SECRETARÍA DE EDUCACIÓN PÚBLICA

UNIVERSIDAD TECNOLÓGICA DE TULA-TEPEJI

Organismo Descentralizado
de la Administración Pública del Estado

PROGRAMA EDUCATIVO DE QUÍMICA

Métodos Estadísticos

Unidad I

Estadística Descriptiva

Cuatrimestre: Enero-Abril 2015

ELABORÓ:
ING. PEDRO GARCÍA BERNAL

Tula de Allende, Hidalgo, 2015

Presentación de la Unidad

La palabra estadística frecuentemente remite conceptualmente a gráficas y tablas; cifras relativas a nacimientos, muertes, impuestos, demografía, ingresos, deudas, créditos, etc. No obstante, para aprovechar las herramientas de análisis estadístico, es necesario comprender qué representa cada concepto y la metodología mediante la cual se obtiene un **dato estadístico.**

Definiciones básicas

La estadística es la ciencia cuyo objetivo es reunir información cuantitativa relacionada a individuos, grupos, series de hechos, entre otros. Gracias al análisis de estos datos se pueden deducir algunos significados precisos o algunas previsiones para el futuro. La estadística, en general, es la ciencia que trata la recopilación, la organización, la presentación, el análisis y la interpretación de datos numéricos con el fin de realizar una toma de decisiones más efectiva.

Suele haber una confusión con los términos asociados con las estadísticas, lo cual es conveniente aclarar debido a que esta palabra tiene tres significados: la palabra estadística, en primer término, se usa para referirse a la información estadística y la descripción de parámetros; también se usa para referirse al conjunto de técnicas y métodos que se utilizan para analizar la información estadística; y el término estadístico, en singular y en masculino, se refiere a una medida derivada de una muestra.

Utilidad e importancia

La estadística resulta muy útil no sólo para recopilar y describir datos, sino también para interpretar la información obtenida, que puede ser aprovechada para demostrar la evolución de un fenómeno a través de cierto tiempo.

En México, el Instituto Nacional de Estadística y Geografía (INEGI) se encarga de recabar información estadística y geográfica de todo el país, en diferentes áreas y contextos. Los datos que publica sirven para dar a conocer, a cualquier persona, la situación en la que se encuentra el área de donde se obtuvo la información.

Los métodos estadísticos se utilizan prácticamente en investigaciones de todas las áreas de conocimiento, tanto en el ámbito académico, como en el profesional y laboral; en todos ellos la finalidad es poder resolver un problema – entendiendo que un *problema* queda definido como la diferencia entre lo real y lo deseado –, en donde la estadística muestra la realidad para que el investigador pueda analizar sus deseos y con ello tomar una decisión.

Conceptos básicos de estadística

La estadística, para su mejor estudio, se ha dividido en dos grandes ramas:

Estadística descriptiva

La función descriptiva de la estadística se enfoca en la presentación y clasificación de los datos obtenidos de la población que se analiza, es decir, describe datos.

Estadística inferencial

Esta aplicación de la estadística busca plantear y resolver problemas específicos y/o hacer previsiones a partir de los datos de una muestra, dado que es muy difícil estudiar a la población completa. Esta rama de la estadística infiere a partir de los datos, entendiendo inferir como la estimación de un resultado.

Todo profesional que utilizará la estadística como herramienta debe conocer los siguientes conceptos básicos.

Población

Conjunto de todos los elementos que permiten resolver un problema y que presentan una característica común determinada, observable y medible. Por ejemplo, si el elemento es una persona, se pueden estudiar las características edad, peso, nacionalidad, sexo, etc. Los elementos que integran una población pueden corresponder a personas, objetos o grupos (por ejemplo, familias, las manzanas de una cosecha, empleados de una empresa, etc.).

Muestra

Cuando es difícil estudiar la población debido a su gran tamaño o que provenga de un proceso que no se detiene (como la producción de un bien), se debe analizar un subconjunto o parte de ésta que la represente, llamado muestra, partiendo del supuesto de que este subconjunto presenta el mismo comportamiento y características que la población. En general, el tamaño de la muestra es mucho menor al tamaño de la población. Por ejemplo, a veces se estudian poblaciones enteras: elecciones, censos; otra veces números “pequeños”: los alumnos de una facultad, los habitantes de una ciudad, los miembros de una asociación, etc.; pero otras muchas veces se estudian muestras.

Individuo

Un individuo o unidad estadística es cada uno de los elementos que componen la población. Debes tener en cuenta que un individuo en estadística puede ser distinto a un individuo como persona. Por ejemplo, en los censos económicos se obtienen datos de los negocios. En este caso cada negocio, que está formado por varias personas, es un individuo de la población.

Muestreo

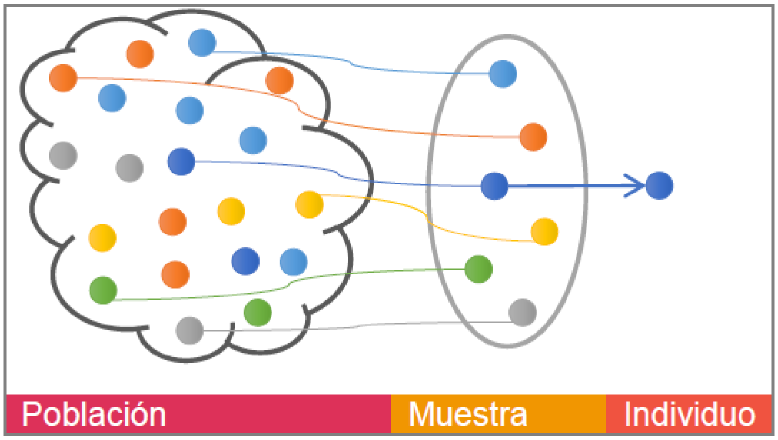
Es el proceso de recabar los datos que se desean analizar, obtenidos de una proporción reducida y representativa de la población.

Dato

El dato es cada uno de los valores que se han obtenido al realizar un estudio estadístico. Por ejemplo: si se lanza una moneda al aire 5 veces obtenemos 5 datos: sol, sol, águila, sol, águila.

