Notarás que mediante esta agrupación podemos obtener una "imagen" inmediata la distribución de los puntajes del C.I. entre nuestros estudiantes de bachillerato. P ejemplo, se observa que hay concentración de frecuencias en los intervalos de cl entre los puntajes de 100 y 119. Se observa también que el número de puntajes los extremos tiende a disminuir. Así, se cumple uno de nuestros objetivos agrupación, que es lograr una colección de puntajes económica y manejable.

Haciendo un análisis sobre los "límites verdaderos" de un número cardin consideremos que el valor "verdadero" de un número es igual a su valor aparente m o menos una mitad de la unidad de medida. Por supuesto, ese razonamiento tamb es válido para estos valores aún después de haberse agrupado en intervalos de clase Así, aunque escribamos los límites de intervalo de clase menor como 80-84, l límites verdaderos del intervalo serán 79.5-84.5 (el límite inferior de 80 y el lím superior de 84 respectivamente.

Debe tenerse en cuenta que los límites verdaderos de un intervalo de clase no son mismo que los límites aparentes. Cuando se calcula la mediana y los percentiles pa una agrupación de datos, haremos uso de los límites verdaderos del intervalos clase.

Frecuencias acumuladas y distribuciones porcentuales acumulativas.

A menudo conviene convertir los datos de una distribución de frecuencias en u distribución de frecuencias acumulada. Además de ayudar a la interpretación de distribución de frecuencias, una distribución de frecuencias acumuladas es bue ayuda para obtener la mediana y varios percentiles de puntajes, como veremos m delante.

La distribución de frecuencias acumuladas se obtiene de una manera muy sencil Observemos los datos en la tabla 4.

En cada entrada de la distribución de frecuencia se indica el número de estudiant de bachillerato que cae dentro de los intervalos de clase. En cada entrada de distribución de frecuencia acumulada se indica el número de todos los casos frecuencias por debajo del límite superior verdadero de ese intervalo. Así, en el terc intervalo de clase partiendo de abajo hacia arriba en la tabla 4, la entrada "13" en distribución de frecuencias acumuladas indica que un total de 13 estudiantes obtu un puntaje más bajo que el límite superior verdadero de este intervalo, es decir 94, La construcción de la distribución de frecuencias acumuladas se obtiene por un simp proceso de adición sucesiva de las frecuencias anteriores al intervalo de cla correspondiente. De este modo, la frecuencia acumulada del intervalo 104,5-109, se logra por una adición sucesiva de 3+5+5+4+12+14=43. Adviértase que frecuencia acumulada correspondiente a la entrada superior es siempre igual a N. no se obtiene este resultado quiere decir que ha habido un error en la suma frecuencias y se deberá revisar el trabajo.

La distribución porcentual acumulativa que se muestra igualmente en la tabla 3.4 se obtiene dividiendo cada valor en la columna de frecuencia acumulada por N y multiplicada por 100. Adviértase que el ingreso superior debe ser 100%, puesto que todos los casos caen bajo el límite superior real del intervalo más alto.

$$\%\ acum = \frac{f\ acum.}{N} x100$$

INTERVALO DE CLASE	*f*	*f* ACUMULADA	% ACUMULADO
150-154	2	110	100
145-149	2	108	98
140-144	3	106	96
135-139	5	103	94
130-134	7	98	89
125-129	9	91	83
120-124	9	82	75
115-119	13	73	66
110-114	17	60	55
105-104	14	43	39
100-104	12	29	26
95-99	4	17	15
90-94	5	13	12
85-89	5	8	7
80-84	3	3	3

Tabla 4
Distribución de frecuencias agrupadas y distribución de frecuencias acumuladas basadas en los datos que aparecen en la tabla 3 N=110

Técnicas de representación gráfica

Hemos examinado algunos de los procedimientos empleados para hacer que adquiera sentido un conjunto de datos no organizados. Como se puntualizó generalmente el trabajo está apenas comenzando cuando tu hayas obtenido la distribución de frecuencias de los datos. El siguiente paso, por lo común, es presentar los datos en forma de dibujo de tal modo que el lector pueda percibir fácilmente los hechos esenciales de una distribución de frecuencias y compararlos con otra si lo desea. Estos dibujos reciben el nombre de gráficas, y no deben considerarse como sustitutos de un tratamiento estadístico de los datos, sino más bien como ayuda para pensar e interpretar problemas estadísticos.

