mayor grado de concentración (menor dispersión) de los valores en torno a su media y otras, por el contrario, presentan un menor grado de concentración (mayor dispersión) de sus valores en torno a su valor central. Por tanto, la curtosis nos informa de lo apuntada (mayor concentración) o lo achatada (menor concentración) que es una distribución.

Tipos de curtosis: dependiendo del grado de curtosis, existen tres tipos;

1. Leptocúrtica: Existe una gran concentración de los valores en torno a su media

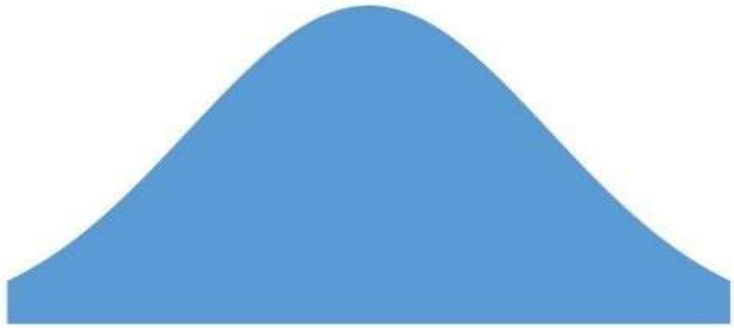


2. Mesocúrtica: Existe una concentración normal de los valores en torno a su media



3. Platicúrtica: Existe una baja concentración de los valores en torno a su media

Instituto Universitario de Tecnología para la Informática –IutepiInstituto Universitario de Tecnología para la Informática - Iutepi



En el siguiente enlace puede visualizar un ejemplo de curtosis:

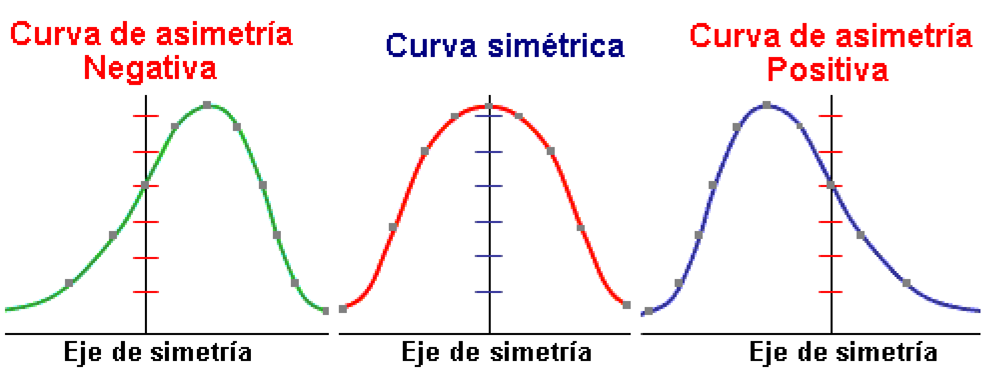
https://www.youtube.com/watch?v=9-nRaVKs5No

Simetría: Se da cuando una distribución se distribuye aproximadamente la misma cantidad de los datos a ambos lados de la media aritmética. No tiene alargamiento o sesgo, se representa por una curva normal en forma de campana llamada campana de Gauss o también conocida como de Laplace, también se dice que una distribución es simétrica cuando su media aritmética, su mediana y su moda son iguales.

La simetría de una distribución de frecuencias hace referencia al grado en que los valores de la variable, equidistantes a un valor que se considere centro de distribución, posee frecuencias similares, es un concepto más intuitivo a nivel visual, especialmente, si se observa una representación gráfica (diagrama de barras, histograma, etc.,) de la distribución de frecuencias. Ésta será simétrica si la mitad izquierda de la distribución es la imagen especular de la mitad derecha.

Asimetría: Esta medida nos permite identificar si los datos se distribuyen de forma uniforme alrededor del punto central (media aritmética). La asimetría presenta tres estados diferentes (ver figura #1), cada uno de los cuales define de forma concisa como están distribuidos los datos respecto al eje de asimetría. Se dice que la ***asimetría es positiva*** cuando la mayoría de los datos se encuentran por encima del valor de la media aritmética, y la ***asimetría negativa*** cuando la mayor cantidad de datos se aglomeran en los valores menores que la media. La curva es ***simétrica*** como se explicó en el párrafo anterior.

