Universidad Autónoma de Entre Ríos
FCYT Facultad de Ciencia y Tecnología

LICENCIATURA EN SISTEMAS DE INFORMACIÓN

MÓDULO 1

ESTADÍSTICA DESCRIPTIVA

Prof. Natalia M Albert

ÍNDICE

LA ESTADÍSTICA Y SU ÁMBITO

Reseña histórica

El vocablo estadística deriva de la voz latina “status” en sus dos sentidos, como estado político y como situación geográfica.

Los orígenes de esta ciencia son limitados, pues en sus principios se entendía como estadística la exposición oral de los hechos sobresalientes de un estado, en especial lo relativo a cantidades de víveres y número de hombres en edad de ir al frente en épocas de guerra. Es decir, que la palabra Estadística se usa, en un principio, para designar a la ciencia que estudia los asuntos del Estado.

En los primeros tiempos, la estadística se limitaba a la recolección de datos que tenían que ver con alguna cuestión determinada. En la actualidad, en cambio, el concepto de Estadística es mucho más amplio, pues se han ido incorporando nuevas técnicas para el análisis de la información recopilada y a partir de este análisis se infieren posibles resultados futuros que sirven como herramientas auxiliares muy valiosas para otras disciplinas.

El origen de la estadística se remonta al siglo XVIII en que, por un lado los juegos de azar, y por otro la ciencia política, impulsaron los estudios de probabilidad que dieron lugar a la teoría en la que hoy se sustenta la estadística. En dicho siglo, el interés en la descripción numérica de fenómenos relativos a ciudades, provincias, países, etc., dio origen a lo que actualmente se conoce como estadística descriptiva.

La Estadística y la Probabilidad son necesarias y están presentes en la mayoría de las otras ciencias, a las que pretende servir dándoles rigor en sus deducciones y claridad en la presentación de resultados. Por este motivo, esta rama de la Matemática es de fundamental importancia en el análisis de otras disciplinas, como por ejemplo Medicina, Ingeniería, Geografía, Economía, etc.

La Estadística brinda un conjunto de métodos sumamente útiles en la investigación, pues tiene mucho que ofrecer al investigador en la planificación, análisis e interpretación de resultados de su tarea.

La estadística descriptiva, como se ha dicho anteriormente consiste en la presentación de datos en tablas y gráficos y la determinación de parámetros característicos de un acontecimiento o experimento, como puede ser, por ejemplo su media o promedio, su valor máximo o mínimo, etc.

Hoy, la estadística va mucho más allá de brindar una herramienta para organizar, mostrar e interpretar datos, es la estadística inferencial la que se ha desarrollado con mayor ímpetu en las últimas décadas y permite extraer conclusiones tomando como base muestras que representan a una población de la cual se investiga una determinada característica. Los resultados obtenidos en el análisis de la muestra se extienden a la población y esta generalización permite predecir comportamientos futuros, lo que está íntimamente ligado a la toma de decisiones. Esto último implica un riesgo, riesgo que debe ser conocido y evaluado, ya que puede generar decisiones incorrectas.

Definición de Estadística

Es factible elaborar varias definiciones de lo que es la Estadística. Por ejemplo, podemos decir:

- Estadística es todo cúmulo de datos cuantitativos relativos a un conjunto de individuos que presentan algún atributo en común o están afectados por causas comunes.

- Por Estadística entendemos aquellos métodos especialmente adecuados para dar significado a un conjunto de datos, afectados por el azar, es decir que no responden a leyes físicas o matemáticas, usando instrumentos de la Matemática.

- Estadística es la recopilación, presentación, análisis e interpretación de datos numéricos extraídos de un conjunto de individuos, que nos permiten formular conclusiones válidas, efectuar decisiones lógicas basadas en dicho análisis y extender los resultados desde un grupo pequeño hacia una población.

La primera acepción del término se corresponde con las “series estadísticas”, las que surgen al hablar de estadísticas de producción, de precios, demográficas, etc. Estas series, una vez elaboradas, sirven para describir al grupo en análisis y constituyen la fase previa para la toma de decisión y la posterior actuación. Es éste el sentido con el que la palabra Estadística se utilizó en épocas remotas. Constituía en dicha época una herramienta muy importante para el desarrollo de la política, sobre todo la económica empresarial y la macroeconómica.

La segunda acepción equivale a la expresión actual “Estadística Matemática”, la cual, apoyándose en el Cálculo de Probabilidades, estudia las leyes de comportamiento de aquellos fenómenos que, no estando sometidos a leyes rígidas, dependen del azar.

En una segunda etapa se generalizan dichas leyes y, basándose en ellas, predice e infiere resultados.

La última acepción engloba los métodos utilizados por la “Estadística descriptiva” y también por la “Estadística Inductiva o Inferencial”, que trataremos más adelante.

En este capítulo trataremos las cuestiones más vistosas de esta rama de la Matemática: las tablas y gráficos estadísticos.

Actualmente es muy común ver un informe de cualquier actividad desarrollada por el hombre acompañado de una tabla que permite visualizar con claridad los resultados obtenidos y un gráfico que muestra con diferentes grafismos y colores la incidencia de los mismos.

Un periodista enriquece su labor con gráficos que exponen los resultados, por ejemplo, de una encuesta; un empresario exhibe el estado financiero de su empresa; los políticos describen el avance de su campaña, etc.

En cualquier diario o periódico, en vez de exponer al público una gran cantidad de datos, los que se vuelven densos y difíciles de ser leídos e interpretados, se publican los mismos ordenados y recopilados en tablas que ayudan a la lectura y comprensión de los gráficos adjuntos.

Según definimos, la Estadística es una ciencia que utiliza conjuntos de datos numéricos para obtener a partir de ellos inferencias basadas en el cálculo de probabilidades. Trabajando con dichos datos el estadístico, usando técnicas apropiadas, trata de simplificar al máximo la información disponible para que sea clara y útil.

Estadística descriptiva e inferencial

De acuerdo a lo expuesto en el apartado anterior, la Estadística presenta dos aspectos:

- La Estadística se refiere, en un sentido técnico, a una rama de la Matemática aplicada que se ocupa de interpretar los resultados numéricos, elaborar tablas y gráficos explicativos y, posteriormente inferir parámetros estadísticos que caracterizan al conjunto total de datos recolectados. El conjunto de estas técnicas se denomina **Estadística Descriptiva**.

- Por otro lado la Estadística es la ciencia que estudia a los fenómenos regidos por el azar, midiendo los riesgos y llevando a valores numéricos los resultados de hechos aleatorios, observando el comportamiento de una muestra y generalizando conclusiones a toda una población . Estas conclusiones fundamentan la toma de decisiones para que las mismas sean las acertadas y minimizar así los riesgos de equivocarse. Fundamentalmente, trata de la generalización hacia las poblaciones de los resultados obtenidos en las muestras y de las condiciones bajo las cuales estas conclusiones son válidas. El conjunto de estas técnicas se llama **Estadística Inferencial**.

EL LENGUAJE DE LA ESTADÍSTICA

Entre los métodos estadísticos encontramos la llamada Teoría de Muestras. Esta teoría se propone establecer resultados válidos para una población numerosa partiendo de la observación del comportamiento de una parte de la misma, generalmente pequeña, llamada muestra, la que bajo ciertas condiciones, resulta representativa de las características de la población en estudio (de todas formas el problema de la representatividad de las muestras merece un cuidado especial en cada caso)

Definiremos a continuación muchos de los términos usados en Estadística para precisar algunos de los más frecuentes:

- Población:

> “Se denomina población al conjunto de **todos** los elementos que cumplen una determinada característica, que deseamos medir o estudiar.”

Ejemplos:

1. Si se pretende estudiar la intención de voto de los españoles, la población será el conjunto de todos los españoles con derecho al voto.
2. Si se desea estudiar el peso de los vacunos de 3 años de la provincia en Entre Ríos, la población está compuesta por todos los vacunos de dicha provincia.
3. Si se quiere investigar la producción lechera de la Pampa Húmeda, la población estará compuesta por todos los tambos productores de leche de esta zona.
4. Si se desea conocer el nivel de satisfacción de los afiliados a una obra social, la población está formado por todos sus afiliados.
5. Si se quiere determinar la vida útil de un lote de lámparas producidas por un turno en una fábrica, la población está constituida por todas las lámparas del lote.

- Muestra:

“Se denomina muestra a cualquier subconjunto de la población.”

Ejemplos:

1. Si queremos conocer la opinión sobre las preferencias deportivas de los 743 alumnos de un colegio puede elegirse una muestra de 40 alumnos, por ejemplo 10 alumnos de cada curso, seleccionados por sorteo.
2. Para estudiar el peso de los vacunos antes mencionados podemos seleccionar al azar 200 vacunos de distintos sectores de la región.
3. En el caso de las lámparas, la muestra serán un conjunto de 10 ó 20 luminarias extraídas al azar del lote.

- Unidad estadística:

“En Estadística se considera unidad estadística a cada individuo de una población.”

Ejemplos:

1. En el caso antes planteado cada vacuno es una unidad estadística.
2. Si se estudiara el salario del personal de una empresa, la población y la muestra serían el personal completo de la empresa y cada empleado es una unidad estadística. Si el personal ascendiera a un número muy grande, podría procederse a hacer un muestreo.
3. En el ejemplo de las lámparas, cada lámpara es una unidad estadística.

