frecuencia variable, con el objeto de tener una sola entrada en la columna de frecuencias para representar el número de veces que se da una calificación dada. Así, en la tabla 5 sabemos, por la columna encabezada por f, que la calificación 8 se presentó 6 veces. Al calcular la media, no es necesario sumar seis veces 8, pues podemos multiplicar la calificación por su frecuencia y obtener el mismo valor de 48. Puesto que cada calificación se multiplica por su correspondiente frecuencia antes de la adición, podemos representar la media de la distribución de frecuencias como sigue:

$$\bar{X} = \frac{\sum fX}{N}$$

Obtención de la media a partir de la distribución de frecuencia agrupada por el método de calificación original:

El cálculo de la media a partir de una distribución de frecuencia implica esencialmente los mismos procedimientos que han sido empleados para las distribuciones de frecuencia no agrupadas. Para empezar, se utiliza el punto medio de cada intervalo para representar todas las calificaciones dentro de ese intervalo. El punto medio de cada intervalo se multiplica por su frecuencia correspondiente, se suma este producto y el resultado se divide entre N. Los procedimientos empleados en el cálculo de la media a partir de una distribución de frecuencias agrupada se indican en la tabla 6.

1 Intervalo clase	2 Frecuencia f	3 Punto medio X	4 Frecuencia X punto medio fX	
125-129	2	127	254	
120-124	5	122	610	$\bar{X} = \frac{\sum fX}{N}$
115-119	8	117	936	
110-114	10	112	1120	$= \frac{10195}{100}$
105-109	15	107	1605	
100-104	20	102	2040	
95-99	15	97	1455	$=101,95$
90-94	10	92	920	
85-89	8	87	696	
80-84	4	82	328	
75-79	3	77	231	
N = 100			$\sum fX=10195$	

Tabla 6
Procedimientos de cómputo para calcular la media a partir de una distribución de frecuencia agrupada

La mediana

En distribuciones de frecuencia agrupadas, la mediana se define como la calificación o calificación potencial de una distribución abajo y arriba de la cual cae la mitad de las frecuencias. Si esta definición les parece ligeramente familiar, no es por casualidad. La mediana es sencillamente un caso especial de un rango percentil; la mediana es la calificación del percentil 50. Debe quedar claro que los procedimientos generalizados expuestos anteriormente para determinar la calificación en varios rangos percentiles, pueden ser aplicados para calcular la mediana.

Ocasionalmente, será necesario la mediana cuando N no sea suficiente para justificar la fusión de los datos en la forma de una distribución de frecuencia o de una distribución de frecuencia agrupadas. Consideremos el siguiente arreglo de calificaciones: 5, 19, 37, 39, 45. Observe que las calificaciones están ordenadas de acuerdo a su magnitud y que N es un número impar. La calificación 37 es la mediana puesto que dos calificaciones caen encima y dos calificaciones caen debajo de ella. Si N es un número par, la mediana es la suma de los dos valores centrales divido por dos. Los dos valores centrales en el arreglo de calificaciones 8, 26, 35, 43, 47, 73, son 35 y 43. La media aritmética de estos dos valores es (35 + 43)/2, o sea 39.

La moda

De todas las medidas de tendencia central, la moda es la más fácilmente se determina, puesto que se obtiene por inspección, y no por cómputo. La moda es simplemente la calificación que ocurre con mayor frecuencia. Para las datos agrupados, la moda se designa como el punto medio del intervalo que contiene la mayor frecuencia. En la tabla 6 la moda es la calificación de 102 puesto que es el punto medio del intervalo 100 - 104 al cual corresponde la mayor frecuencia.

En algunas distribuciones, que no consideramos aquí, podrá haber dos puntos máximos que produzcan la apariencia de dos jorobas, similares a las de la espalda de un camello. Tales distribuciones se denominan bimodales. Una distribución que contenga más de dos jorobas se llama multimodal.

En general, la media aritmética es la media preferida para representar la tendencia central a causa de las diversas propiedades deseables que posee. Para empezar, la media es uno de los miembros del sistema matemático, que permite su empleo en análisis estadísticos más avanzados.