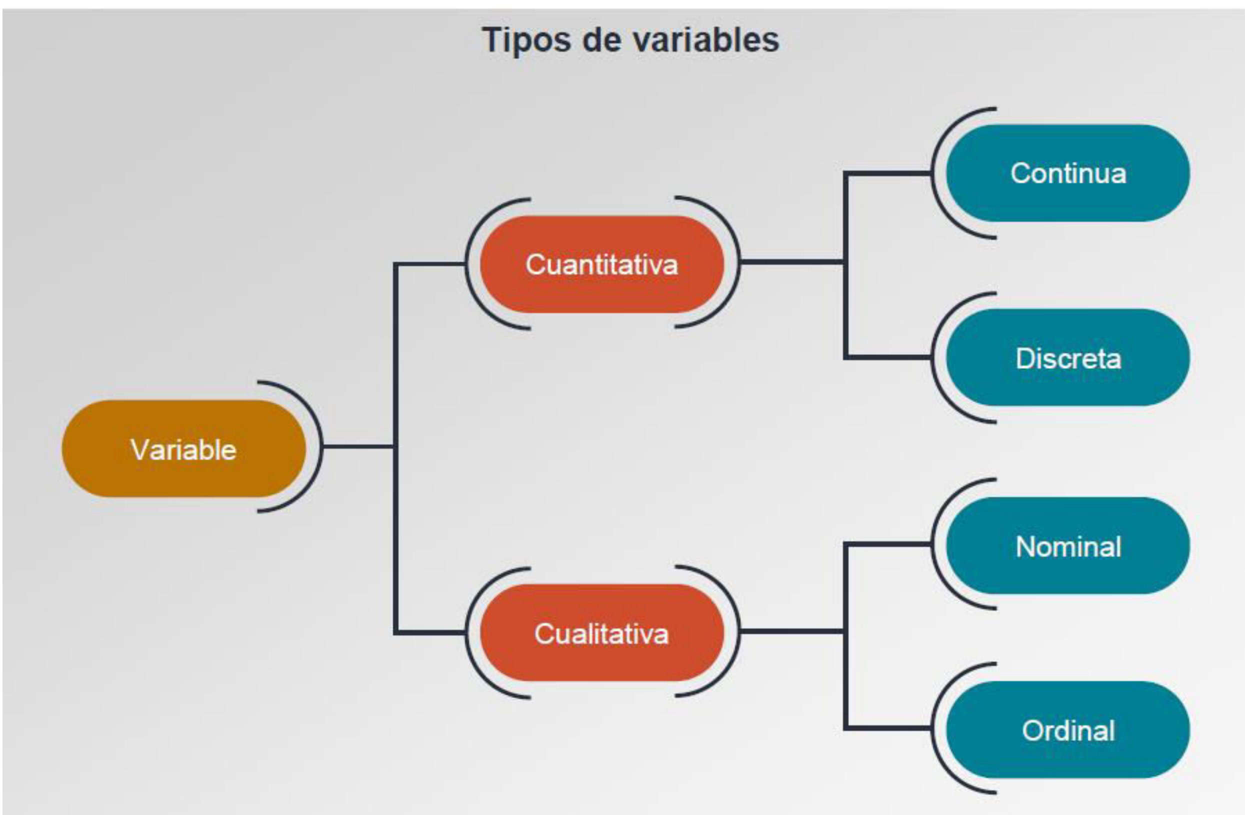
Variable

Se llama variable a una característica que se observa en una población o muestra, la cual se desea estudiar. La variable puede tomar diferentes valores dependiendo de cada individuo y se puede clasificar en cuantitativa y cualitativa.



Tipos de variables

A su vez las variables se dividen en distintos tipos, después del siguiente esquema se ofrece una breve descripción de cada tipo de variable.



Cuantitativa: Variable cuantitativa es aquella que se expresa en valores numéricos. Se subdividen en discreta y continua.
Discreta: Es una variable expresada con valores enteros. Ejemplo: número de hijos de una familia, número de alumnos de un curso, número de empleados en una empresa.
Continua: Es una variable que puede tomar cualquier valor dentro de un intervalo. Ejemplo: peso (de 50 a 57.9 kg), estatura (de 1.70 a 1.76 m), sueldo (de $15,030.40 a $25,299.90).
Cualitativa: Variable cualitativa es aquella que describe cualidades. No son numéricas y se subdividen en nominal y ordinal.
Nominal: Son variables presentadas sin orden ni jerarquía. Ejemplo: estado civil, preferencia por una marca, sexo, lugar de residencia.
Ordinal: Son variables organizadas de acuerdo con una clasificación. Ejemplo: grado de estudios, días de la semana, calidad de la atención, nivel socioeconómico.

DISTRIBUCIÓN DE FRECUENCIAS

Una Distribución o Tabla de Frecuencias es la representación conjunta de los datos en forma de tabla o subgrupo de datos correspondientes a un fenómeno en estudio, y su ordenamiento en base al número de observaciones que corresponden a cada dato o a cada grupo de datos, adecuados según cronología, geografía, análisis cuantitativo o cualitativo. Los principales elementos de una tabla estadística son: Título, unidades, encabezado, cuerpo o contenido, nota de pie y referencias. Se elabora colocando en la primera columna los datos diferentes o subgrupos de datos (llamados clases o intervalos de clase) y en la columna siguiente el número de observaciones que corresponden a cada dato o a cada grupo de datos (llamada frecuencia). Una tabla de este tipo dará, en forma abreviada, una información completa acerca de la distribución de los valores observados. Estas tablas facilitan el uso de los métodos gráficos y aritméticos.
La presentación de los datos en forma ordenada, por medio de una tabla, dependerá de los datos de que se trate, y si estos son cualitativos o cuantitativos como se muestra a continuación:

Datos	Ordenamiento
Cualitativos	Alfabético A – Z
	Alfabético Z – A
	Del más al menos repetido
	Del menos al más repetido
Cuantitativos	Creciente (menor al mayor)
	Decreciente (mayor al menor)

EJEMPLO 1. Se preguntó a un grupo de alumnos de primer año de un colegio en específico, por la asignatura de su preferencia, arrojándose los siguientes resultados:

Asignaturas

Mate Social Taller Quím. Infor Mate Inglés
Mate Quím. Infor Inglés Ética Inglés Social
Inglés Ética Mate Taller Quím. Mate Taller
Social Mate Inglés Infor Inglés Ética Infor
Mate Inglés Infor Ética Quím. Taller Inglés
Social Inglés Ética Taller Infor Quím. Taller
Taller Infor Mate Quím. Infor Mate Infor
Inglés

Distribución de frecuencias

Asignatura	Repeticiones (frecuencias)
Ética y valores	5
Informática	9
Inglés	10
Matemáticas	9
Química	6
Sociales	4
Taller de lectura	7
Total	50

EJEMPLO 2. En la UTTT se realizó un experimento sobre el coeficiente intelectual (C.I.) de sus alumnos, para lo cual aplicó un examen de C.I. a un grupo de 20 alumnos escogidos al azar, obteniendo los siguientes resultados:

119, 109, 124, 119, 106, 112, 112, 112, 112, 109, 112, 124, 109, 109, 109, 106, 124, 112, 112,106.