- Carácter estadístico:

“Cada una de las propiedades o aspectos que pueden estudiarse en los individuos de una población recibe el nombre de carácter o estadístico”

Ejemplos:

1. En el caso de los vacunos antes examinado, el peso de cada vacuno es el carácter estadístico.
2. En el caso de los salarios de los empleados de una empresa, el monto de cada salario es el carácter estadístico.
3. Para el lote de lámparas, la duración de su vida antes de fallar es el carácter analizado.

Su clasificación es:

1) Carácter estadístico cualitativo o atributo: es aquel no susceptible de ser medido ni contado. Por ejemplo, estado general de un animal, la nacionalidad, profesión o sexo de una persona, la opinión sobre un producto.

Consideremos un carácter cualquiera, como por ejemplo el "gusto". Este carácter, al ser observado por un individuo, puede presentar cuatro posibilidades, es decir, es posible percibir cuatro sensaciones diferentes: dulce, amargo, salado y ácido. Las distintas categorías de un atributo se llaman modalidades. Por ejemplo si estudiamos el atributo "estado general de un vacuno", las modalidades pueden ser "bueno" o "malo". Si estudiamos una población en relación con el atributo nacionalidad, podemos considerar dos modalidades, "argentinos" o extranjeros".

Otros ejemplos son: las razas, los tipos de clima, los idiomas, las preferencias, etc.

Las diversas modalidades de un carácter deben cubrir todas las posibilidades que éste puede presentar y deben ser exhaustivas. Es decir que un individuo no puede presentar a la vez más de una de ellas y además debe presentar alguna de ellas.

Por lo tanto, al estudiar algún carácter, como por ejemplo la raza, se deberán considerar todas las posibles modalidades del carácter (todas las posibles razas), con objeto de poder clasificar a todos los individuos que se observen.

En el proceso de medición de estas variables, se pueden utilizar dos escalas:

Escalas nominales: ésta es una forma de observar o medir en la que los datos se ajustan por categorías que no mantienen una relación de orden entre sí (color de los ojos, religión, profesión, presencia o ausencia de un factor de riesgo o enfermedad, etcétera).

Escalas ordinales: en las escalas utilizadas, existe un cierto orden o jerarquía entre las categorías (grados de fatiga, estadio de un tumor, nivel de conocimientos en una disciplina, estado sanitario de una población, etcétera.)

2) <u>Carácter estadístico cuantitativo</u>: es el que surge de un proceso de medición o conteo.

Se distingue:

- <u>Carácter estadístico discreto y continuo:</u>

> “El carácter estadístico es **discreto** sólo si entre dos valores consecutivos del mismo no puede existir un valor intermedio. Esto indica que su recorrido es un conjunto definido en el campo de los $\mathbb{N}$ o $\mathbb{Z}$.”

> “El carácter estadístico es **continuo** si entre dos valores cualesquiera de su recorrido puede existir siempre uno intermedio. Esto indica que su recorrido es un intervalo incluido en el conjunto de los números reales.”

Razonemos un poco ¿???!!!!!

A continuación se dan varios ejemplos de variables estadísticas, ¿puede usted clasificarlas?

1) Se estudiaron 30 variedades de maíz híbrido, clasificando a los granos de la siguiente manera: semi-dentado; liso-semi-dentado, dentado y liso

....

2) Se estudió el número de cerrojos rechazados en cada lote de producción

....

3) En un equipo deportivo se tomaron 10 jugadores al azar y se les midió su estatura

....

4) En un examen de selección se analizó el nivel de conocimiento de inglés de los aspirantes clasificándolo en:
excelente, muy bueno, bueno, regular y malo

....

5) La población de una región se clasificó por su religión

....

6) Se realizó un conteo de hojas enfermas en plantas florales de un invernáculo

....

7) Para abonar el salario familiar en una institución educativa se clasificó a personal según el número de hijos en edad escolar

....

8) Se analizó el rendimiento en quintales por hectárea de la producción de soja

....

SERIES ESTADÍSTICAS O DISTRIBUCIONES ESTADÍSTICAS

Se denominan así al ordenamiento sistemático de los datos recolectados. Se clasifican en:

a) Series cronológicas: el ordenamiento se realiza teniendo en cuenta el tiempo. Por ejemplo: producción de trigo en la República Argentina entre los años 1977-1987.

b) Serie espacial o geográfica: en este caso el fenómeno en estudio es estudiado a través del territorio o del espacio. Por ejemplo: índice de delitos en las distintas provincias argentinas en el año 1999.

c) Serie o distribución de frecuencias: es la correspondencia que hay entre cada valor de la variable (que se denomina solamente variable para la variable discreta y clase para variables continuas)y su respectivo número de observaciones o frecuencias.

Dado un universo o población, si a cada unidad estadística le consideramos un atributo, sea éste susceptible de medición o de diferenciación y establecemos una relación entre los distintos valores de este atributo (X) y el número de individuos que lo poseen, es decir su frecuencia (f), obtenemos una serie de frecuencias.

Ejemplos:

1) Carácter estadístico cualitativo: Nivel de conocimiento de inglés de 100 ingresantes a una Universidad. (Escala ordinal)

Nivel	f_i
Bueno	62
Regular	25
Malo	13
Total	100

2) Carácter estadístico cuantitativo discreto: número de hijos, en edad escolar, por obrero en cierta institución.

Número de hijos: X	Número de familias f
0	8
1	13
2	20
3	30
4	5
	$\Sigma = 76$

3) <u>Carácter estadístico cuantitativo continuo</u>: tiempo de supervivencia de los peces en agua contaminada.

Tiempo en horas: X	f
$x < 300$	20
$300 \leq x < 310$	60
$310 \leq x < 315$	105
$315 \leq x < 320$	10
$x \geq 320$	5
	$\Sigma = 200$

Razonemos un poco ¿???!!!!!

Analicemos un artículo extraído del diario LA NACIÓN del día 20 de febrero de 2000 en el cual se puede leer:

“En los últimos años, la cantidad de usuarios de Internet en el país creció rápidamente. La explosión que tuvo La Red en tan sólo cinco años no se compara con los sesenta que le llevó a la radio y los quince, a la televisión, para alcanzar los sesenta millones de usuarios que hoy posee Internet a nivel mundial.”

Además se predice:

“Las transacciones comerciales crecerán un 149 % anual para el año 2003; los analistas pronostican el éxito para inversores con una presencia temprana en la Red.”

En dicho artículo se presenta una tabla de la cual podemos extraer los siguientes datos:

VENTAS en la WEB *(en millones de dólares)*

	1998
ARGENTINA	12,8
BRASIL	93,0
CHILE	6,0
COLOMBIA	7,7
MÉXICO	21,8
VENEZUELA	5,6
OTROS	19,7
TOTAL de AMÉRICA LATINA	166,6

¿Qué tipo de serie es ésta? ¿Por qué?

El mismo artículo muestra también el crecimiento real y estimado en la República Argentina de comercio electrónico:

VENTAS en la WEB *(en millones de dólares para nuestro país)*

1997	3,3
1998	12,8
1999	48,7
2000	149,2
2001	355,6
2002	699,2
2003	1223,5

¿Cómo clasificaría esta serie? ¿Por qué?

ESQUEMA DE REALIZACIÓN DE UN TRABAJO ESTADÍSTICO

La Estadística es una herramienta de gran aplicación en la práctica diaria y en general para abordar un problema e iniciar su investigación deben seguirse los siguientes pasos:

1 ESPECIFICACIÓN DEL PROBLEMA

En esta etapa se deben:

- Establecer con precisión el problema a tratar: si el problema no puede definirse y acotarse con claridad, existen muy pocas probabilidades de resolverlo.
- Definir los objetivos del trabajo: estos pueden contener la forma en que las preguntas deben ser contestadas, la hipótesis que se va a probar o los efectos que se desean estimar. Los objetivos deben detallarse con exactitud, ya que serán la base para la organización del plan experimental.
- Seleccionar el material experimental: es decir cuál es la población en estudio y la muestra que se seleccionará, así como también la confección del cuestionario o el diseño del experimento.

2 RECOLECCIÓN Y ORDENACIÓN DE DATOS

Como la Estadística trata de la interpretación de información numérica recopilada, lo primero que se deben obtener son esos datos, (la temperatura de nuestra ciudad, los precios de artículos de consumo, etc.).

Para recopilar los datos es necesario proceder con el mayor orden posible y, por lo tanto tener en cuenta los siguientes aspectos:

- Fijar los procedimientos para realizar el experimento. Por ejemplo, en el caso de aplicar un cuestionario, no es lo mismo entrevistar a los individuos en la calle que en los domicilios; en la experimentación se deben señalar las pautas para la ejecución de las tareas de campo o de laboratorio.
- Tener a disposición todos los elementos requeridos para recoger dichos datos, como pueden ser: libretas de campos, elementos de medición, tablas o grillas para ser completadas, formularios pre - impresos, etc.
- Examinar el tipo de datos requeridos, es decir si son cuantitativos o cualitativos.
- Disponer los datos en forma creciente o decreciente, según convenga, para que sean de fácil ubicación y análisis.
- Encontrar el rango de variación de los datos recolectados para hallar entre qué valores máximos y mínimos se hallan comprendidos.