Otra característica importante de la media implica que es la más estable o la más confiable de las medidas de tendencia central. si tomáramos muestras repetidas de una población dada, la media generalmente mostraría una menor fluctuación que la correspondiente a la mediana o a la moda. En otras palabras, la media generalmente proporciona una mejor estimación del parámetro correspondiente a la población.

Por otra parte, hay algunas situaciones en que se prefiere la mediana como medida de tendencia central. Cuando la distribución es simétrica, la media y la mediana son idénticas. En estas circunstancias deberá emplearse la media. Sin embargo, como hemos visto, cuando la distribución es visiblemente asimétrica la media proporciona una estimación falsa de la tendencia central. El ingreso anual por familia es una variable comúnmente estudiada en donde se prefiere usar la mediana en vez de la media, porque la distribución de esta variable está sesgada en dirección de las salarios altos, con el resultado de que la media sobreestima el ingreso obtenido por la mayoría de las familias.

La mediana es también la medida que se elige en las distribuciones en las cuales hay valores indeterminados. Para ilustrar, cuando hacemos caminar ratas en un laberinto, habrá ocasiones en que una o varias simplemente no caminen. Sus tiempos son por lo tanto, indeterminados. Sus "calificaciones" no pueden despreciarse simplemente ya que la característica de su no-caminata será de una significación considerable al evaluar los efectos de la variable independiente. En estas circunstancias, la mediana deberá ser empleada como medida de tendencia central.

La moda es la medida apropiada siempre que se desee una estimación aproximada rápida de la tendencia central, o cuando estamos interesados únicamente en el caso típico. La moda se usa rara vez en las ciencias de la conducta.

Medidas de dispersión

En discusiones anteriores, decíamos que una calificación por sí misma carece de significado, y sólo lo adquiere cuando se compara con otras calificaciones y otros estadísticos. Así, si conocemos la media de la distribución de una variable dada, podemos determinar cuándo una calificación particular es mayor o menor que dicha media. Pero ¿cuándo mayor o menor? Resulta evidente, a esta altura, que una medida de la tendencia central, como la media, sólo proporciona una cantidad limitada de información. Para describir una distribución en forma más completa o para interpretar con más detalle una calificación, necesitaremos información adicional acerca de la dispersión de las calificaciones con respecto a nuestra medida de tendencia central.

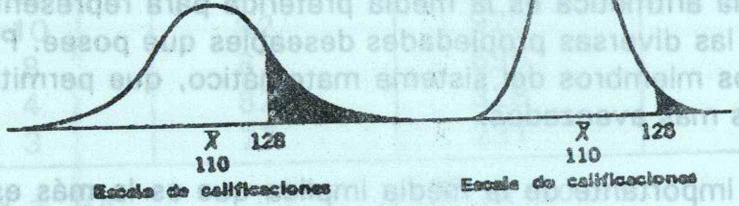

Fig. 6
Dos polígonos de frecuencia con medias idénticas pero diferente dispersión o variabilidad.

Consideremos en la gráfica de la figura 6 las partes (a) y (b). En ambos ejemplos de polígono de frecuencia, la media es exactamente la misma. Sin embargo, hay que

anotar la diferencia en la interpretación de la calificación 128. En (a), debido a que las calificaciones están bastante dispersas con respecto a la media, una calificación de 128 puede considerarse sólo moderadamente alta. Un buen número de individuos en la distribución obtuvo una calificación por encima de 128, como lo indica el área a la derecha de 128. En (b), por el contrario, las calificaciones están estrechamente distribuidas alrededor de la misma media. En una distribución más homogénea. En consecuencia, la calificación 128 está localizada casi en el máximo de la distribución y puede, por lo tanto, considerarse como una calificación muy alta.