Toda vez que se tienen los datos, se ordenan de menor a mayor o viceversa.

106, 106, 106, 109, 109, 109, 109, 109, 112, 112, 112, 112, 112, 112, 112, 119, 119, 124, 124, 124

Datos	Repeticiones
106	3
109	5
112	7
119	2
124	3

Frecuencia absoluta, absoluta acumulada, relativa y relativa acumulada.

Frecuencia Absoluta de un dato es el número de veces que se repite ese dato, también se presenta la frecuencia absoluta de un intervalo que se refiere al número de datos que pertenecen a ese intervalo. La denotaremos por *f*.

Frecuencia Absoluta Acumulada: Hasta un dato específico, es la suma de las frecuencias absolutas de todos los datos anteriores, incluyendo también la del dato mismo del cual se desea su frecuencia acumulada. De un intervalo es la suma de las frecuencias absolutas de todos los intervalos de clase anteriores, incluyendo la frecuencia del intervalo mismo del cual se desea su frecuencia acumulada. La denotaremos por f_a. La última frecuencia absoluta acumulada deberá ser igual al número total de datos.

Frecuencia Relativa: De un dato, se obtiene al dividir la frecuencia absoluta de cada dato entre el número total de datos. De un intervalo se obtiene al dividir la frecuencia absoluta de cada intervalo entre el número total de datos. La denotamos por f_r.

Frecuencia Relativa Acumulada: Hasta un dato específico, es la suma de las frecuencias relativas de todos los datos anteriores, incluyendo también la del dato mismo del cual se desea su frecuencia relativa acumulada. De un intervalo es la suma de las frecuencias relativas de todos los intervalos de clase anteriores incluyendo la frecuencia del intervalo mismo del cual se desea su frecuencia relativa acumulada, La denotaremos por f_{ra}. La última frecuencia relativa acumulada deberá ser igual a la unidad.

Construcción de distribución o tabla de frecuencias para datos no agrupados y agrupados.

Datos no agrupados

Datos diferentes: Consideraremos como un dato diferente, a cada uno de los distintos datos que se presentan en la muestra, los denotaremos por x_i . Y al número total de datos diferentes lo denotaremos por *m*.

Datos no Agrupados: Cuando el tamaño de la muestra (n) es finito y el número de datos diferentes es pequeño (consideraremos pequeño k ≤ 10), es fácil hacer un análisis de los datos tomando cada uno de los datos diferentes y ordenándolos tomando en consideración las tablas anteriores.

Asignatura de Preferencia				
x_i	f	f_a	f_r	f_{ra}
Ética y valores	5	5	0.1	0.1
Informática	9	14	0.18	0.28
Inglés	10	24	0.2	0.48
Matemáticas	9	33	0.18	0.66
Química	6	39	.12	0.78
Sociales	4	43	0.08	0.86
Taller de lectura	7	50	0.14	1
Total	50		1	

Coeficiente Intelectual				
x_i	f	f_a	f_r	f_{ra}
106	3	3	0.15	0.15
109	5	8	0.25	0.40
112	7	15	0.35	0.75
119	2	17	0.10	0.85
124	3	20	0.15	1
Total	20		1	

Ahora resulta un poco inoperante el realizar cálculos repetitivos, sobre todo cuando se trata de una infinidad de datos o cuando el tamaño de la muestra es considerablemente grande, por lo que se utiliza el agrupar los datos en subgrupos llamados intervalos o clases.

Datos agrupados

Datos Agrupados: Cuando el tamaño de la muestra es considerable o grande y los datos numéricos son muy diversos (n>15), conviene agrupar los datos de tal manera que permita establecer patrones, tendencias o regularidades de los valores observados. De esta manera podemos condensar y ordenar los datos tabulando las frecuencias asociadas a ciertos intervalos de los valores observados.
Intervalos de Clase: Son los intervalos en los que se agrupan y ordenan los valores observados. Cada uno de estos intervalos está delimitado (acotado) por dos valores extremos que les llamamos límites.

Pasos a seguir para construir intervalos de frecuencia.

1. Determinar la cantidad de intervalos apropiada.
   La selección del número adecuado de intervalos y los límites entre ellos dependen del criterio o experiencia de quien realiza el estudio. Sin embargo, existen reglas empíricas para calcular el número de intervalos; la más empleada es la Regla de Sturges, cuya expresión es: K= 1 + 3.3 Log n
   Dónde: K=Número de intervalos el cual siempre debe ser un número entero. Razón por la cual se deberá redondear el resultado al entero más cercano.
   n= Número de datos.
   Log = logaritmo en base 10.

   Otra regla utilizada es la de Velleman que establece que el número de Intervalos se obtiene de la raíz cuadrada del número de datos; es decir K= $\sqrt{n}$, recomendable para tamaños de muestra pequeños (n< 50)

El número de intervalos determinado mediante cualquier regla se aproxima al valor entero más cercano pero deberá ser responsabilidad de quien realiza el estudio, pudiendo utilizar éste en ocasiones uno menor o mayor al obtenido por cualquier regla, si esto le permite tener intervalos con la misma amplitud. Sin embargo, la mayoría de las reglas subestiman el número de intervalos.

2. Calcular el rango de los datos.
   Llamamos rango al número de unidades de variación presente en los datos recopilados y se obtiene de la diferencia entre el dato mayor y el dato menor. Se representa con la letra *R.*
   R = dato mayor – dato menor.