3 ORGANIZACIÓN DE DISTRIBUCIONES DE FRECUENCIAS

Supongamos que se dispone de un registro de datos correspondiente al relevamiento de 100 hogares sobre diferentes aspectos, tales como el número de hijos vivos; superficie de vivienda, sus ingresos mensuales, etc. realizado en una determinada zona del país durante el año 1999.

De los datos recopilados se desea obtener información tal como valor máximo y mínimo, valor promedio, valor más frecuente, etc.

Para facilitar el trabajo de búsqueda y no tener que repetirlo cada vez que analicemos un aspecto del registro, es conveniente organizarlos de alguna manera sistemática; es decir de mayor a menor o viceversa. Este ordenamiento, según vimos en el punto anterior agiliza la búsqueda, pero en el caso que analizamos ahora, en el cual el número de datos es muy grande, deja de tener utilidad. Cuando se dispone de una gran cantidad de valores recopilados es conveniente condensar, simplificar o resumir la totalidad de las observaciones.

Este es el objetivo de las distribuciones de frecuencias, las que presentan los datos mediante tablas o cuadros que se conocen con el nombre de tablas de frecuencias.

Si en nuestro universo establecemos una relación entre un atributo susceptible de ser medido o diferenciado X_i y el número de individuos de nuestro universo que lo posee f_i encontramos una distribución de frecuencias.

Distribución de frecuencias para la variable aleatoria cualitativa (atributo)

La distribución de frecuencias para un atributo muestra el número de observaciones para cada una de las clases o categorías del mismo.

Para el ejemplo que muestra el nivel de conocimiento de inglés de 100 ingresantes a una Universidad, se confecciona una tabla en la que se muestran:

- Frecuencias absolutas: f_i número total de observaciones que pertenecen a cada clase o categoría.

- Frecuencias relativas: f_{i_r} relación entre la frecuencia absoluta de cada modalidad y el número total de observaciones. $f_{ir} = \dfrac{f_i}{n}$; siendo ***n*** el número total de casos.

La frecuencia relativa es para las series de frecuencias un concepto análogo al de probabilidad en las variables aleatorias discretas.

- Frecuencias relativas porcentuales: $f_{i_r}\%$ expresión porcentual de la frecuencia relativa.

Para nuestro ejemplo, que muestra el nivel de conocimiento de inglés de 100 ingresantes a una Universidad:

Nivel	f_i	f_{ir}	$f_{ir}\%$
Bueno	62	0,62	62
Regular	25	0,25	25
Malo	13	0,13	13
Total	100	1	100

Distribución de frecuencias para la variable aleatoria cuantitativa discreta

La distribución de frecuencias para una VAD muestra el número de observaciones para cada uno de los valores de la misma.

Para el ejemplo que analiza el número de hijos por obrero en una institución que antes definimos es posible confeccionar una tabla en la que se muestran:

- Frecuencias absolutas, frecuencias relativas y frecuencias relativas porcentuales: según lo definido anteriormente.

- <u>Frecuencias acumuladas</u>: F_k es la suma de las frecuencias absolutas hasta un determinado valor de la variable inclusive. $F_k = \sum_{i=1}^{k} f_i$
- <u>Frecuencias acumuladas relativas</u>: $F_{k\,r}$ es la relación entre las frecuencias acumuladas y el número total de casos. $F_{kr} = \frac{F_k}{n}$
- <u>Frecuencias acumuladas relativas porcentuales:</u> $F_{k_r}\%$ expresión porcentual de la frecuencia relativa acumulada.

Para el ejemplo considerado, es decir el número de hijos en edad escolar de los obreros de una Institución:

X	f_i	f_{ir}	$f_{ir}\%$	F_k	F_{kr}	$F_{kr}\%$
0	8	0,105	11	8	0,105	11
1	13	0,171	17	21	0,276	28
2	20	0,263	26	41	0,539	54
3	30	0,395	40	71	0,934	93
4	5	0,066	6	76	1	100
	Σ = 76	1	100			

<u>Distribución de frecuencias para la variable aleatoria cuantitativa continua</u>

Como la VAC puede tomar un sinnúmero de valores entre dos cualesquiera de su recorrido, ésta se presenta en una distribución de frecuencias agrupada en intervalos llamados intervalos de clase; de manera tal que cada valor individual estará contenido en alguno de ellos.

Definiremos a continuación los elementos necesarios para confeccionar una tabla de frecuencias para el caso de VAC.

- <u>Rango</u>: se denomina así a la diferencia entre el mayor y menor valor observado de la serie de datos.

$$R = X_{max} - X_{min}$$

El intervalo $X_{max} - X_{min}$ se llama recorrido de la VAC.

El rango se divide en varios intervalos conocidos como intervalos de clase.

- Intervalos de clase: se llama de esta manera a cada uno de los subintervalos en que queda dividido el recorrido de la VAC.

- Número de intervalos de clase: el número óptimo de intervalos de clase depende de la cantidad de observaciones realizadas, aunque no existe un criterio o regla para fijarlo. Solamente se puede decir que la cantidad de intervalos no debe ser tan elevada como para que se anule el objetivo de resumir o condensar datos, ni tan pequeña que se pierda información valiosa. En la generalidad de los casos se sugiere no trabajar ni con menos de 5 ni con más de 20 intervalos.

Empíricamente se han determinado fórmulas que dan un indicio del número óptimo a tomar, siendo ellas:

"$K = \sqrt{N}$" ò "$K = \sqrt{n}$" siendo ***N*** el número total de casos de la población y ***n*** el número total de individuos en una muestra extraída de una población.

Fórmula de Sturger: "$K = 1 + 3{,}3.\log N$"siendo ***N*** el número total de casos de la población.

- Amplitud o tamaño de clase: en una distribución de frecuencias los intervalos pueden tener todos la misma amplitud o distinta. En general conviene, para simplificar cálculos, que los intervalos tengan todos la misma amplitud; aunque en muchos casos la distribución en intervalos de diferente amplitud da una idea más clara del suceso.

La amplitud se obtiene efectuando el cociente entre el rango y el número de intervalos considerados.

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

- Límites de clase: son los extremos de un intervalo de clase. La determinación de la cantidad de cifras significativas de los mismos depende de los valores alcanzados por la variable, pudiendo o no coincidir el límite superior de una clase con el inferior del siguiente.

Es conveniente que ninguno de los valores de la variable coincida con uno de los extremos, pero si eso llega a suceder se debe indicar expresamente sí se lo incluye como límite superior o inferior.

- Límites reales de clase: *(LRI - LRS)* el límite real superior de una clase es igual a la semisuma del límite superior de dicha clase más el inferior de la siguiente. El límite real inferior de una clase es la semisuma del límite inferior de una clase y el superior de la anterior.

- Marca de clase: (M_i) es el punto medio de cada intervalo de clase y se obtiene efectuando el promedio de los límites de clase o de los límites reales de clase.

La marca de clase es considerada como el valor representativo del conjunto de datos incluidos dentro de cada intervalo de clase. Esto indica que al considerar a la marca de clase como un valor típico de cada intervalo se supone que todas las observaciones de un intervalo han tenido un mismo valor, el correspondiente a la marca de clase.

Ejemplos:

Se dan a continuación los datos recopilados en la medición de 110 niños entre 0 y 6 meses de vida. La VAC identifica la talla en centímetros de los niños.

El valor máximo es de 83,5 cm y el mínimo de 58,5 cm siendo, por lo tanto, el rango de R = 25 cm.

El número de intervalos de clase es de : $K = \sqrt{110} \cong 10$ y la amplitud c = 25/10 = 2,5 cm.

Adoptamos, para facilitar las cuentas c = 3 cm.

La tabla de frecuencias resulta:

X_i	M_i	f_i	f_{ir}	$f_{ir}\%$	F_k	F_{kr}	$F_{kr}\%$
56-59	57,5	3	0,0273	2,73	3	0,0273	2,73
59-62	60,5	4	0,0364	3,64	7	0,0636	6,36
62-65	63,5	13	0,1182	11,82	20	0,1818	18,18
65-68	66,5	18	0,1636	16,36	38	0,3454	34,54
68-71	69,5	27	0,2454	24,54	65	0,5909	59,09
71-74	72,5	20	0,1818	18,18	85	0,7727	77,27
74-77	75,5	10	0,091	9,10	95	0,8636	86,36
77-80	78,5	8	0,073	7,30	103	0,9363	93,63
80-83	81,5	5	0,045	4,50	108	0,9818	98,18
83-86	84,5	2	0,018	1,80	110	1	100
Σ		110	1	100			

4 PRESENTACIÓN DE DATOS POR MEDIO DE GRÁFICOS

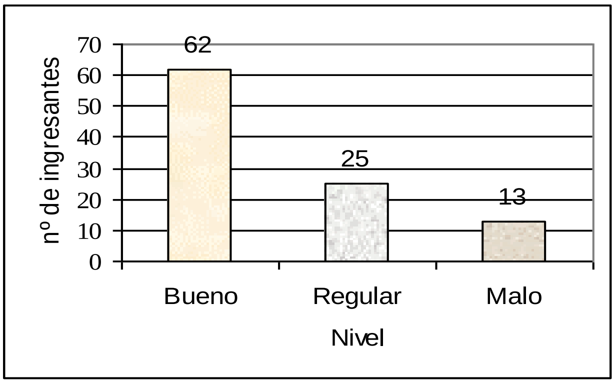
La claridad que requiere un gráfico implica elaborarlo con precisión y seleccionar el tipo más conveniente. Una gráfica sencilla, atractiva y bien trazada, que presente un número limitado de datos y brinde un resumen de la información recopilada es más fácil de comprender que un cuadro. Así como cada gráfico tiene sus ventajas, posee también desventajas, tales como la posibilidad de expresar los datos solamente en forma aproximada y brindar una idea general de la situación pero no sus detalles. Existen diferentes clases de gráficos:

- Diagrama de barras

Se llama diagrama de barras al gráfico que asocia a cada valor de la variable una barra, generalmente vertical, proporcional a la frecuencia con que se presenta.