Fig. # 1. Ejemplo de Asimetría y Simetría



Instituto Universitario de Tecnología para la Informática –IutepiInstituto Universitario de Tecnología para la Informática - Iutepi

Probabilidades

Entre los primeros teóricos de la probabilidad se encuentran Jacob Bernoulli (1654–1705), Abraham de Moivre (1667–1754), el reverendo Thomas Bayes (1702–1761) y Joseph Lagrange (1736–1813) quienes desarrollaron fórmulas y técnicas para el cálculo de la probabilidad. En el siglo XIX, Pierre Simón, Márquez de Laplace (1749–1827), unificó todas estas primeras ideas y compiló la primera teoría general de probabilidad. La teoría de probabilidad fue aplicada con éxito en las mesas de juego y, lo que es más importante en nuestro estudio, en problemas sociales y económicos, muchos centros de aprendizaje estaban estudiando la probabilidad como una herramienta para el entendimiento de los fenómenos sociales.

En la actualidad, la teoría matemática de la probabilidad es la base para las aplicaciones estadísticas tanto en investigaciones sociales como en la toma de decisiones. La probabilidad es una parte de nuestras vidas cotidianas, en la toma de decisiones personales y administrativas, nos enfrentamos a la incertidumbre y utilizamos la teoría de la probabilidad, admitamos o no el uso de algo tan complejo. Cuando escuchamos una predicción de un 70% de posibilidades de lluvia, cambiamos nuestros planes de salir de día de campo y nos quedamos en casa divirtiéndonos con juegos de mesa.

Los Administradores que se encargan del inventario de ropa de moda para mujer deben preguntarse sobre las posibilidades de que las ventas alcancen o excedan un cierto nivel. Vivimos en un mundo que es incapaz de predecir el futuro con total certidumbre.

Antes de profundizar en la forma como se utilizan las probabilidades, es necesario conocer de cierta manera de donde provienen. Hay tres formas de calcular o estimar la probabilidad. El enfoque clásico o a priori proveniente de los juegos de azar o definición clásica de Laplace que se emplea cuando los espacios muéstrales son finitos y tienen resultados igualmente probables; la definición empírica, a posteriori o frecuencia relativa de ocurrencia de un evento con respecto a un gran número de ensayos repetidos y por último. La probabilidad mide la frecuencia con que se obtiene un resultado (o conjunto de resultados) al llevar a cabo un experimento aleatorio, del que se conocen todos los resultados posibles, bajo condiciones suficientemente estables. La probabilidad es simplemente que tan posible es que ocurra un evento determinado.

Cuando no estamos seguros del resultado de un evento, podemos hablar de la probabilidad de ciertos resultados: que tan común es que ocurran.

Definiciones básicas de probabilidad:

Probabilidad: Es una medida numérica de la posibilidad de que ocurra un evento. Por tanto, las probabilidades son una medida del grado de incertidumbre asociado con cada uno de los eventos previamente enunciados.

Evento: En teoría de la probabilidad, un evento es uno o más posibles resultados de hacer algo. Por ejemplo, lanzar una moneda al aire, si cae cara es un evento, y si cae sello (cruz) es otro evento, elegir un estudiante de 100, para que responda una pregunta, cuando escuchamos las pocas gratas predicciones del índice de mortalidad en accidentes de tránsito, esperamos no ser uno de tales eventos.

Eventos exhaustivamente colectivos: lista de eventos que representa todos los resultados posibles de un experimento.

Eventos mutuamente excluyentes: Eventos que no se pueden presentar juntos.

Experimento: En el contexto de la probabilidad, un experimento es definido como un proceso que genera resultados definidos, es la actividad que se origina de dichos eventos, como el de lanzar una modera al aire es el experimento, lanzar un dado, determinar el número de alumnos que aprobaron estadística, etc.

Espacio Muestral: Es el conjunto de todos los resultados posibles de un experimento, el ejemplo de lanzar una moneda, su espacio muestral se denota con la letra S, y viene siendo: S = {cara, sello}

Instituto Universitario de Tecnología para la Informática –IutepiInstituto Universitario de Tecnología para la Informática - Iutepi

Diagrama de Venn: Representación gráfica de los conceptos de probabilidad en la que el espacio muestral está representado por un rectángulo y los eventos que suceden en el espacio muestral se representan como parte de dichos rectángulos.