3. Obtención de la amplitud o anchura que tendrá cada intervalo.
   Se encuentra dividiendo el rango por el número de intervalos. Se representa con la letra A de tal manera que:

$$Ac = \frac{R}{K}$$

4. Construcción de los intervalos.
   Los intervalos de clase son conjuntos numéricos y deben ser excluyentes y exhaustivos; es decir, si un dato pertenece a un intervalo determinado, ya no podrá pertenecer a otro, esto quiere decir excluyentes y además todos y cada uno de los datos deberá estar contenido en alguno de los intervalos, esto les da el valor de exhaustivos.
   Las dos caracteres mencionadas anteriormente se logran construyendo intervalos cerrados por la izquierda y abiertos por la derecha; esto se simboliza a través del uso de corchetes y paréntesis respectivamente. Por razones naturales, el último intervalo será cerrado por ambos extremos.

   El primer intervalo se construye de la siguiente manera: Habrá de iniciar con el dato menor, el cual será el extremo inferior del intervalo; el otro extremo se obtiene de la suma del dato menor y la amplitud, con este mismo valor iniciamos el segundo intervalo, del cual el segundo extremo se encuentra sumando al valor anterior la amplitud y este proceso se repite sistemáticamente hasta completar el total de intervalos indicado por la regla elegida, por ejemplo la de Sturges.

   Los valores extremos o límites de intervalo.
   Los intervalos de clase deben estar definidos por límites que permitan identificar plenamente si un dato pertenece a uno u otro intervalo. Estos límites son los valores extremos de cada intervalo.
   Límite inferior: Es el extremo menor de cada intervalo y lo denotaremos por Li .
   Límite superior: Es el extremo mayor de cada intervalo y lo denotaremos por Ls..
   También será muy útil conocer y calcular la Marca de Clase de cada intervalo: Se refiere al Punto Medio del intervalo y a través del representaremos a todo el intervalo, lo denotaremos por MC y una de las maneras de calcularla es promediando los valores límite de cada intervalo, es decir:

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

EJEMPLO 3. Un grupo de investigadores pertenecientes a la secretaría de seguridad pública, tomó una muestra aleatoria de las velocidades (km/h) registradas por 30 vehículos en el trayecto Hermosillo – Ures, con el fin de establecer nuevos límites máximos de velocidad para una carretera. La muestra arrojo los datos siguientes:

90, 99, 104, 99, 119, 98, 95, 112, 95, 120, 100, 90, 116, 96, 114, 108, 98, 118, 100, 106, 114, 100, 112, 106, 100, 115, 111, 105, 114, 97

Toda vez que se tienen los datos, se recomienda ordenarlos de menor a mayor o viceversa

90, 90, 95, 95, 96, 97, 98, 98, 99, 99, 100, 100, 100, 104, 105, 106, 108, 111, 112, 112, 114, 114, 115, 116, 118, 119, 120

Ahora llevamos a la práctica los pasos descritos anteriormente para la construcción de los intervalos.
Primero obtendremos el número de intervalos que vamos a utilizar, para lo cual empleamos la Regla de Sturges:

$$K = 1 + 3.3\text{Log}(30) = 1 + 3.3\,(1.4771212547) = 1 + 4.87$$
$$= 5.87 \approx 6$$

Segundo, calculamos el rango de variación, R = 120 – 90 = 30

Tercero, obtenemos la amplitud de cada intervalo de clase como sigue:

$$Ac = \frac{30}{6} = 5$$

Finalmente construimos los intervalos, el primero de ellos inicia con 90 que es el extremo inferior que, sumado a 5 obtenemos 95, que será el extremo superior; este extremo será el inferior del segundo intervalo; y al sumar nuevamente la amplitud tendremos 100 que será el extremo superior y así sucesivamente hasta completar los 6 intervalos., que se muestran enseguida:

[90 – 95), [95 – 100), [100 – 105), [105 – 110) [110 – 115) y [115 – 120]

Los corchetes expresan que el valor extremo se incluye en el intervalo y los paréntesis dan a entender que el valor extremo del intervalo no se incluye en el.

Para la construcción de distribuciones de frecuencias contabilizamos el número de datos que le corresponden a cada intervalo; es decir obtenemos las frecuencias absolutas y de estas podemos generar los demás tipos de frecuencias y presentarlas en una tabla de resumen como la que a continuación se muestra:

Distribuciones de frecuencias para las velocidades

Intervalos de Clase	f	f_a	f_r	f_{ra}	m_c
[90 – 95)	2	2	0.07	0.07	92.5
[95 – 100)	8	10	0.27	0.34	97.5
[100 – 105)	5	15	0.17	0.51	102.5
[105 – 110)	4	19	0.13	0.64	107.5
[110 – 115)	6	25	0.20	0.84	112.5
[115 – 120]	5	30	0.16	1.00	117.5
Total	30		1.00		

REPRESENTACIÓN GRÁFICA

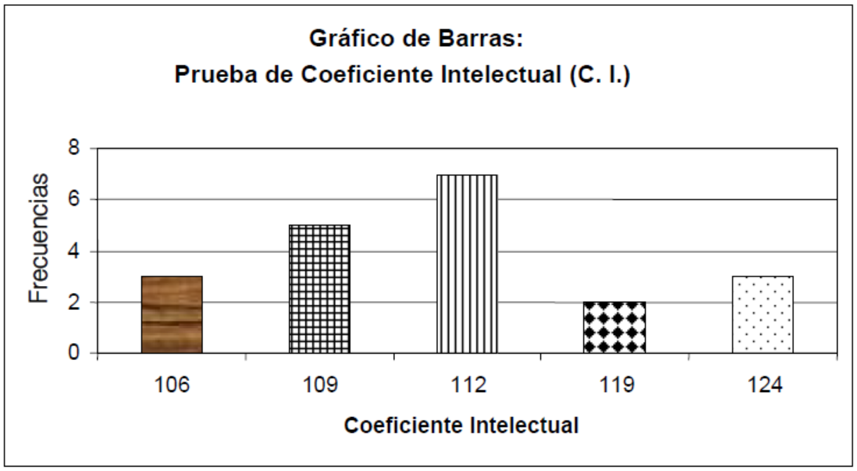
Toda vez que se ha hecho el análisis de frecuencias, existe en estadística, un conjunto de imágenes gráficas, las cuales combinando distintos tipos de colores, sombreados, puntos, líneas, símbolos, números o texto, etcétera, y un sistema de referencia (coordenadas), nos permite la representación en forma más resumida y total del experimento o fenómeno en estudio. Los gráficos son muy útiles como apoyos e incluso sustitutos de las tablas o distribuciones y como una herramienta para el análisis de los datos, lo que los convierte en el medio más efectivo para la presentación, descripción, resumen y análisis de la información

En este curso se abordarán los gráficos estadísticos como un vehículo de presentación y herramienta de análisis que permita observar tendencias presentes en los datos obtenidos al realizar un estudio.