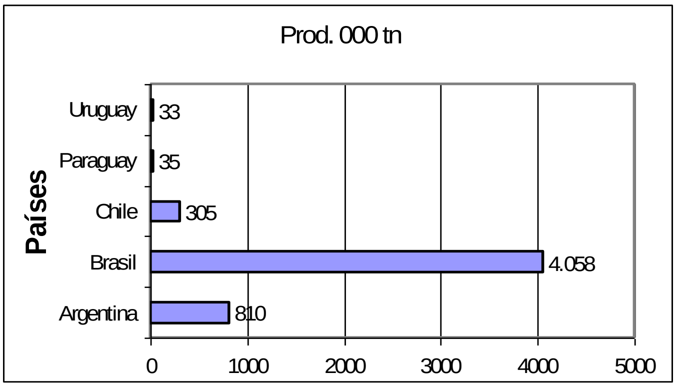
Para la VA cualitativa resulta:

Nivel	f_i
Bueno	62
Regular	25
Malo	13
Total	100



Si se hace diagramas sobre datos geográficos, generalmente su usan barras horizontales, pues de esta manera se pueden incluir más detalles. En el siguiente ejemplo se brindan datos sobre la producción avícola, en toneladas, de los países del Mercosur-1996:

Pollos de engorde	*Prod.* 000 *tn*
Argentina	810
Brasil	4.058
Chile	305
Paraguay	35
Uruguay	33
Total	5.241



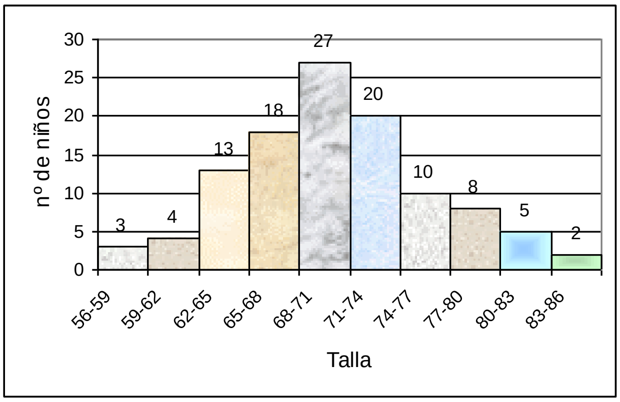
• Histogramas

Los histogramas son diagramas de barras para las variables continuas, es decir para las que se agrupan en intervalos de clase.

Un histograma es un conjunto de rectángulos cuyas bases coinciden con el tamaño de clase, sus puntos medios son las marcas de clase y sus extremos son los límites reales de los intervalos. En general las alturas corresponden a la frecuencia absoluta de cada intervalo, aunque en algunas ocasiones la frecuencia absoluta se relaciona con el área de cada rectángulo.

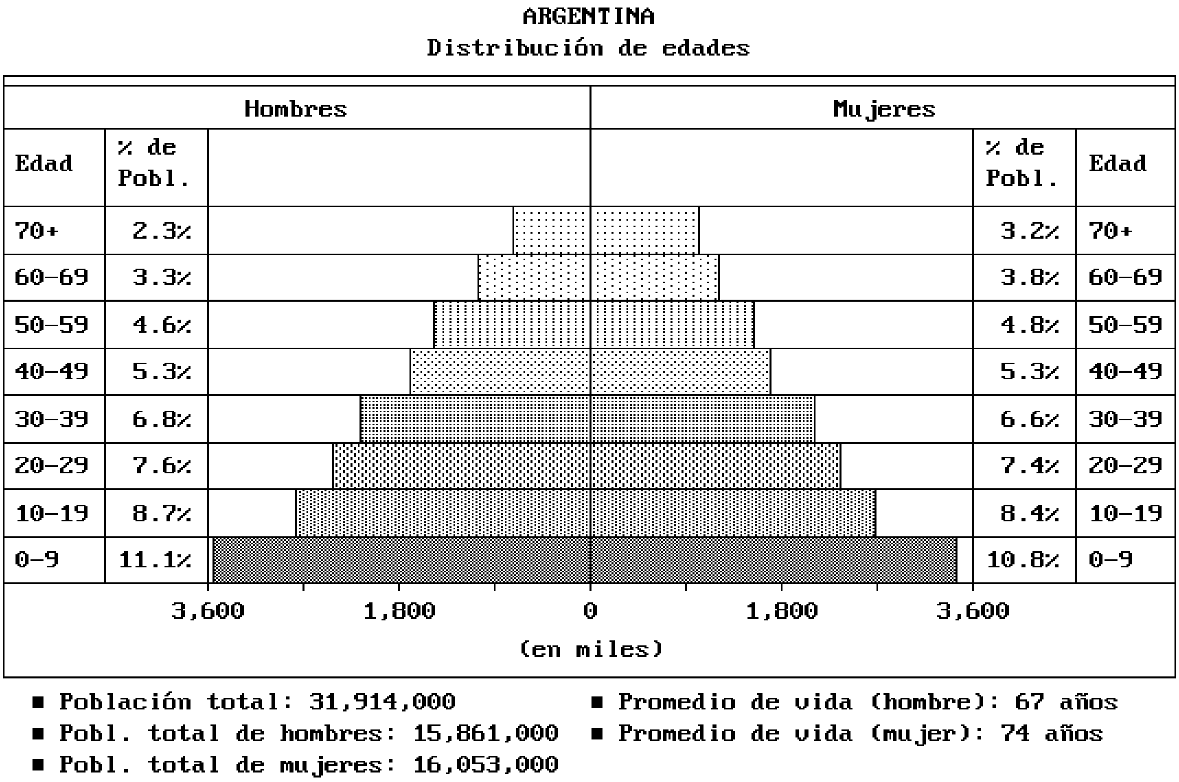
Para el ejemplo es:

X_i	f_i
56-59	3
59-62	4
62-65	13
65-68	18
68-71	27
71-74	20
74-77	10
77-80	8
80-83	5
83-86	2
Σ	110



• Pirámides de población

Un caso particular de histograma son las pirámides poblacionales, en las cuales se representan mediante barras horizontales y separadas por sexo las magnitudes correspondientes a los distintos grupos de edades. En el siguiente ejemplo se muestra una pirámide de población correspondiente a nuestro país, en la que se ha adoptado una amplitud de clase de 10 años, pero el tamaño de las mismas puede ser anual, quincenal, etc. según convenga en cada caso.

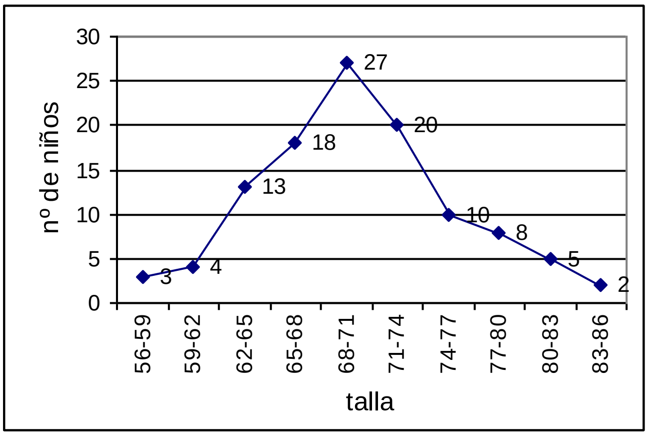


Fuente: Programa PCGlobe-1992

- <u>Polígono de frecuencias</u>

Si en un histograma se proyectan las marcas de clase sobre las bases superiores de los rectángulos, y a los puntos obtenidos se los une mediante tramos rectos, se obtiene una poligonal cuyos vértices representan las frecuencias absolutas de cada una de las clases. Esta poligonal recibe el nombre de polígono de frecuencias y, en numerosas ocasiones permite visualizar con mayor claridad el comportamiento de una variable.

Para nuestro ejemplo sobre la talla de los recién nacidos será:



- Diagrama de sectores o diagramas de torta

En los diagramas de sectores cada suceso está representado por un sector circular de una amplitud proporcional a su frecuencia.

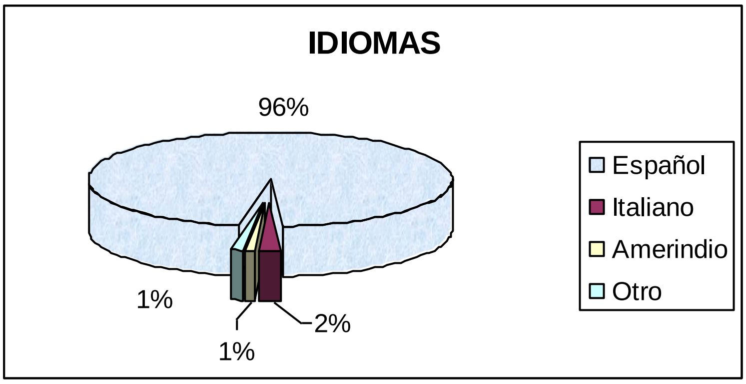
La amplitud de cada sector circular se obtiene mediante una simple regla de tres.