Árbol de probabilidades: Representación gráfica que muestra los resultados posibles de una serie de experimentos y sus respectivas probabilidades.

Tabla N°13.Ejemplo de Experimento y Espacio Muestral

Experimento	Resultado del experimento
Lanzar una moneda	Cara, cruz(sello)
Tomar una moneda para inspeccionarla	Con defecto, sin defecto
Realizar una llamada de ventas	Hay compra, no hay compras
Lanzar un dado	1,2,3,4,5,6
Jugar un partido de futbol	Ganar, perder, empatar

Definiciones de probabilidad clásica:

El planteamiento clásico define la probabilidad de que un evento ocurra como:

$$\textbf{Probabilidad de un Evento} = \frac{\textbf{Número de resultados en los que se presenta el evento}}{\textbf{Número total de resultados posibles}}$$

Ejemplo: Calcular la probabilidad de obtener cara al lanzar una moneda, aplicando la fórmula se obtiene

$$P(\text{cara}) = \frac{1}{(1+1)}$$

1 → Número de resultados posibles en un lanzamiento en los que se presenta el evento (en este caso, el # de resultados que producirán una cara)

(1 + 1) → Número total de resultados posibles en un lanzamiento (una cara y un sello)

P(cara) = ½ = 0,5

Instituto Universitario de Tecnología para la Informática –IutepiInstituto Universitario de Tecnología para la Informática - Iutepi

Planteamiento de frecuencia relativa:

En el siglo XIX, los estadísticos británicos, interesados en la fundamentación teórica del cálculo del riesgo de pérdidas en las pólizas de seguros de vida y comerciales, empezaron a recoger datos sobre nacimientos y defunciones. En la actualidad, a este planteamiento se le llama frecuencia relativa de presentación de un evento y define la probabilidad como: la frecuencia relativa observada de un evento durante un gran número de intentos o la fracción de veces que un evento se presenta a la larga, cuando las condiciones son estables.

Ejemplo: suponga que en una compañía de seguros sabe, por la información obtenida de los datos actuariales registrados, que de los hombres de 40 años de edad, 60 de cada 100.000 morirán en un período de un año. Calculando la probabilidad queda determinada

P(muerte de 40 años) = 60/100.000 = 0,00006

Planteamiento de probabilidad subjetiva:

Las probabilidades subjetivas están basadas en las creencias de las personas que efectúan la estimación de probabilidad. De hecho, la probabilidad subjetiva se puede definir como la probabilidad asignada a un evento por parte de un individuo, basada en la evidencia que se tenga disponible. Esta evidencia puede presentarse en forma de frecuencia relativa de presentación de eventos pasados, o puede tratarse simplemente de una creencia meditada.

Las asignaciones de probabilidad subjetiva se dan con más frecuencia cuando los eventos se presentan sólo una vez o un número reducido de veces. Por ejemplo, usted tiene encomendada la tarea de entrevistar y elegir a un nuevo trabajador social. Su población se ha reducidos a sólo tres personas, cada una de éstas tiene buena apariencia, alto nivel de actividad, bastante confianza en sí misma, buen registro de logros pasados y buena disposición de ánimo para enfrentar los retos que se presentan. ¿Cuáles son las posibilidades de que cada candidato se relacione exitosamente con los clientes? El responder a esta pregunta y escoger a uno de los tres requerirá que usted asigne una probabilidad subjetiva al potencial de cada persona que solicita el puesto.

Reglas de probabilidad:

Los valores de probabilidad se encuentran en una escala de 0 a 1. Los valores cercanos a 0 indican que las posibilidades de que ocurra un evento son muy pocas. Los cercanos a 1 indican que es casi seguro que ocurra un evento. Otras probabilidades entre cero y uno representan distintos grados de posibilidad de que ocurra un evento. Por ejemplo, si considera el evento “que llueva mañana”, se entiende que si el pronóstico del tiempo dice “la probabilidad de que llueva es cercana a cero”, implica que casi no hay posibilidades de que llueva. En cambio, si informan que la probabilidad de que llueva es 0.90, sabe que es muy posible que llueva, lo cual representa un 90% de posibilidad. La probabilidad de 0,50 indica que es igual de posible que llueva como no llueva.