Así pues, puede verse que para la interpretación de calificaciones individuales, debemos hallar una información auxiliar que acompañe a la media o a la mediana. Esta información adicional debe, en cierta forma, indicar el grado de dispersión de las calificaciones alrededor de la medida de tendencia central. Hay 5 medidas de dispersión o variabilidad: el rango, el rango intercuartil,, la desviación media, la varianza y la desviación estándar. De las cinco, discutiremos aquí la desviación estándar pues es la medida de dispersión más adecuada tanto para la estadística descriptiva como para la inferencial. En inferencia estadística avanzada, como en análisis de varianza, la varianza llegará a ser la medida de dispersión más adecuada.

Afortunadamente la desviación estándar, basada en la elevación al cuadrado de las calificaciones de desviación (diferencia entre un valor particular y la media aritmética), es de un extraordinario valor desde tres puntos de vista diferentes. (1) La desviación estándar representa dispersión de calificaciones por lo que la variabilidad de diferentes distribuciones puede compararse en términos de la desviación estándar (s). (2) La desviación estándar permite la interpretación precisa de las calificaciones dentro de una distribución. (3) La desviación estándar, como la media, es miembro de un sistema matemático que permite su utilización en consideraciones estadísticas más avanzadas. Tendremos más que decir acerca de los aspectos interpretativos de /s/ una vez que hayamos mostrado cómo se calcula.

La varianza se define verbalmente como la suma de los cuadrados de las desviaciones de las observaciones con respecto a la media dividida por N. Simbólicamente se representa así:

$$s^2 = \frac{\sum (X - \bar{X})^2}{N}$$

La desviación estándar es la raíz cuadrada de la varianza y se define así:

$$s = \sqrt{\frac{\sum (X - \bar{X})^2}{N}}$$

Para datos presentados en forma de distribución de frecuencia, la fórmula de la desviación estándar es:

$$s = \sqrt{\frac{\sum f\,(X - \bar{X})^2}{N}}$$

Observa que la f aparece en la fórmula para recordarle que cada $(X\text{-}\bar{X})^2$ se debe multiplicar por su correspondiente frecuencia antes de ser sumada. Aún cuando estemos utilizando una serie de calificaciones simples, esta fórmula es la más general porque la frecuencia de cada calificación es uno, es decir, $f = 1$. Por ésta razón, debemos considerar la f implícita aún cuando no se de.

O bien, la fórmula para calcular s con datos originales agrupados es

$$s = \sqrt{\frac{\sum fX^2}{N} - \bar{X}^2}$$

La tabla 7 resume el procedimiento para calcular s, que utiliza datos originales agrupados. Obsérvese que los valores fX^2 de la última columna, se pueden obtener elevando X al cuadrado y multiplicando por f, o simplemente multiplicando fX por X, como este último método requiere un paso menos y propicia en menor grado el error, los autores lo prefieren para obtener los valores de fX^2.

1 Intervalo de clase	2 f	3 Punto medio del intervalo	4 $f\chi$	5 $f\chi^2$	$\bar{X} = \frac{990}{66} = 15.00$
26-28	1	27	27	729	$s = \sqrt{\frac{\sum fX^2}{N} - \bar{X}^2}$
23-25	4	24	96	2304	
20-22	7	21	147	3087	$= \sqrt{\frac{16488}{66} - 15^2}$
17-19	12	18	216	3888	
14-16	18	15	270	4050	$= \sqrt{249.82 - 225,00}$
11-13	11	12	132	1584	
8-10	9	9	81	729	$= \sqrt{24.82}$
5-7	3	6	18	108	
2-4	1	3	3	9	
	N = 66		$\Sigma f\chi = 990$	16488	s = 4.98

Tabla 7
Procedimientos para calcular la desviación estándar a partir de distribuciones de frecuencia, usando, el método de datos originales.

Resumamos el cálculo de la desviación estándar a partir de una distribución de frecuencia:

Paso 1 Seguir todos los pasos necesarios para calcular la media a partir de datos agrupados.

Paso 2 Agregar una columna adicional, encabezada $f\chi^2$.

Paso 3 Multiplicar los valores de la columna $f\chi$ por los correspondientes valores de la columna χ y poner los resultados en la columna $f\chi^2$.