Se utilizan muchos tipos de gráficas para representar datos. Hay gráficas circulares, pictóricas, de barras y de otros tipos. Al utilizar gráficas, suponemos que es posible hacer una idea rápida sobre varias características importantes de una distribuc observando con rapidez la gráfica. Para ilustrar esto, vamos a ver lo que se con como gráfica de barras, demostrando que se puede desarrollar para convertirse en gráfica más general, denominada histograma.

Primeramente, vamos a presentar datos reunidos sobre los pesos de los jugadores fútbol americano de una escuela. En el cuadro 2 la columna de la izquierda prese los pesos en kilogramos y la columna de la derecha da los números de alumnos tenían ese peso.

PESO EN KILOGRAMOS	FRECUENCIA
50	2
55	4
60	7
65	9
70	12
75	15
80	10
85	5
90	1

Cuadro 2

Observa que los pesos se dan a intervalos de 5 kilogramos y que el número total alumnos es de 65. A continuación, vamos a representar esos datos utilizando u gráfica de barras.

La observación rápida de la figura 1 revela los hechos siguientes:

1. El peso que se presenta con mayor frecuencia es el de 75 kilogramos
2. El peso que ocurre con menor frecuencia es el de 90 kg.
3. Todos los pesos dados se representan mediante números, diferentes estudiantes.
4. El peso menor es de 50 kilogramos y el peso mayor es de 90 kilogramos.

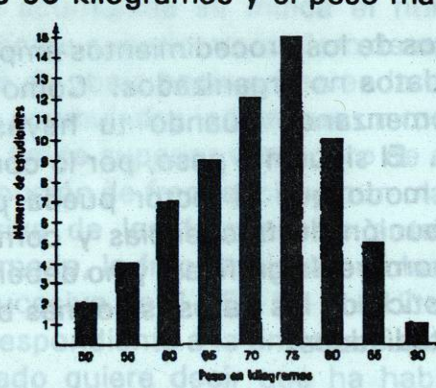

Fig. 1

A continuación, vamos a manipular la gráfica de la figura 1. Dividimos a la mitad todos los intervalos entre pesos de 5 kilogramos y conectamos los límites de esos intervalos recién creados por medio de líneas de puntos. Observa que los múltiplos de 5 kilogramos se transforman en los puntos medios de intervalos recién creados. Esto se muestra en la figura 2.

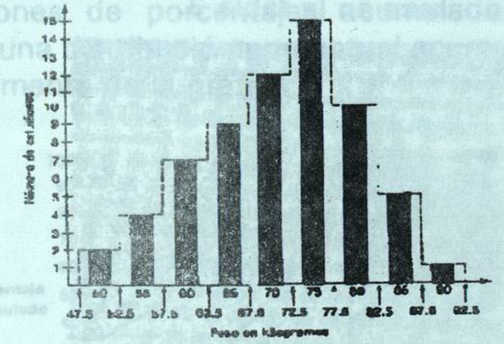

Fig. 2

La etapa siguiente es una gráfica en la que las únicas líneas son las que se muestran por medio de líneas de puntos en la figura 2. Esa gráfica se presenta en la figura 3.

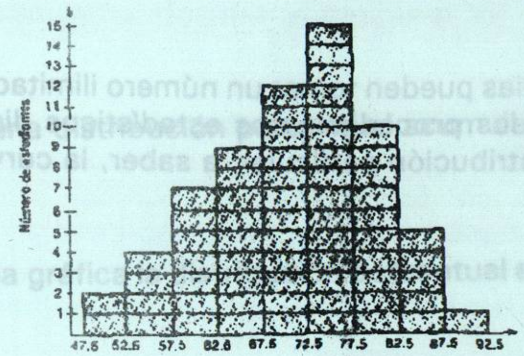

Fig. 3

En la figura 3 aparece otro cambio de la figura 2. Trazamos líneas de puntos para mostrar frecuencias. Podemos utilizar esas líneas de puntos para representarnos los números de estudiantes como áreas. La superficie de cada pequeño rectángulo representa un estudiante. Además, debemos observar que el peso es una cantidad continua. Nuestro histograma refleja correctamente este hecho. Por ejemplo, los estudiantes que caen en el intervalo entre 72.5 y 77.5 kilogramos se representan en el área de ese intervalo. Hay 15 de esos alumnos. Así pues, un histograma es un tipo muy funcional de gráfica, que se puede utilizar para representar datos continuos, tales como peso, tiempo y altura.