Probabilidad compuesta:

- ✓ Los experimentos compuestos están formados por dos o más experimentos simples.
- ✓ Utilizamos diagramas de árbol para obtener el conjunto de resultados posibles.
- ✓ La probabilidad de un camino es igual al producto de las probabilidades de las ramas de dicho camino.

Tipos de eventos o sucesos:

Sucesos independientes: extracción con devolución (con reemplazo)

Dos sucesos A y B son independientes si la realización de A no condiciona la realización del suceso B, es decir, p(B/A) = p(B)

$$\boxed{P\ (A \cap B) = p(A) * p(B)}$$

Instituto Universitario de Tecnología para la Informática –IutepiInstituto Universitario de Tecnología para la Informática - Iutepi

Sucesos dependientes: extracción sin devolución (sin reemplazo)

Dos sucesos A y B son dependientes si la realización de A condiciona la realización de B, es decir, $p(B/A) \neq p(B)$

$$\boxed{P\ (A \cap B) = p(A)*p}$$

Ejemplo de probabilidad compuesta:

En una caja hay 5 bolas: 3 azules y 2 verdes. Se extrae una bola, se anota el color y se repite el mismo proceso otra vez. a) ¿Cuál es la probabilidad de obtener 2 bolas azules? b) ¿Cuál es la probabilidad de que la 1ra sea verde y la 2da sea azul? Con devolución y sin devolución para cada alternativa.

Con devolución: Sucesos independientes

a) Probabilidad de obtener 2 bolas azules
Evento A: 1ra extracción obtener bola azul $\longrightarrow$ $p(A) = 3/5$
Evento B: 2da extracción obtener bola azul $\longrightarrow$ $P(B) = 3/5$

$$P(A \cap B) = P(A)*P(B) = 3/5*3/5 = 9/25$$

b) Probabilidad de obtener en la 1ra extracción verde y 2da extracción azul
Evento A: 1ra extracción obtener bola verde $\longrightarrow$ $p(A) = 2/5$
Evento A: 1ra extracción obtener bola azul $\longrightarrow$ $p(B) = 3/5$

$$P(A \cap B) = P(A)*P(B) = 2/5*3/5 = 6/25$$

Sin devolución: Sucesos dependientes

a) Probabilidad de obtener 2 bolas azules
Evento A: 1ra extracción obtener bola azul $\longrightarrow$ $p(A) = 3/5$
Evento B: 2da extracción obtener bola azul $\longrightarrow$ $P(B) = 2/4$

$$P(A \cap B) = P(A)*P(B/A) = 3/5*2/4 = 6/20 = 3/10$$

b) Probabilidad de obtener en la 1ra extracción verde y 2da extracción azul
Evento A: 1ra extracción obtener bola verde $\longrightarrow$ $p(A) = 2/5$
Evento A: 1ra extracción obtener bola azul $\longrightarrow$ $p(B) = 3/4$

$$P(A \cap B) = P(A)*P(B/A) = 2/5*3/4 = 6/20 = 3/10$$

Sucesos mutuamente excluyentes:

Se dice que 2 eventos son mutuamente excluyentes si no tienen puntos muéstrales en común. Dos o más eventos son mutuamente excluyentes o disjuntos, si no pueden ocurrir simultáneamente. Es decir, la ocurrencia de un evento impide automáticamente la ocurrencia del otro evento (o eventos).

Por ejemplo, al lanzar una moneda al aire, solo puede ocurrir que salga cara o sello pero no los dos a la vez, es decir, sale cara, o sale sello(cruz).

La intersección de dos o más sucesos mutuamente excluyentes, es el conjunto vacío, $A \cap B = 0$

Instituto Universitario de Tecnología para la Informática –IutepiInstituto Universitario de Tecnología para la Informática - Iutepi

Ley de adicción para eventos mutuamente excluyentes:

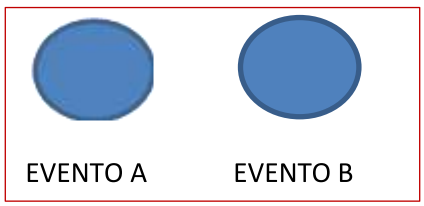
Probabilidad de que A suceda

P(AUB) = P(A) + P(B) ⇨ P (AoB) = P(A) + P (B)

abilidad de que B suceda.