Una manera sencilla de diferenciar los segmentos es sombreándolos de claro a oscuro, siendo el de mayor tamaño el más claro y el de menor tamaño el más oscuro.

Presentación de Datos: Después de la Organización de los datos y su presentación en Tablas Estadísticas, la información contenida en una tabla estadística también se puede presentar mediante gráficas, siendo las más comunes para variables discretas (datos no agrupados) las de: Barras y circulares o de pastel; y para variables continuas (datos agrupados) el histograma, polígono de frecuencias y ojiva. Estos gráficos no son los únicos para la presentación y análisis de datos estadísticos, pero si los más comunes y utilizados.

Gráfica de Barras: es un método gráfico que consta de dos ejes: Uno horizontal, en el que se representan los valores (Eje de los datos) utilizando barras verticales en forma rectangular y de la misma amplitud, y un eje vertical, en el cual la frecuencia representa la altitud que tendrá la barra rectangular (Eje de las frecuencias), las barras van separadas la misma distancia unas de otras y para distinguirlas puede utilizarse distintos colores o entramados según se considere.



Gráfica Circular de Pastel o también llamada del 100%: este gráfico se utiliza fundamentalmente, para representar distribuciones de frecuencias relativas (es decir, porcentajes % o proporciones) haciendo corresponder la medida de la frecuencia relativa con la medida del ángulo en grados; es decir, si el 100 % de los datos son 360º de la circunferencia, a cada 1% le corresponderán 3.6º; así, para obtener la medida del ángulo del sector, multiplicamos la frecuencia correspondiente por 3.6º. Al utilizar este gráfico se aconseja no sobrepasar los 10 elementos, y ordenar los sectores de acuerdo a una de dos formas, ya sea siguiendo el orden que se les dé a los datos o empezando del mayor al menor segmento, iniciando a partir de las 12 horas y en el sentido de las manecillas del reloj. Por último, si el texto que representa cada sector no puede colocarse dentro del mismo, se elabora una leyenda que se coloca fuera del segmento, unidos por una flecha.

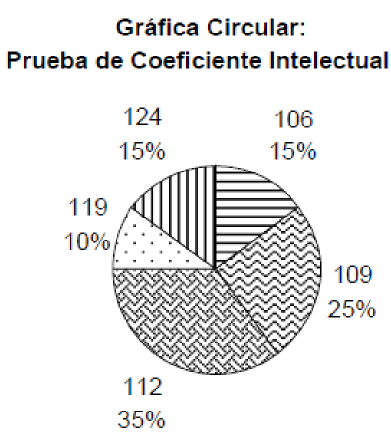


Gráfico de menor a mayor dato

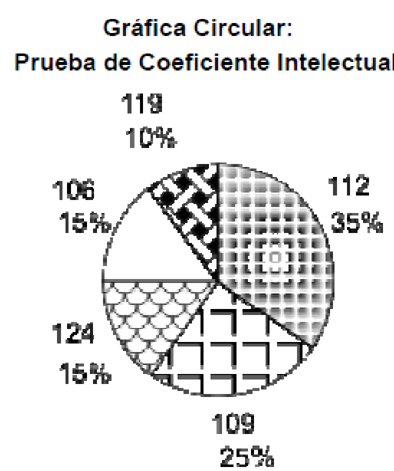
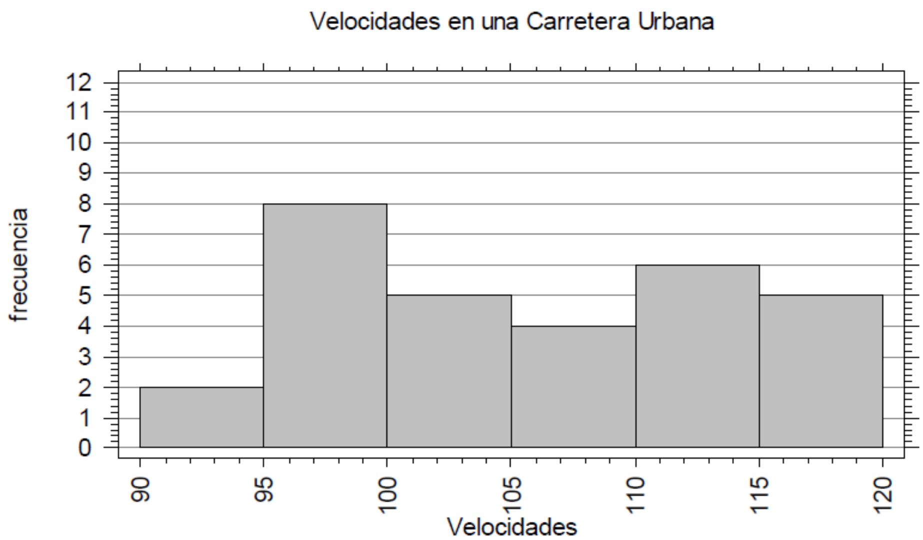
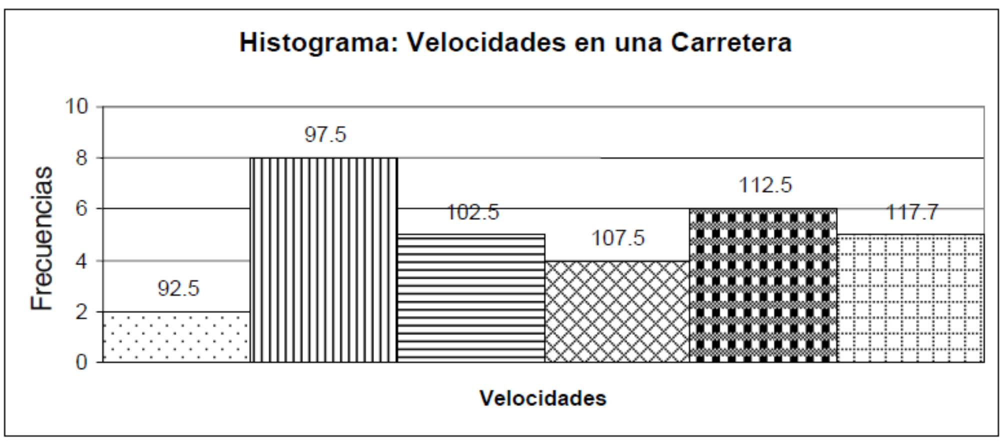