Paso 4 Sumar los valores de la columna $f\chi^2$ para obtener $\Sigma f\chi^2$

Paso 5 Substituir los valores $\Sigma f\chi^2$, N, $\bar{X}$, en la fórmula y efectuar las operaciones indicadas para encontrar la desviación estándar.

Una regla práctica para calcular la desviación estándar es que el cociente de dividir el rango por al desviación estándar, raramente es menor que 2 o mayor que 6. En nuestro ejemplo anterior, el cociente es 27/4,98, es decir 5,42. Si obtenemos una desviación estándar que nos proporciona un cociente mayor que 6 o menor que 2, casi seguramente hemos cometido un error. Si es mayor que 6, revisemos la columna $f\chi^2$ y la posición del punto decimal. Si es menor que 2, revisemos la posición del punto decimal.

Una buena comprensión del significado de la desviación estándar depende del conocimiento de la relación existente entre la desviación estándar y la distribución normal. Así, para poder interpretar las desviaciones estándar calculadas en este capítulo, es indispensable explorar la relación existente entre las calificaciones originales, la desviación estándar y la distribución normal. Presentamos este material a continuación.

La desviación estándar y la distribución normal estándar

Anteriormente notamos que las calificaciones obtenidas a partir de las escalas empleadas por los científicos que estudian el comportamiento, en general, carecen por sí mismas de significado a menos que se comparen con la distribución de las calificaciones de algún grupo de referencia. Claro está que las calificaciones obtenidas a partir de cualquier escala, incluyendo las empleadas por los físicos, llegan a tener un mayor significado cuando se comparan con un grupo de referencia, sea de personas o de objetos. Así, si nos dijeran que un pescador canadiense obtuvo una trucha de 22 kilos de peso, nos sorprenderíamos o no, según nuestros conocimientos acerca del peso usual de esta clase de pez. Sin embargo, una vez establecido un grupo de referencia, dicha medición adquiere sentido. Puesto que la mayoría de las truchas pesan alrededor de 4,5 kilos y en raras ecuaciones alcanzan los 10 kilos, el logro de nuestro pescador apócrifo se debe considerar exagerado.

Al interpretar una calificación única, queremos colocarla en algún lugar con respecto a la serie de calificaciones de un grupo de referencia. Anteriormente se aprendió a ubicar una calificación en su posición respectiva, determinando su rango percentil. Se recordará que el rango percentil de una calificación nos dice el porcentaje de calificaciones que son menores. Otro camino para la interpretación de una calificación única, podría ser observarla en relación con un punto central, como la media. Por tanto una calificación de 20 en una distribución con una media de 23 se puede expresar como -3. Finalmente, podremos expresar la desviación de la calificación en términos de unidades de desviación estándar. Así, si nuestra desviación estándar es 1,5 la calificación de 20 estaría dos desviaciones estándar por debajo de la media (es decir, -3/1,5 = -2). Este proceso de dividir la desviación de una calificación con respecto a la media por la desviación estándar, se conoce como la transformación hacia calificaciones estándar z, y se define como:

$$z = \frac{X-\bar{X}}{s}$$

Se notará que cada calificación en la distribución se puede transformar en una calificación ż, en donde cada ż representará la desviación de una calificación específica, con respecto a la media expresada en unidades de desviaciones estándar.

¿Cuál es la importancia de la transformación a una calificación ż? Simplemente ésta: si la población de las calificaciones de una variable dada es normal, podemos expresar cualquier calificación como un rango percentil refiriendo nuestra ż a la distribución normal estándar. Además, puesto que las calificaciones ż representan números abstractos (adimensionales) en oposición a los valores concretos de las calificaciones originales (centímetros, kilos, coeficientes, I.Q., etc.), podemos comparar la posición de un individuo en una variable, con su posición en una segunda variable. Para entender esta importante característica de las calificaciones ż, tenemos que referirnos a la distribución normal estándar.

La distribución normal estándar tiene un área total igual a 1.00. Hay una proporción fija de casos entre una línea vertical, u ordenada, erigida en cualquier punto y una ordenada erigida en cualquier otro punto.