Probabilidad de que suceda A o B

Fig. # 2. Diagrama de Venn



Fuente: García, D(2019)

Ejemplo:

Cinco estudiantes por igual capacidad, esperan la fecha en que se les hará una entrevista para trabajar en el verano, la compañía solicitante ha anunciado que contratará a sólo uno de los cinco, mediante una elección aleatoria. El grupo está formado por los estudiantes siguientes: Bill, Helen, John, Sally y Walter. Determine: a) ¿Cuál es la probabilidad de que John sea elegido? b) ¿Cuál es la probabilidad de que John o Sally sean elegidos?

a) P(John) = 1/5 = 0,2 Es una probabilidad sencilla

b) Evento A: John sea elegido
Evento B: Sally sea elegida
P (John o Sally) = P(A) + P(B) = 1/5 + 1/5 = 2/5 = 0,4

Sucesos no excluyentes o solapados:

Dos o más eventos o sucesos son no excluyentes, o conjuntos, cuando es posible que ocurran ambos. Esto no indica que necesariamente deban ocurrir estos eventos en forma simultánea.

Por ejemplo si consideramos en un juego de domino sacar al menos un blanco y un seis, estos eventos son no excluyentes porque puede ocurrir que salga el seis blanco.

Ley de adicción para eventos mutuamente excluyentes:

Si dos eventos no son mutuamente excluyentes, es posible que ambos se presenten al mismo tiempo, en tales casos, debemos modificar la regla de la adición. Por ejemplo, cuál es la probabilidad de sacar un as o un corazón de un mazo de barajas?, los eventos as y corazón pueden presentarse juntos, ya que se puede sacar un as de corazón de un mazo de barajas, cuya fórmula sería:

Instituto Universitario de Tecnología para la Informática –IutepiInstituto Universitario de Tecnología para la Informática - Iutepi

Probabilidad de que A suceda

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

Probabilidad de que A y B sucedan juntos

Probabilidad de que Se presente A o B

Probabilidad de que B suceda

Ejercicios:

1) Se tiene un mazo de barajas, se desea determinar la probabilidad de extraer una carta de as o corazón rojo.

Total de barajas son 52
- Picas = 13
- Corazones Rojos = 13
- Corazones Negros = 13
- Trébol = 13

As, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K

Evento A: extraer un as $P(A) = 4/52$

Evento B: extraer un corazón rojo $P(B) = 13/52$

La probabilidad de que suceda A y B: $P(A \cap B) = 1/52$

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

$P(A \cup B) = 4/52 + 13/52 - 1/52 = 16/52$

2) Los empleados de una cierta compañía han elegido a cinco de ellos para que los representen en el consejo administrativo y de personal sobre productividad. Los perfiles de los cinco elegidos son:
1. hombre de 30 años
2. hombre de 32 años
3. mujer de 45 años
4. mujer de 20 años
5. hombre de 40 años

Este grupo decide elegir un vocero, la elección se efectúa sacando de un sombrero uno de los nombres impresos. De desea determinar la probabilidad de que el vocero sea mujer o cuya edad esté por arriba de 35 años?

Evento A: sea mujer $P(A) = 2/5$

Evento B: mayor de 35 años $P(B) = 2/5$

La probabilidad de que sea mujer y mayor de 35 años $P(A \cap B) = 1/5$

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

$P(A \cup B) = 2/5 + 2/5 - 1/5 = 3/5$

Instituto Universitario de Tecnología para la Informática –IutepiInstituto Universitario de Tecnología para la Informática - Iutepi

Bibliografía:

- Walpole – Myers. Probabilidad y Estadística, para Ingenieros y Ciencias.
- Rafael Valera Ibarra. Estadística Básica
- Alan Waster, de Mc Graw – Hill, estadística para administradores.

Instituto Universitario de Tecnología para la Informática –IutepiInstituto Universitario de Tecnología para la Informática - Iutepi