Ejemplos:

Para la siguiente tabla:

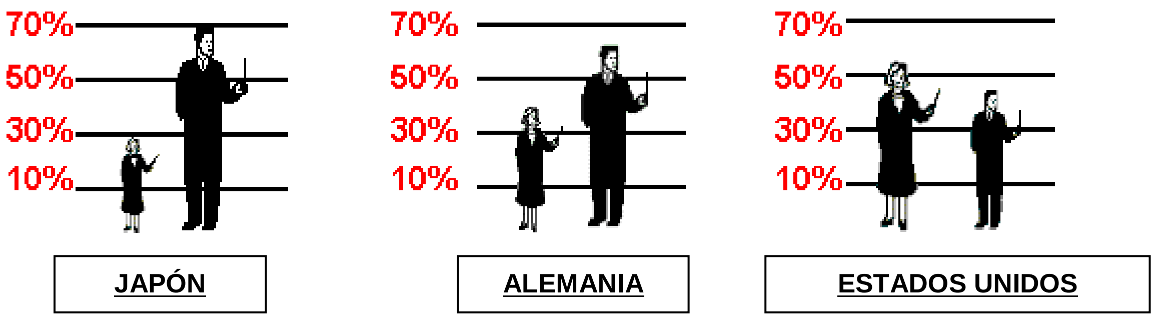
ARGENTINA					
Idiomas	%	Grupos étnicos	%	Religiones	%
Español	96	Europeo	85	Católica	93
Italiano	2	Amerindio/Mestizo	15	Protestante	2
Amerindio	1			Judía	1
Otro	1			Otra	4

Fuente: Programa PCGlobe-1992



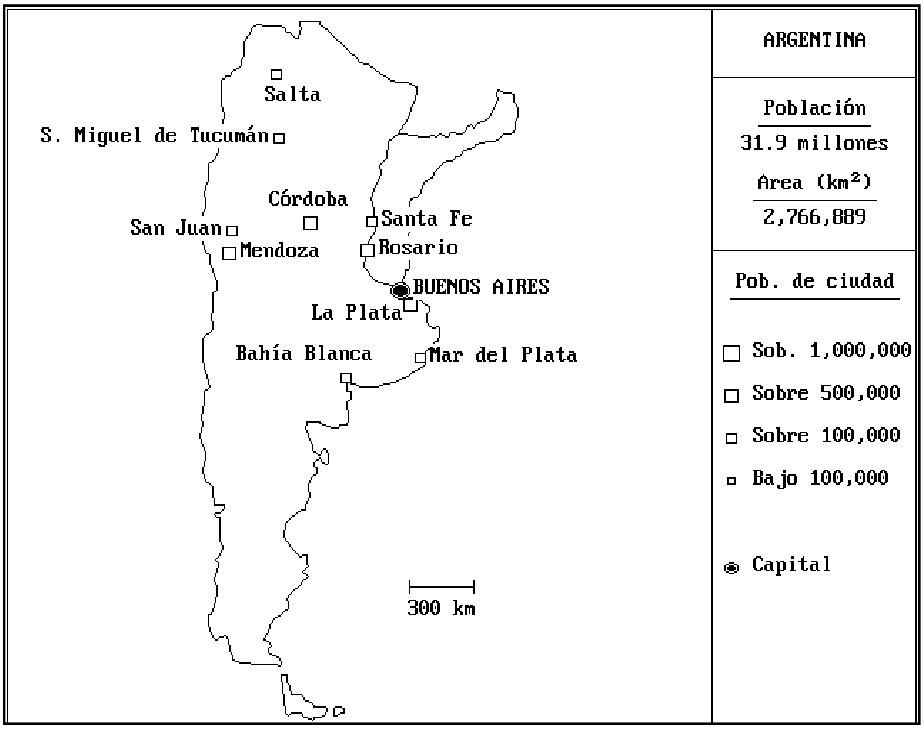
- Pictogramas

Un pictograma es una representación en la que junto al gráfico de los datos aparece una imagen relativa a lo que se mide.



- Cartogramas

Se llama cartograma a la representación de datos sobre un mapa.



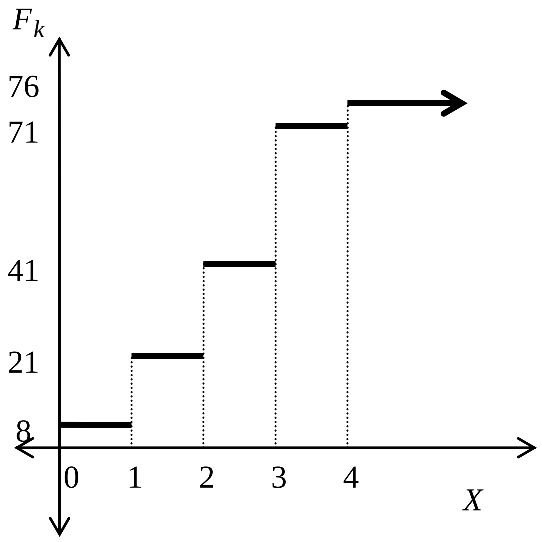
Fuente:Programa PCGlobe-1992

- Diagrama de frecuencias acumuladas

Esta representación gráfica se utilizar para mostrar frecuencias acumuladas de las variable aleatoria discretas. El modelo es una función "escalonada", es decir con tramos constantes entre dos valores consecutivos de la variable, cuyos valores se corresponden con las frecuencias absolutas acumuladas o relativas acumuladas hasta el valor considerado.

Para el ejemplo que muestra el número de hijos en edad escolar de los empleados de una fábrica es:

X	f_i	F_k	F_{kr}	$F_{kr}\%$
0	8	8	0,105	11
1	13	21	0,276	28
2	20	41	0,539	54
3	30	71	0,934	93
4	5	76	1	100
	$\Sigma = 76$			

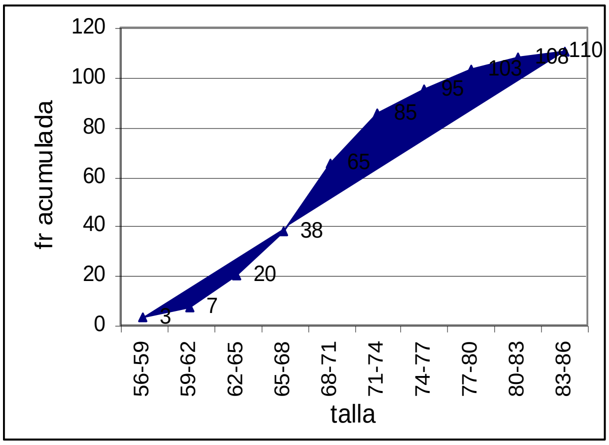


- <u>Ojiva o polígono de frecuencias acumuladas</u>

A veces resulta necesario conocer cuantos datos o que proporción de ellos son menores que un determinado valor. Para lo que es necesario mostrar las frecuencias acumuladas de una serie de datos.

Se llama ojiva al diagrama que muestra las frecuencias acumuladas tomadas sobre los límites reales de cada clase, o sobre la marca de clase. En el eje horizontal se anotan los valores de los datos y en el eje vertical las frecuencias acumuladas, las frecuencias relativas acumuladas o las porcentuales acumuladas.

En el ejemplo de la talla de los niños es:



5 OBTENCIÓN DE PARÁMETROS CARACTERÍSTICOS

* En 1890 la esperanza de vida de los españoles era de 34,7 años. En 1985 esta esperanza de vida era de 75,8 años. Estos datos son suficientes para constatar no sólo que los españoles de hoy viven más del doble que los de principios de siglo, sino que las condiciones económicas han mejorado, que el sistema sanitario se ha perfeccionado y que la alimentación es más equilibrada.
(Fuente: "Matemáticas 2" – Martínez Mediano; Cuadra López y Jiménez Villanueva)

¿Qué información brinda el ejemplo anterior? ¿Qué conclusión se puede extraer de los datos dados?

Se llaman parámetros característicos a todos aquellos valores que describen de manera precisa a un conjunto de datos. Existen distintos tipos de parámetros:

1- Medidas de centralización:

En la mayoría de los casos los datos de una serie de frecuencias tienden a agruparse alrededor de un punto central, como por ejemplo su promedio. Estos valores centrales son útiles pues representan a todos los valores de dicha serie y por esto se los conoce también como medidas de posición.

- Media aritmética: La media aritmética o simplemente media es la medida de centralización que se usa con mayor frecuencia y no es otra cosa que el promedio de una serie de datos.

La MA se calcula sumando el valor de todos los datos y dividiendo el resultado por el número total de ellos. Si la variable toma los valores $X_1 ; X_2 ; X_3 ; ... ; X_i$

En símbolos: $$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Ejemplo:

Si los datos son: 25 12 23 28 17 15 $$\overline{X} = \frac{25+12+23+28+17+15}{6} = 20$$

Si los datos están agrupados en una serie de frecuencias se calcula su MA de la siguiente manera:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i . f_i}{n}$$

X_i el valor ***i*** de la variable aleatoria
f_i la frecuencia de dicho valor
n el número total de casos

Para el ejemplo correspondiente al carácter estadístico cuantitativo discreto: número de hijos, en edad escolar, por obrero en cierta institución.