Podemos transformar rápidamente el histograma en otra forma empleada comúnmente para las representaciones gráficas, llamada polígono de frecuencias mediante la unión de los puntos medios de las barras a base de segmentos de rectas. Sin embargo, no es necesario elaborar un histograma antes de construir un polígono de frecuencias. Todo lo que debe hacerse es marcar un punto en el sitio correspondiente a las cimas de las barras y luego unir estos puntos. Cuando dos o más distribuciones de frecuencia se comparan entre sí, el polígono de frecuencias proporciona una visión más clara. La figura 4 muestra un polígono de frecuencias basado en la distribución de frecuencias agrupada que aparece en la tabla 3.

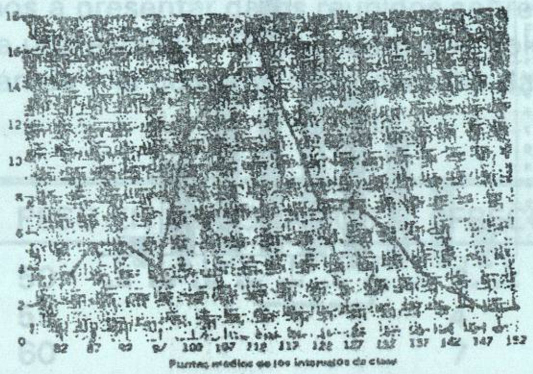

Fig. 4
En los polígonos de frecuencias, la frecuencia no puede ser representada por la altura de la ordenada, sino solamente por el área entre las ordenadas.

Los polígonos de frecuencias pueden tomar un número ilimitado de formas diferentes. No obstante, muchos de los procedimientos estadísticos discutidos en este texto, suponen una forma de distribución particular, a saber, la curva normal en "forma de campana".

Percentiles

Supongamos que un hermano o hermana menor regresa de la escuela y exclama:"Obtuve una calificación de 127 puntos en mi prueba de aptitud". ¿Cuál es tu reacción? ¿Elogiarlo por haber obtenido tan magnífica calificación? ¿Regañarlo por no haber conseguido una nota más alta? ¿O no decir nada hasta conocer mejor la distribución de las calificaciones dentro de su clase o grupo? Si usted ha comprendido el curso hasta este punto, no cabe duda que elegirá la última de estas tres alternativas.

Debe quedar claro que una calificación por sí misma carece de significado. Únicamente lo adquiere cuando se compara con alguna base o escala patrón. De esta manera, si su hermanito voluntariamente agrega:"El 79% de los estudiantes obtuvo una calificación menor que la mía", estará proporcionando un marco de referencia para interpretar su calificación. Realmente habrá citado el rango percentil de su calificación. El rango percentil de una calificación representa, por tanto, el porcentaje de los casos de un grupo que alcanzó valores menores que el citado. Así, si decimos que una calificación de 127 tiene un rango percentil de 79 significa que 79% del grupo obtuvo una calificación menor que 127. Incidentalmente, cada calificación se

considera como un punto hipotético sin dimensión, de manera que es igualmente significativo decir que el 21% del grupo en comparación obtuvo una calificación superior a 127.

En discusiones anteriores aprendimos a construir distribuciones de frecuencia acumuladas y distribuciones de porcentajes acumulados. Si tuviéramos que representar gráficamente una distribución porcentual acumulativa podríamos leer los rangos percentiles directamente de la gráfica[2]

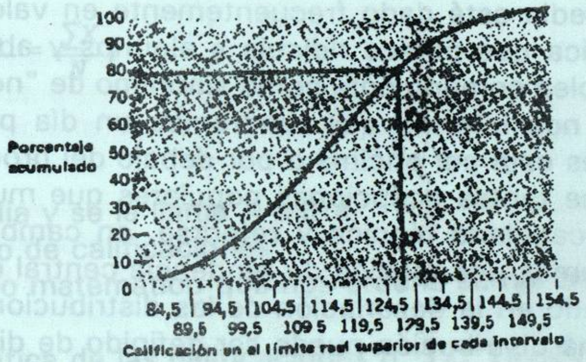

Fig. 5
Representación gráfica de una distribución porcentual acumulativa.

La figura 5 muestra en forma gráfica la distribución porcentual acumulada presentada en la tabla 4.

Como ilustración, imaginemos que queremos determinar el rango percentil de una calificación de 127. Localizamos el 127 sobre la abscisa y levantamos una perpendicular hasta el punto en el cual intercepta a la curva. Leyendo directamente en la escala de la izquierda, vemos que el rango percentil es aproximadamente 79. Por otra parte, si queremos conocer la calificación de un percentil dado, podremos invertir el procedimiento. Por ejemplo, ¿cuál es la calificación con un percentil de 90? Situamos el percentil 90 en la ordenada, leemos directamente a la derecha hasta que intercepte la curva, construimos una línea perpendicular a las abscisas, y leemos el valor en la escala de calificaciones. En el presente ejemplo puede verse que la calificación correspondiente al percentil de 90 es aproximadamente 135.