(La proporción de casos entre dos valores dados de la variable es una constante). Tomando unos cuantos puntos de referencia a lo largo de una curva normal, podemos enunciar lo siguiente:

1. Entre la media (z = 0) y una calificación estándar(z = 1) por encima de la media se encuentra el 34.13% de todos los casos. Análogamente, el 34.13% de todos los casos se encuentra entre la media y una calificación estándar por debajo de la media. Dicho de otra manera, 34.13% del área bajo la curva se encuentra entre la media y una calificación estándar por encima de la media, y 34.13% del área está comprendida entre la media y menos una calificación estándar.

2. Entre la media y 2 calificaciones estándar por encima de la media, se encuentra el 47.72% de los casos. Puesto que la curva normal es simétrica, 47.72% del área también está comprendida entre la media y -2 calificaciones estándar.

3. Finalmente entre la media y 3 calificaciones estándar por encima de la media se encuentra el 49.87% de los casos. Análogamente, el 49.87% de los casos está ubicado entre -3 calificaciones estándar. Estas relaciones se ilustran en la fig. 7.

Ahora, al transformar las calificaciones de una variable normalmente distribuida en calificaciones ż expresamos en realidad estas calificaciones en unidades de la curva normal estándar. Para cualquier valor dado de x con una cierta proporción de área más allá de este, existe un valor correspondiente de ż con la misma proporción de área más allá de él. Por ejemplo, si tenemos una población en la cual $\bar{X} = 30$ y s = 10, la z de la calificación en la media ($\bar{X} = 30$) será igual a cero, y las calificaciones z que están 1 desviación estándar por encima y por debajo de la media (X = 40 y X = 20) serán + 1.00 y -1.00, respectivamente.

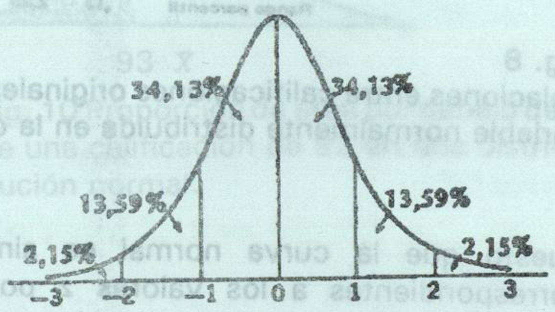

Fig. 7
Areas entre puntos seleccionados bajo la curva normal.

Para obtener claridad en la exposición, restringimos la anterior discusión del área bajo la curva normal a algunos puntos escogidos. Para efectos prácticos, sin embargo, es posible determinar el porcentaje de las áreas entre dos puntos cualesquiera, haciendo uso de los valores tabulados del área bajo la curva normal. (tabla A). La columna de la izquierda encabezada por ż representa la desviación respecto a la media expresada en unidades de desviación estándar. Refiriéndonos al cuerpo de la tabla, podemos determinar la proporción del área total que se encuentra entre una calificación dada y la media (columna B), y el área más allá de una calificación dada (columna C). Así, si un individuo obtuvo una calificación de 24.65 en una variable normalmente distribuida con $\bar{X} = 16$ y s = 5, su calificación ż sería:

$$z = \frac{24,65-16}{5} = 1.73$$

Remitiéndonos a la comuna B en la tabla A, encontramos 0.4582 o sea que 45.82%[3] del área está situada entre dicha calificación y la media. Puesto que en una distribución simétrica 50% del área también está situada por debajo de la media, podemos concluir que el 95.82% del área total está ubicada por debajo de una calificación de 24.65. Nótese que ahora podemos interpretar esta calificación como un rango percentil de 95.82.

Supongamos que otro individuo obtuvo una calificación de 7,35 en la misma variable normalmente distribuida. Su calificación ż sería:

$$z = \frac{7,35-16}{5} = -1,73$$

[3] Las áreas bajo la curva se expresan como proporciones del área. Para convertirlas en porcentajes del área, multiplica por 100 o simplemente, mueve la coma del decimal dos lugares hacia la derecha.