Número de hijos: X	Número de familias f
0	8
1	13
2	20
3	30
4	5
	Σ = 76

$$\overline{X} = \frac{0.8+1.13+2.20+3.30+4.5}{76} = 2{,}14$$

Si los datos se presentan en intervalos se debe tomar como $\boldsymbol{X_i}$ a la marca de cada clase, siendo este valor el representativo de toda la clase.

$$\overline{X} = \frac{\sum_{i=1}^{n} M_i \cdot f_i}{n}$$

$\boldsymbol{M_i}$ el valor ***i*** de la marca de clase
$\boldsymbol{f_i}$ la frecuencia de dicho valor
n el número total de casos

Para el ejemplo que muestra la talla en centímetros de 100 niños:

X_i	M_i	f_i
56-59	57,5	3
59-62	60,5	4
62-65	63,5	13
65-68	66,5	18
68-71	69,5	27
71-74	72,5	20
74-77	75,5	10
77-80	78,5	8
80-83	81,5	5
83-86	84,5	2
Σ		110

$$\overline{X} = \frac{57{,}5.3+60{,}5.4+63{,}5.13+66{,}5.18+69{,}5.27+72{,}5.20+75{,}5.10+78{,}5.8+81{,}5.5+84{,}5.2}{110} = 70{,}21$$

Si se trabaja con población, en vez de una muestra, la media poblacional se denota con la letra griega μ .

- <u>La mediana:</u> la mediana es el valor que se encuentra en el centro o punto medio de una secuencia ordenada de datos; es decir que divide al conjunto en dos partes iguales. Por esto, se denomina mediana al valor central de los datos cuando éstos se han dispuestos ordenadamente de menor a mayor, sin importar qué valores toma dicha variable. (La mediana deja al 50% de los valores por debajo de la misma y el 50% restante por encima de ella.)

Si el número de datos es impar su cálculo es directo, pero si la cantidad de datos es par su valor se obtiene haciendo la semisuma de los dos centrales.

<u>Ejemplo:</u>

Para los datos dados, previo ordenamiento de los mismos resulta:

12 15 17 23 25 28 $\quad$ $\text{Me} = \dfrac{17+23}{2} = 20$

Si en cambio tenemos:

12 15 17 23 25 $\quad$ Me = 17

Si los datos se presentan en intervalos de clase la fórmula para su cálculo es:

$$Me = LRI + \frac{c.\left(\frac{n}{2} - \sum f \quad ant\right)}{f_{Me}}$$

siendo:

LRI: límite real inferior de la clase que contiene a la mediana
c: tamaño de clase
$\sum f\ ant$: sumatoria de las frecuencias anteriores a la clase que contiene a la Me
$\boldsymbol{f_{Me}}$: frecuencia de la clase mediana

Busque en la bibliografía sugerida por la cátedra la demostración de la fórmula anterior.

- <u>La moda</u>: la moda es el valor de un conjunto de datos que se repite con mayor frecuencia. Esta medida es la más adecuada si se trabaja con datos cualitativos.

Para el ejemplo no existe un valor que se presente con mayor frecuencia, por lo tanto no presenta moda. En los casos en que dos valores se repiten mayoritariamente se dice que la distribución es bimodal y si existen más de dos valores se llama multimodal.

Si los datos se presentan en intervalos de clase la fórmula para su cálculo es:

$$M_o = LRI + \frac{c\,.\,\Delta_1}{\Delta_1 + \Delta_2}$$

LRI: límite real inferior de la clase modal
c: tamaño de clase
$\boldsymbol{\Delta_1}$ = diferencia entre la frecuencia absoluta de la clase modal y la frecuencia de la clase inmediata anterior
$\boldsymbol{\Delta_2}$ = diferencia entre la frecuencia absoluta de la clase modal y la frecuencia de la clase inmediata posterior

Busque en la bibliografía sugerida por la cátedra la demostración de la fórmula anterior.

2 - <u>Medidas de dispersión:</u>

Las medidas de dispersión completan el análisis de una serie de datos pues determinan la mayor o menor separación de los datos con respecto a su valor central. Es decir que indican el alejamiento de los valores de la variable con respecto a sus medidas de centralización; siendo por lo tanto una forma de evaluar la heterogeneidad de los datos.

<u>Medidas de dispersión absolutas:</u>

- <u>El rango</u>: es la diferencia entre el mayor y el menor de los valores que toma nuestra variable. Si bien esta medida brinda una primera idea de la heterogeneidad de los valores de la VA, presenta el inconveniente de que toma solo los valores extremos y nada dice sobre los intermedios.

$$R = X_{max} - X_{min}$$

- <u>Los desvíos</u>: el desvío de cada valor de la VA con respecto a la media aritmética es igual a la diferencia entre dicho valor y la media de un conjunto de datos.

$$d_i = X_i - \overline{X}$$

- <u>La desviación media</u>: <u>es</u> el promedio de los valores absolutos de los desvíos antes definidos.

$$DM = \frac{\sum_{i=1}^{n} |\,d_i\,|}{n}$$

Para datos agrupados en series de frecuencias $DM = \dfrac{\sum_{i=1}^{n} |\,d_i\,|\ .\ f_i}{n}$

- <u>La varianza</u>: se define como el promedio de los cuadrados de los desvíos con respecto a la media aritmética.

Podemos simbolizarla como: $Var(X) = \frac{\sum_{i=1}^{n} (d_i)^2}{n} = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n}$

Si los datos se repiten con una determinada frecuencia absoluta f_i cada uno de ellos resulta: $Var(X) = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2 . f_i}{n}$

Si se trabaja con una población la varianza se denota con σ_X^2, pero si se analizan datos muestrales su símbolo es s^2.

- <u>La desviación típica o estándar:</u> esta es la medida de uso más frecuente y se calcula mediante:

  $\sigma = \sqrt{Var(X)}$ para población $s = \sqrt{Var(x)}$ para muestras

<u>Medidas de dispersión relativas:</u>

- <u>El Coeficiente de Variación:</u> en muchos casos es de utilidad efectuar comparaciones entre medidas. La medida de variación relativa utilizada más comúnmente es el llamado coeficiente de variación. Es la desviación típica expresada en porcentaje de la media aritmética.

  En símbolos: $C.V. = \frac{\sigma}{\overline{X}} .100$

Su utilidad consiste en que nos permite comparar la dispersión o variabilidad de dos o más grupos. Así, por ejemplo, si tenemos el peso de 5 pacientes (70, 60, 56, 83 y 79 kilogramos) cuya media es de 69,6 kilogramos, y su desviación típica (s) = 10,44 y la tensión arterial de los mismos (150, 170, 135, 180 y 195 mmHg) cuya media es de 166 mmHg y su desviación típica de 21,3. La pregunta sería: ¿qué distribución es más dispersa, el peso o la tensión arterial? Si comparamos las desviaciones típicas observamos que la desviación típica de la tensión arterial es mucho mayor; sin embargo, no podemos comparar dos variables que tienen escalas de medidas diferentes, por lo que calculamos los coeficientes de variación:

$CV = \frac{10,44}{69,6} = 0,15 = 15\%$ para la variable peso

$CV = \frac{21,30}{166} = 0,128 = 12,8\%$ para la tensión arterial

Observando los resultados se puede afirmar que la variable peso tiene mayor dispersión.

3 - Medidas de concentración:

- Los cuartiles: son los valores de la variable aleatoria que dividen al conjunto en cuatro grupos iguales.

  Si los datos están agrupados en intervalos de clase, se obtienen mediante:

$$Q_i = LRI + \frac{\left(\frac{i.n}{4} - \sum f \quad ant\right).c}{f_{Q_i}}$$

donde:

LRI: límite real inferior de la clase que contiene al cuartil
$f_{Q\ i}$: frecuencia de la clase del cuartil ***i***.
c: tamaño de clase

- Los deciles: son los valores de la variable aleatoria que dividen al conjunto en diez grupos iguales.

  Si los datos están agrupados se obtienen mediante:

$$D_i = LRI + \frac{\left(\frac{i.n}{10} - \sum f \quad ant\right).c}{f_{D\ i}}$$

donde:

LRI: límite real inferior de la clase que contiene al decil
$f_{D\ i}$: frecuencia de la clase del decil ***i***.
c: tamaño de clase

- Los percentiles: son los valores de la variable aleatoria que dividen al conjunto en cien grupos iguales.

  Si los datos están agrupados se obtienen mediante:

$$P_i = LRI + \frac{\left(\frac{i.n}{100} - \sum f \quad ant\right).c}{f_{P\ i}}$$

donde:

LRI: límite real inferior de la clase que contiene al percentil
$f_{P\ i}$: frecuencia de la clase del percentil ***i***.
c: tamaño de clase

Forma de una distribución:

La desviación típica y las demás medidas de dispersión miden la disgregación de una distribución de frecuencias; sin embargo existen otras características de la forma de la misma que quedan sin determinar.

- El sesgo: mide el grado de concentración de los datos de una distribución a un lado y otro de la media; expresando, por lo tanto la asimetría de la misma.

Si una distribución es unimodal y simétrica la media aritmética, la mediana y la moda coinciden, $\overline{X} = Me = Mo$. Pero si resulta asimétrica, la relación entre estas medidas cambia.

Diremos que una distribución es *asimétrica a la derecha* si las frecuencias (absolutas o relativas) descienden más lentamente por la derecha que por la izquierda.