Gráfico circular de mayor a menor porcentaje

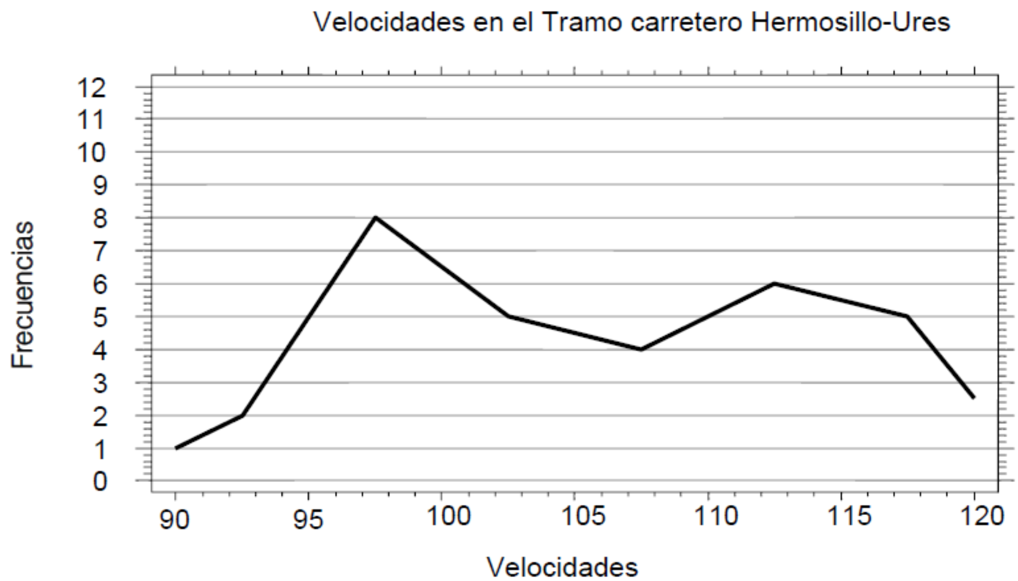
Histograma: es una gráfica en forma de barras que consta de dos ejes, uno horizontal, llamado eje de la variable en observación, en donde situamos la base de una serie de rectángulos o barras contiguas; es decir, que no van separadas, y que se rotula con los límites inferiores de cada clase o intervalo excepto el último que deberá llevar también el límite superior, centradas en la marca de clase. Y un eje vertical llamado eje de las frecuencias, en donde se miden las alturas que vienen dadas por la frecuencia del intervalo que representa. Todos los intervalos deben tener la misma longitud.
Veámoslo a través de un ejemplo, cuando las amplitudes de los intervalos son iguales:

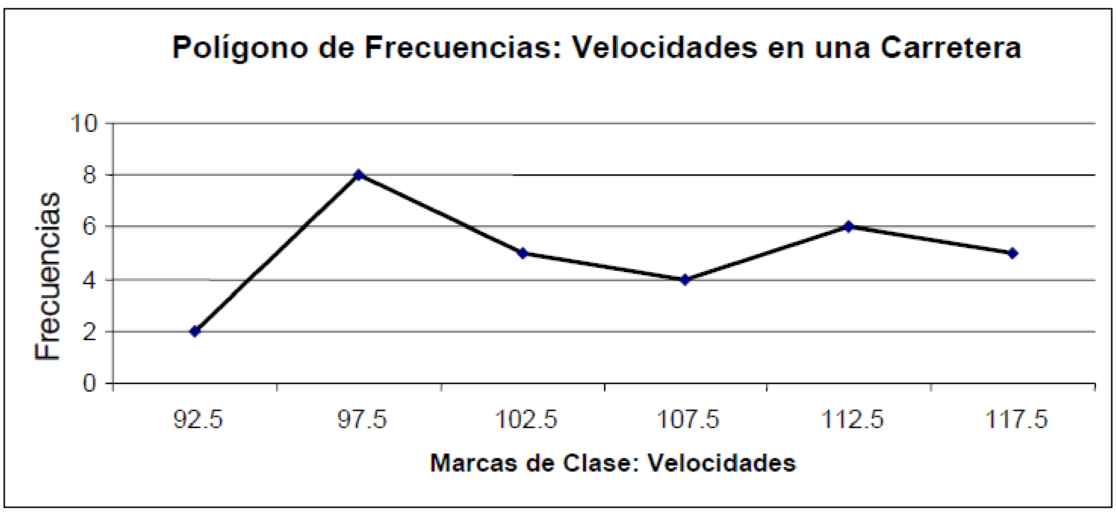




Polígono de Frecuencias: es una gráfica del tipo de las gráficas de líneas trazadas sobre las marcas de clase, (de ahí el nombre de polígono), y se traza uniendo con segmentos de recta, de izquierda a derecha, las parejas ordenadas que se forman, al considerar como abscisa la marca de clase (eje horizontal) y como ordenada la frecuencia del intervalo representado (eje vertical); la primera y última parejas ordenadas se unen mediante un segmento de recta al eje horizontal, con las que serían la marca de clase anterior y posterior respectivamente si estas existieran.

Este tipo de gráfico adquiere mayor importancia cuando se quiere mostrar en un mismo gráfico más de una distribución o una clasificación cruzada de una variable continua con una discreta, situación que no se puede observar en uno de los gráficos presentados anteriormente por la forma de construcción del mismo gráfico.