2 Observa que lo contrario también es verdad, o sea, que dado un rango percentil, podemos leer las correspondientes calificaciones.

Medidas de tendencia central

Una de las mayores fuentes de confusión entre la gente neófita, y quizás, la causa de su suspicacia al juzgar la estadística más como un arte que como una ciencia, es la ambigüedad en el uso del término "promedio". Los sindicatos y las gerencias hablan de los salarios promedio, y citan frecuentemente valores numéricos que muestran entre sí un grave desacuerdo. Los programas de televisión y los comerciales, según se dice, están preparados teniendo en mente el promedio de televidentes. Los políticos se interesan vivamente en los puntos de vista de los votantes promedio; el tamaño de la familia promedio está dado frecuentemente en valores fraccionarios, como abstracción estadística que parece ridícula a algunos y absurda a otros. El término "promedio" se emplea comúnmente como sinónimo de "normal". El servicio meteorológico de la T.V. nos informa que tendremos un día promedio o que la precipitación pluvial del mes está por encima o por debajo del promedio. En efecto, el término "promedio" tiene tantas acepciones populares que muchos Estadísticos prefieren suprimirlo del vocabulario técnico y referirse, en cambio, a medidas de tendencia central. Definiremos una medida de tendencia central como un índice de localización central empleado en la descripción de las distribuciones de frecuencia. Puesto que el centro de una distribución puede ser definido de diferentes maneras, habrá también diferentes medidas de tendencia central. Trataremos tres de las medidas de tendencia central más frecuentemente empleadas: la media, la mediana y la moda.

A lo largo de la primera parte de la sección, nos ocupamos primeramente en organizar los datos en una forma útil y significativa. Sin embargo, para ir más lejos deseamos disponer nuestros datos en forma tal que puedan presentarse proposiciones cuantitativas. Una distribución de frecuencia representa una organización de los datos, pero no nos permite por sí misma establecer proposiciones cuantitativas, ya sea describiendo la distribución o comparando dos o más distribuciones.

Hay dos características que se presentan en múltiples distribuciones de frecuencia que los Estadísticos han notado y para las cuales han desarrollado métodos cuantitativos de descripción: (1) Con frecuencia los datos se acumulan alrededor de un valor central situado entre los dos extremos de la variable que se estudia. (2) Los datos pueden tender a dispersarse y distribuirse alrededor de un valor central, en forma tal que esta tendencia puede ser especificada cuantitativamente.

La capacidad de localizar un punto de tendencia central, particularmente cuando al mismo tiempo existe una descripción de la dispersión de calificaciones con respecto a ese punto, puede ser muy útil para los Estadísticos. Por ejemplo, podrán reducir una masa de datos a un simple valor cuantitativo que llegará a ser comprendido y comunicado a otros especialistas.

Anteriormente expresamos que con frecuencia se pide al Estadístico comparar las medidas obtenidas a partir de dos o más grupos de materias con el propósito de establecer inferencias acerca de los efectos de una variable independiente. Las medidas de tendencia central simplifican mucho la tarea de establecer conclusiones.

La media aritmética

Usted probablemente estará íntimamente familiarizado con la media aritmética, porque siempre que obtiene un "promedio" de calificaciones sumando las cifras, y dividiéndolas por su número total está calculando la media aritmética. En pocas palabras, la media es la suma de las calificaciones o de los valores de una variable dividida por su número.

Expresado en forma algebraica:

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{N} = \frac{\sum X}{N},$$

Donde:

$\bar{X}$ = La media y se le llama X barra,
N = Número de calificaciones
Σ = El verbo matemático que nos ordena sumar todas las observaciones.

Así, la media aritmética de las calificaciones 8, 12, 15,19, 24 es $\bar{X} = \frac{78}{5} = 15{,}60$

X	f	fx	
12	1	12	
11	2	22	
10	5	50	$\bar{X} = \frac{\sum fX}{N}$
9	4	36	
8	6	48	$\bar{X} = \frac{232}{29}$
7	4	28	
6	3	18	
5	2	10	$\bar{X} = 8{,}00$
4	2	8	
N =	29	ΣfX	= 232

Tabla 5
Procedimientos de cómputo para calcular la media con distribuciones de frecuencia no agrupadas.

Obtención de la media de distribuciones de frecuencia no agrupadas:

Recordemos que hemos construido con anterioridad una distribución de frecuencias, como un medio para eliminar la repetición constante de calificaciones que ocurre con