Si las frecuencias descienden más lentamente por la izquierda que por la derecha diremos que la distribución es *asimétrica a la izquierda.*

En una distribución con **sesgo a la derecha**, la media aritmética es la mayor de las tres, pues en ella influyen los valores muy altos de la variable y queda la mediana entre la moda y la media. $Mo < Me < \overline{X}$

En una distribución **sesgada a izquierda**, la media es el menor valor, ubicándose la mediana entre ella y la moda, que resulta ser el mayor valor. $\overline{X} < Me < Mo$

Existen varias maneras de determinar la forma de una distribución:

- Coeficiente de asimetría de Pearson:

Se define como:

$$A_s = \frac{\overline{X} - M_o}{s}$$

$\overline{X}$: media aritmética
M_o : moda
s: desviación estándar

Siendo cero cuando la distribución es simétrica, positivo cuando existe asimetría a la derecha y negativo cuando existe asimetría a la izquierda.

- Coeficiente de Asimetría de Fisher

Se basa en las diferencias $X_i - \overline{X}$. Como analizamos al estudiar los desvíos, su media es siempre nula; si las elevamos al cuadrado, serían siempre positivas, por lo que tampoco servirían, por lo tanto se elevan esas diferencias al cubo. Para evitar el problema de la unidad, y hacer que sea una medida adimensional y por lo tanto relativa, dividimos por el cubo de su desviación típica. Con lo que resulta la siguiente expresión:

$$A_s = \frac{\sum_{i=1}^{n} \frac{\left(X_i - \overline{X}\right)^3 . f_i}{n}}{s^3}$$

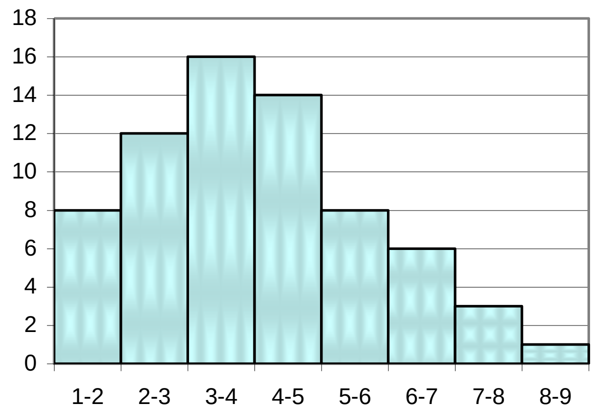
X_i : valores de la variable
$\overline{X}$: media aritmética
s: desviación estándar
f_i : frecuencia absoluta de cada valor de la variable

n : cantidad total de datos

La interpretación del coeficiente de Fisher es la misma que la del coeficiente de Pearson, si la distribución es simétrica vale cero, siendo positivo o negativo cuando exista *asimetría a la* derecha *o izquierda respectivamente.*

Ejemplo:

Se muestra el histograma correspondiente a una distribución con $\overline{X} = 4{,}044$, $M_o = 3{,}667$ y $s = 1{,}702$; por lo que $A_s = 0{,}22$.



Se observa que la distribución presenta un sesgo a derecha y que el coeficiente de Pearson es positivo.

ACTIVIDAD 1

1) La fabricación en un taller determina cierto porcentaje de piezas defectuosas, las que se rechazan. Se ha observado 20 lotes diferentes de 100 piezas c/u obteniéndose los siguientes resultados:

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

Siendo X el n° de piezas rechazadas por lote:

1.1* Elaborar la distribución de frecuencias para la variable aleatorias discreta.
1.2* Representar mediante gráficos las frecuencias absolutas y las acumuladas.
1.3* Hallar media, mediana, moda, varianza y desviación estándar.

2) El siguiente conjunto de datos representa pesos, medidos en Kg., de paquetes que contienen piezas de máquinas. Los datos se presentan ordenados en forma creciente.

92.3	94.0	94.4	95.7	96.1	96.4	97.2	97.5	97.9	98.3
98.4	98.5	98.6	98.9	99.3	99.6	100.0	100.0	100.1	100.3
100.5	100.7	100.8	101.1	101.2	101.5	102.1	102.5	102.9	103.8

2.1* Elabore una distribución de frecuencias para los datos anteriores usando 92.0 como límite de clase inferior de la primera clase y 2.0 como tamaño de clase.
2.2* Calcule la mediana y la moda.

3) En la siguiente tabla, se han borrado algunos de los resultados obtenidos al lanzar un dado 30 veces.

Número	1	2	3	4	5	6
Frecuencia	7	5	6	7		

3.1* ¿Cuáles son las frecuencias faltantes si la media aritmética es 3?
3.2* Determinar la mediana, moda y la desviación estándar.

4) La tabla siguiente muestra una distribución de frecuencia de diámetros (en cm) de los árboles tomados a la altura del pecho en un monte forestal:

Diámetro (en cm)	N° de árboles
50 a 59	8
60 a 69	10
70 a 79	16
80 a 89	14
90 a 99	10
100 a 109	5
110 a 119	2

Calcular:
4.1* límite inferior de la 6° clase

4.2* límite superior de la 4° clase

4.3* marca de clase de la 3° clase

4.4* límites reales de la 5° clase

4.5* tamaño del 5° intervalo de clase

4.6* frecuencia de la 3° clase

4.7* frecuencia relativa de la 3° clase

4.8* intervalo de clase de mayor frecuencia

5) Construir un histograma y un polígono de frecuencia con los datos del ejercicio anterior.

6) Construir una distribución de frecuencia acumulada y una ojiva.

7) Obtener $\overline{X}$, M_e , M_o , Q_1 , Q_3 , D_8 , P_{14} , P_{74} , $Var(x)$ y σ_X .

8) Determinar la asimetría de la distribución.

9) De las bases de datos del Seguro Social se extrajo la siguiente información: de un total de 1965 trabajadores que pidieron su jubilación entre los 60 y 89 años de edad, se habían clasificado en grupos quinquenales, y se disponía de las siguientes cifras:

Rango de edades	Nº trabajadores jubilados
60-64	373
65-69	733
70-74	489
75-79	240
80-84	102
85-89	28
Total	

Calcular el cuartil Q_1 y el decil D_6 , e interpretar los resultados.

10) Hallar los coeficientes de Pearson y de Fisher extrayendo conclusiones sobre la forma de la distribución.

11) Del informe de una empresa transportista sobre la frecuencia de carga mensual de combustible de 50 camiones, se obtuvo la siguiente distribución de frecuencias relativas:

X	2	4	7	5	3	6	8
f_r	0,04	0,18	0,16	0,24	0,06	0,20	0,12

11.1* Calcular el porcentaje de transportistas que carga combustible por mes:
a) 5 veces b) menos de 5 veces c) no menos de 6 veces d) a lo sumo 6 veces.

11.2* Calcular cuántos transportistas cargan combustible por mes:
a) 3 veces b) menos de 3 veces c) no menos de 7 veces d) a lo sumo 6 veces
e) entre 3 y 7 veces inclusive.

12) Los siguientes datos recopilados en campaña muestran las alturas, en metros, de 40 tallos de plantas de maíz:

1,06	1,16	1,21	0,96	1,17	1,11	1,03	1,11	1,20	1,26
1,14	1,15	1,07	1,18	1,22	0,97	1,20	1,11	1,14	1,09
1,18	1,12	1,15	1,05	1,24	1,12	1,19	1,03	1,19	1,10
1,33	1,04	1,18	1,12	1,19	1,08	1,27	1,30	1,13	1,13

12.1* Realizar la correspondiente distribución de frecuencias.
12.2* Graficar el histograma y su correspondiente polígono de frecuencias.
12.3* Graficar la ojiva.
12.4* Obtener, la media aritmética, la mediana, la moda, el cuartil tres, el decil ocho y el percentil veinticuatro.
12.5* Obtener la varianza y la desviación estándar.
12.6* Hallar el coeficiente de sesgo mediante método de Pearson y extraer conclusión.

13) En un campo de la ciudad de Santa Fe se realizó un experimento sobre 30 tortas de girasol. Se midió el diámetro de dichas tortas y los datos obtenidos fueron los siguientes:

10,5	20,2	15,4	11,6	13,8	20,5	14,0	16,5	17,0	12,0
15,4	21,3	22,0	23,1	15,9	16,5	13,0	18,0	16,0	17,3
17,8	19,0	13,4	17,3	14,2	15,0	16,2	19,0	15,5	16,0

13.1* Resuma esta información en una distribución de frecuencias.
13.2* ¿Cómo representaría esta información gráficamente? Hágalo de todas las maneras posibles.
13.3* Determine las medidas de centralización: media aritmética, mediana y moda.
13.4* Determine las medidas de dispersión absolutas: rango, desvíos, desviación media, varianza y desviación estándar.
13.5* Halle el coeficiente de variación.
13.6* Obtenga los cuartiles 1º y 3º, los deciles 1º y 8º y los percentiles 25º y 78º.
13.7* Encuentre los coeficientes de sesgo e indique la forma de la distribución.