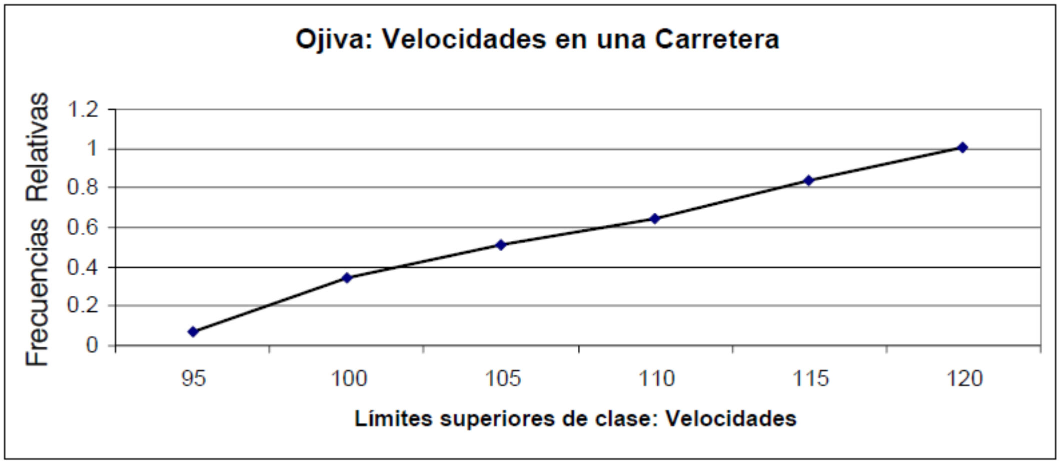




Gráfica de Frecuencias Acumuladas u Ojiva: Es un gráfico que igual al histograma y polígono de frecuencias se utiliza para el análisis y representación de variables continuas, sólo que en vez de utilizar las frecuencias absolutas, por sus características se construye uniendo con segmentos de recta, de izquierda a derecha, las parejas ordenadas que se forman, al considerar como abscisa los límites superiores de cada intervalo (eje horizontal) y como ordenada las frecuencias relativas acumuladas hasta cada intervalo representado (eje vertical).

Existen dos tipos de ojivas, las llamadas de mayor que, iniciando en la frecuencia más alta 1 hacia la más baja 0 y las llamadas de menor que, iniciando en la frecuencia más baja 0 hacia la más alta 1.

El gráfico ojiva representa mayor importancia cuando se trata de comparar las observaciones de una misma característica en dos experimentos distintos, ya que no se puede ejecutar comparaciones sobre frecuencias absolutas, es necesario una comparación sobre frecuencias relativa; además permite ver cuantas observaciones se hallan por arriba o debajo de ciertos valores establecidos.



MEDIDAS DE TENDENCIA CENTRAL Y DISPERSIÓN

Las medidas de tendencia central, de centralización o posición nos facilitan información sobre un conjunto o serie de datos que estamos analizando una vez que estos datos fueron recopilados u organizados ya sea en una investigación documental o en una investigación de campo.

Normalmente, la variable que se intenta medir es conocida en algunas ocasiones de manera insuficiente. Esto no significa que no se tenga algún conocimiento global de valores que pueda asumir, sino que es necesario conocerla mejor para tomar alguna decisión de importancia. Por ejemplo, si se desea comparar las estaturas de alumnos varones de dos planteles de Bachilleres, de antemano se sabe que éstas variables estarán siempre entre 140 cm. y 210 cm. En general, nadie dudaría de un amplio margen tan ancho de estaturas. Sin embargo, este conocimiento no es lo suficientemente preciso para hacer la comparación deseada. Es indispensable afinarlo más para cada uno de los dos planteles, interesa donde están centradas las estaturas, que tanta variabilidad tiene, etcétera. De los muchos aspectos de los datos, que intentamos representar numéricamente con estadísticas, dos son los más importantes:

Medidas de Centralización

- a) Media Aritmética o promedio ($\bar{x}$)
- b) Mediana $(\tilde{x})$
- c) Moda $(\hat{x})$

Medidas de Variabilidad o Dispersión

- a) Rango (R)
- b) Desviación Media (D. M)
- c) Varianza (S^2)
- d) Desviación estándar o típica (S)

Medidas de tendencia central para datos no agrupados
Llamaremos datos no agrupados a los que no aparecen resumidos en distribuciones de frecuencias.

a) Media Aritmética. La medida más evidente que podemos calcular para describir un conjunto de observaciones numéricas es su valor medio. La media no es más que la suma de todos los valores de una variable dividida entre el número total de datos de los que se dispone. Siendo su fórmula la siguiente:

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

Donde:

$\sum$ Símbolo de sumatoria que indica que se deberá sumar todos los valores que toma la variable numérica X.

X Cada uno de los datos obtenidos de la muestra.

n Número total de datos.

Como ejemplo, consideremos 10 alumnos de bachillerato cuyas edades en años son: 14, 15, 16, 18, 17, 14, 15, 15, 18 y 17. La media de edad de estos jóvenes es:

$$\bar{x} = \frac{14+15+16+18+17+14+15+15+18+17}{10}$$

$$\bar{x} = \frac{159}{10} = 15.9 \quad \text{años}$$

b) Mediana

Otra medida de tendencia central o de centralización que se utiliza habitualmente es la mediana. Es el dato o valor equidistante o que se encuentran más en medio de todo el conjunto de datos numéricos, se representa con el símbolo: *x*~

La mediana del ejemplo anterior sería el valor que deja a la mitad de los datos por encima de dicho valor y a la otra mitad por debajo de él, es decir el 50 % por arriba y el 50% por debajo del valor mediana.

Para obtener la mediana para datos no agrupados primeramente deberemos ordenar los datos en forma ascendente o descendente observando la siguiente secuencia de datos:

Por ejemplo:

Los pesos en kg de ocho alumnos de bachillerato son los siguientes:

52, 60, 58, 54, 72, 65, 55 y 76

Ordenación ascendente: 52, 54, 55, 58, 60, 65, 72, 76

En este caso podemos observar que se tienen dos datos centrales a saber el 58 y el 60, la mediana que se ubica en medio, se obtiene del promedio de los dos datos anteriores es decir:

$$\tilde{x} = \frac{58+60}{2} = 59 \text{ kg}$$

Si al ejemplo anterior le agregamos el número 78 que representa el peso de un estudiante entonces la mediana se determinará como el dato u observación que se encuentra en el medio, como ahora el número de datos es impar, entonces la mediana es uno de los datos presentes en la muestra, para este caso, al ordenarlos, tenemos:

52, 54, 55, 58, 60, 65, 72, 76, 78

Entonces la mediana $(\tilde{x})$ es: 60 kg

Si la media y la mediana son iguales, la distribución o conjunto de datos de la variable es simétrica.

La mediana es muy sensible a la variación de los datos; pero menos sensible a los valores extremos. Geométricamente la mediana es el valor de variable (se ubica en el eje horizontal) que corresponde a la vertical que divide al histograma en dos secciones cuya áreas son de igual magnitud.