14) A continuación se dan dos tablas que indican la distribución de la población obrera, según el nivel de sus salarios, en dos ciudades del interior del país:

Ciudad A

Salarios mensuales (en cientos de pesos)	Nº de obreros
2-4	3
4-6	15
6-8	18
8-10	20
10-12	16
12-14	14
14-16	4
	90

Ciudad B

Salarios mensuales (en cientos de pesos)	Nº de obreros
2-4	7
4-6	8
6-8	10
8-10	15
10-12	9
12-14	8
14-16	6
16-18	2
	65

14.1* Obtenga la media, mediana y moda, en cada caso y analice en forma crítica los resultados.
14.2* Encuentre sus desvíos y las desviaciones medias.
14.3* Halle las varianzas y las desviaciones estándares.
14.4* Encuentre los coeficientes de variación y extraiga conclusiones sobre la dispersión de los grupos.
14.5* Analice la forma de las distribuciones.

15) La siguiente tabla muestra la distribución de puntajes de un examen de ingreso laboral:

Puntaje	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
Nºalumnos	7	16	21	34	50	37	20	15

Calcular el puntaje mínimo necesario para ingresar al trabajo sabiendo que solo es admitido el 30 % de los postulantes.

16) Completar la siguiente tabla de frecuencias y calcular mediana y coeficiente de sesgo.

Variable X	f absoluta	F acumulada	f relativa	Desvíos
12 - 15	5			
		12		
18 – 21		21		
21 - 24	9		22,5%	

17) Seleccionar la opción correcta y justificar.

Pablo es uno de los 120 alumnos que cursan II año en una Facultad de Ciencias. Al leer el listado de notas correspondientes a la evaluación de Probabilidad y Estadística, comprueba que solo 14 alumnos tienen mejor nota que él y comenta:

17.1* He superado el noveno decil.
17.2* Mi nota es el percentil 91
17.3* He superado el percentil 86
17.4* No llegué al tercer cuartil.

18) Un cereal para desayuno incluye 200 pasas de uva en cada caja. Una muestra de 30 cajas de la producción del día de ayer mostró el siguiente número de pasas:

200	200	202	204	207	197
199	201	206	193	197	194
202	201	196	197	202	200
199	196	198	201	200	202
198	203	205	202	200	198

18.1 Elabore una distribución de frecuencias completa.
18.2 Obtenga la media y la moda.
18.3 Halle la desviación estándar.

19) Los ingresos mensuales en pesos de un grupo de empleados públicos son:

Ingresos	fi	fr	Fk	Fkr
300-350	10			
350-400		0,10		
400-450	15			
450-500				0,74
500-550			50	

19.1* Completar la tabla de distribución de frecuencias.
19.2* Calcular la mediana y la desviación típica.

20) Los datos que se muestran a continuación corresponden a los gastos de electricidad en un mes determinado, en pesos, de una muestra aleatoria de 30 familias de una localidad. Elaborar una distribución de frecuencias con intervalos de clase donde el límite inferior de la primera clase sea 55 y el tamaño de clase igual a 5.

96	59	60	78	80	85
72	87	91	73	83	68
89	70	90	96	78	73
69	67	87	91	75	72
79	77	93	88	85	79

Hallar la media aritmética, el desvío estándar, rango, mediana, moda y percentil 90.

21) Las notas obtenidas por 50 alumnos de un curso en un examen son las siguientes:

6	6	2	8	9	3	6	1	7	9
10	4	4	6	3	7	3	2	6	4
7	6	6	10	3	3	9	2	1	3
7	7	4	10	10	8	4	1	4	6
1	4	9	1	2	1	2	5	4	9

21.1* Si los alumnos que obtuvieron menos de 6 puntos deben rendir un examen en marzo, ¿cuántos alumnos tienen que rendir dicho examen?
21.2* Calcula la media aritmética.
21.3* Si se premió con un campamento de dos días al 50% de los alumnos que habían obtenido mayor nota, ¿cuáles son las notas de los alumnos premiados?

22) Se hace un censo para saber cuántos ambientes tienen las viviendas de 20 personas que trabajan en una oficina. Se obtienen los siguientes datos:

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

22.1* Calcula la desviación media.
22.2* ¿Cómo justifica teóricamente la afirmación: “los empleados de la oficina, en su mayoría, viven en viviendas de 3 ambientes”.
22.3* Muestre los datos en un diagrama apropiado.

23) Damos a continuación los promedios de Matemática de una división de 5º año:

4.33	5	7	9.66	10	8	6
2	1.33	3	4	4.5	7.5	6.66
5.5	9	4	1	10	4.5	7.33
7	8	6	6	2	5	5.5

23.1* Construye la distribución de frecuencia, considerando 5 intervalos de clase, comenzando en cero y con una amplitud de 2.
23.2* Si la nota de aprobación de la asignatura es 6, ¿considera usted que el 50 % de los alumnos aprobó?. Justifique
23.3* ¿A partir de qué promedio se encuentra el 25% de lo mejores alumnos?

24) Se midieron los niveles de colinesterasa en un recuento de eritrocitos en μmol/min/ml de 34 agricultores expuestos a insecticidas agrícolas, obteniéndose los siguientes datos:

Intervalo	7,5-9	9-10,5	10,5-12	12-13,5	13,5-15	15-16,5
Frecuencia	3	8	10	10	1	2

24.1* Calcule la media aritmética, la mediana y la moda.
24.2* Extraiga, mediante la comparación de sus valores, una conclusión sobre la forma de la distribución.
24.3* Halle el coeficiente de sesgo y compare con el resultado del punto anterior.

Respuestas a ejercicios

1)

1.3* $\bar{X} = 3{,}15$ $Me = 2{,}5$ $Mo = 1\ ;\ 2$ $Var(X) = 4{,}93$ $s = 2{,}22$

2)

2.2* $Me = 99{,}71\,kg$ $Mo = 100{,}67\,kg$

3)

3.1* $f(5) = 3$ $f(6) = 2$

3.2* $Me = 3$ $M0 = 1\ ;\ 4$ $Var(x) = 1{,}528$ $\sigma_X = 0.633$

4) 4.1* 100 cm 4.2* 89 cm 4.3* 74,5 cm 4.4* 89,5 - 99,5 cm

4.5* 10 cm 4.6* 16 4.7* 0,25 4.8* 3º

7) $\bar{X} = 79{,}27cm$ $M_e = 78{,}56cm$ $M_o = 77cm$ $Q_1 = 67{,}75cm$

$Q_3 = 90{,}25cm$ $D_8 = 93{,}5cm$ $P_{14} = 60{,}6cm$ $P_{74} = 89{,}6cm$

$Var(x) = 243{,}4cm^2$ $s = 15{,}6cm$

8) $As = 0{,}14$

9) $Q_1 = 65{,}31$ *años* $D_6 = 70{,}25$ *años*

10) $A_s = 0{,}36$ (Pearson) $A_s = 0{,}706$ (Fisher)

11)

11.1* a) 24% b) 28% c) 48% d) 72%

11.2* a) 3 b) 2 c) 14 d) 36 e) 42

12)

12.4* $\bar{X} = 1.1375\,m$ $Me = 1.135\,m$ $Mo = 1.1375\,m$
$Q_3 = 1.185\,m$ $D_8 = 1.202\,m$ $P_{24} = 1.078\,m$

12.5* $Var(x) = 0.006\ m^2$ $\sigma_X = 0.0812\,m$ 12.6* $A_S = 0$

13)

13.3* $\overline{X} = 16,6cm$ $Me = 16,1cm$ $Mo = 15,4cm$

13.4* $R = 12,6cm$ $DM = 2,53cm$ $Var(x) = 9,55cm$ $s = 3,09cm$ 13.5* $CV = 18,6\%$

13.6* $Q_1 = 14,4\text{cm}$ $Q_3 = 17,95cm$ $D_1 = 12,8cm$ $D_8 = 19,1cm$ $P_{25} = 14,4cm$ $P_{78} = 18,62cm$

13.7* $As = 0,34$ (Pearson) a la derecha $As = 0,267$ (Fisher) a la derecha

14)

14.1* Ciudad A: $\overline{X} = 8,978$ $Me = 8,9$ $Mo = 8,67$

Ciudad B: $\overline{X} = 9,12$ $Me = 9$ $Mo = 8,9$

14.2* Ciudad A: $DM = 2,5$ Ciudad B: $DM = 3,04$

14.3* Ciudad A: $Var(x) = 9,46$ $s = 3,07$ Ciudad B: $Var(x) = 14,25$ $s = 3,77$

14.4 Ciudad A: $CV = 34,18$ Ciudad B: $CV = 41,33$

14.5* Ciudad A: $As = 0,09$ Ciudad B: $As = 0,058$

15) $D_7 = 73,24$

16) Me= 20,66 $As = -1,63$ sesgo a izquierda

17) Opción correcta 17.3*

18)

18.2* $\overline{X} = 200\,pasas$ $Mo_1 = 200$ y $Mo_2 = 202$ 18.3* *s=3,25pasas*

19)

19.2* $Me = 433,33$ $s = 71,67$

20) $\overline{X} = 80,16$ $s = 10,306$ $R = 37$ $Me = 79,16$ $Mo_1 = 76$ $Mo_2 = 88,33$ P_{90}=93,75

21)

21.1* 26 21.2* $\overline{X} = 5,1$ 21.3* $x \geq 5$

22)

22.1* $DM = 1,04$

23)

23.2* $Q_2 = Me = 5,4$ no aprobó 23.3* $Q_3 = 7,14$

24)

24.1* $\overline{X} = 11,42$ $Me = 11,42$ $Mo = 12$

24.2* a cargo del alumno

24.3* $As = -0,314$