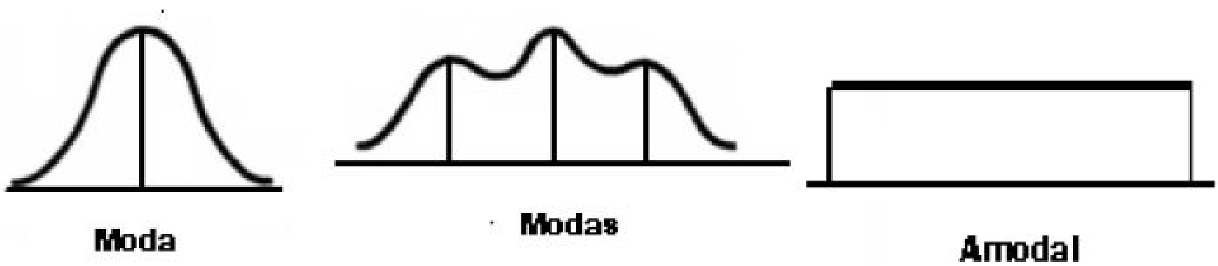
Cuando algunos valores de un conjunto de datos u observaciones son muy grandes o pequeños con respecto a los demás, entonces la media aritmética se puede distorsionar y perder su carácter representativo, en esos casos es conveniente utilizar la mediana como medida de tenencia central.

La moda, representada por el símbolo *x*) o también por (Mo) se suele definir como el valor más frecuente. En el caso de una serie de datos no agrupados, es el valor de la variable que más se repite.
Ejemplos 1: en el caso del ejemplo anterior, 5, 21, 32, 59, 60, 60,61, 64, 71, 80. La moda será:
x) = 60 años. (Unimodal)

Ejemplo 2: determinar la moda del siguiente conjunto de datos 1, 2, 3, 4, 4, 5, 2, 1, 3, 4, 2, 3, 4, 6, 3
x) = 3 y 5 (bimodal)

Ejemplo 3: determinar la moda del siguiente conjunto de datos 1, 2, 3, 4, 5, 6, 7, 8, 9
En este caso, como ningún dato se repite será amodal.

Gráficamente eso se puede reflejar mediante el análisis de un histograma de frecuencias



En el caso en que la distribución o conjunto de datos tenga una moda se dirá que el conjunto de datos es unimodal; si tiene dos modas se llamara bimodal, más de dos modas se le llamará polimodal. y En caso que no tenga ninguna moda se denominara amodal.

Medidas de Variabilidad o Dispersión.

a) Rango(R). Es una medida razonable de Variabilidad llamada también en algunas ocasiones amplitud, representa el número de unidades de variación de los datos numéricos, se obtiene restando el valor más bajo de un conjunto de observaciones del valor más alto.

b) Varianza cuyo símbolo es (S^2) es la media de las desviaciones al cuadrado, calculada usando n o n-1 como divisor, dependiendo si es varianza poblacional o muestral respectivamente. Su expresión es la siguiente:

$$s^2 = \sum \frac{(x - \overline{x})^2}{n-1}$$

Ejemplo: Cinco alumnos obtuvieron las siguientes calificaciones en el segundo examen parcial de Matemáticas Tres: 75, 85, 60, 95 y 85. Determinar la varianza.

$$s^2 = \frac{(75-80)^2 + (85-80)^2 + (60-80)^2 + (95-80)^2 + (85-80)^2}{4} = \frac{700}{4} = 175$$

Las unidades de varianza son cuadráticas, 175 puntos cuadráticos de calificación, no concuerdan con las originales y en ocasiones como esta, resulta un valor muy grande, razones por las cuales se utiliza otra medida de dispersión que veremos enseguida.

c) Desviación típica o estándar cuyo símbolo es (S) La desviación estándar es simplemente la raíz cuadrada positiva de la varianza. Su expresión es:

$$s = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}}$$

En el ejemplo anterior la desviación estándar de las calificaciones es:

$$s = \sqrt{175} \approx 13.22 \text{ puntos}$$

La varianza y la desviación estándar miden la dispersión promedio alrededor de la media; es decir, como las observaciones mayores fluctúan por encima de ésta y como las observaciones menores se distribuyen por debajo de ésta.

Como medidas de variabilidad más importantes, conviene destacar algunas características de la varianza y la desviación estándar o típica.

Medias de tendencia central para datos agrupados.

MEDIA ARITMÉTICA PARA DATOS AGRUPADOS

Para calcular esta medida de centralización o tendencia central se tomaran en cuenta las frecuencias absolutas y la marca de clase de cada clase; mediante la siguiente fórmula:

$$\overline{X} = \frac{\sum fmc}{n}$$

Donde:

$\overline{X}$ = Media aritmética

$\sum$ = Sumatoria

mc= Marca de clase de cada intervalo.

f = Frecuencia absoluta

n= Número de datos (sumatoria de las frecuencias absolutas) de la distribución.

La tabla estadística que a continuación se muestra, resume las estaturas de todos los alumnos de un grupo de quinto semestre:

Estaturas	Frecuencia	mc
[1.52, 1.60)	5	1.56
[1.60, 1.68)	17	1.64
[1.68, 1.76)	14	1.72
[1.76, 1.84)	10	1.80
[1.84, 1.92)	3	1.88

$$\bar{x} = \frac{5(1.56) + 17(1.64) + 14(1.72) + 10(1.80) + 3(1.88)}{49} =$$

$$\frac{7.8 + 27.88 + 24.08 + 18 + 5.64}{49} = \frac{83.4}{49} \approx 1.7$$

MEDIANA PARA DATOS AGRUPADOS.
Para determinar la mediana nos apoyaremos en la siguiente fórmula:

$$\tilde{x} = \hat{L}_i + \left(\frac{\frac{\sum n}{2} - \sum f_a anteriores}{f_{mediana}} \right) \bullet A$$

Donde:

$\hat{L}_i$ = Límite inferior de la mediana

$\sum n$ = Suma total de frecuencias absolutas

$\sum f_a anteriores$ = Suma de todas las frecuencias absolutas que anteceden a la mediana

$f_{mediana}$ = Frecuencia de la mediana

A= Amplitud del intervalo de clase

MODA PARA DATOS AGRUPADOS.

$$\hat{x} = \hat{L}i + \left(\frac{d_a}{d_a + d_p} \right) A$$

$\hat{x}$ = Moda para datos agrupados

$\hat{L}_i$ = Límite inferior modal

d_a= Diferencia anterior

d_p= Diferencia posterior