Rev. EDUCACION U.C.R., Vol. VI, No. 1, 1982. Págs. 47–61

METODOS CUANTITATIVOS PARA MEDICION EDUCATIVA

Juan Manuel Esquivel

INTRODUCCION

Este trabajo tiene dos propósitos: el primero es presentar técnicas estadísticas y de análisis de los resultados de mediciones; el segundo es demostrar que estas técnicas pueden ser aplicadas para mejorar las pruebas y su desarrollo.

Las técnicas aquí presentadas son sencillas. Los datos obtenidos en la aplicación de pruebas hechas por el docente, no deben ser tratados con técnicas sofisticadas; pero han de ser suficientemente válidas para que, al usarse e interpretarse, beneficien el proceso de enseñanza y aprendizaje.

El trabajo se dividirá en dos secciones: tratamiento de los resultados en una prueba y análisis de los ítemes.

Tratamiento de los resultados de una prueba

Cuando un docente administra una prueba escrita o de ejecución, obtiene puntajes que son el resultado de la suma de los puntos obtenidos por cada estudiante en la prueba. Estos puntajes se encuentran en desorden, por lo que la primera tarea es organizarlos. Con este fin, se ordenan de mayor a menor puntaje o viceversa y se apunta el número de veces que se repite cada puntaje. Esta forma de organizar los puntajes se les llama distribución de frecuencias.

Lo anterior se observa en las siguientes tablas. Así, en la Tabla 1 aparecen los resultados de una prueba (en desorden), practicada a un grupo de 32 estudiantes y en la Tabla 2, estos mismos datos organizados en forma de una distribución de frecuencias.

TABLA No.1
PUNTAJES CORRESPONDIENTES A 32 ESTUDIANTES

32, 36, 17, 18, 25, 38, 35, 25, 22, 24, 26, 28, 10, 13, 30, 20, 25,18,
28, 31, 25, 11, 15, 36, 16, 34, 30, 31, 29, 11, 34, 16.

TABLA No.2
DISTRIBUCION DE FRECUENCIAS PARA LOS PUNTAJES DE LOS 32 ESTUDIANTES

PUNTAJE	MARCAS	FRECUENCIA	
X		f	
38	I	1	
36	II	2	
35	I	1	
34	II	2	
32	I	1	
31	II	2	
30	II	2	Mo = 25
29	I	1	Md = 25
28	II	2	
26	I	1	$\bar{X}$ = 24.66
25	IIII	4	
24	I	1	
22	I	1	
20	I	1	
18	II	2	
17	I	1	
16	II	2	
15	I	1	
13	I	1	
11	II	2	
10	I	1	
		N = 32	

Medidas de tendencia central

Las medidas de promedios son muy útiles para la interpretación de las mediciones hechas en el trabajo de clase. Por este medio, el docente puede comparar los resultados de la aplicación de una misma prueba a dos o más grupos, o puede comparar los resultados que produjo una prueba, cuando fue aplicada durante dos años consecutivos a grupos de un mismo nivel. Los promedios, entonces, sirven como punto de referencia; de esta manera, se dice que el estudiante o el grupo está "por debajo" o "por encima" de alguna de las medidas de tendencia central (promedios).

En este estudio, se contemplan tres promedios: la moda, la mediana y la media aritmética.

La moda (Mo)

La moda es el puntaje que más veces se repite, o sea, el de mayor frecuencia. En la distribución de la Tabla 1, la moda es el puntaje 25, pues se repite cuatro veces.

La mediana (Md)

Es el puntaje bajo el cual, está el 50% de los puntajes. Para encontrar la mediana se necesita tener los puntajes ordenados de mayor a menor, tal y como aparece en la Tabla 2. Si el número de puntajes (N) es impar, para encontrar la mediana debemos agregar una unidad a N y dividir el resultado entre 2, así:

$$a = \frac{N + 1}{2} \rightarrow \quad (1)$$

Luego se debe contar desde el puntaje más bajo hacia arriba hasta encontrar el puntaje "a". Ese será la mediana. Por ejemplo, para los siguientes puntajes: 35, 33, 32, 30, 27, 24, 21, 18, 17, 15, 13; la mediana se encontrará así:

$$a = \frac{N + 1}{2} = \frac{11 + 1}{2} = 6$$

Si contamos seis puntajes a partir del más bajo, la mediana será 24.

Si el número de puntajes (N) es par, "a" se encontrará, simplemente, dividiendo N entre 2, así:

$$a = \frac{N}{2} \quad (2)$$

Pero en este caso, después de contar de abajo hacia arriba y de encontrar el puntaje "a", se debe sumar a "a", el siguiente puntaje y dividir entre 2, para obtener la mediana. En esos puntajes: 19, 17, 16, 14, 11, 10, 9, 9, 8, 6.

$$a = N/2 = 10/2 = 5$$

El quinto puntaje de abajo hacia arriba es 10 y el siguiente es 11, por lo tanto:

$$\frac{10 + 11}{2} = \frac{21}{2} = 10.5 \text{ es la mediana}$$

La media Aritmética

Para obtener la media aritmética se suman todos los puntajes de la distribución y luego esa suma se divide entre el número de puntajes (N).

Media aritmética ($\bar{X}$) =

$$\frac{\text{Suma de los puntajes} \quad (\Sigma X)}{\text{Número de puntajes} \quad (N)} \quad (3)$$

La suma de los puntajes de la Tabla 2 es igual a 789, por lo tanto:

$$\bar{X} = \frac{789}{32} = 24.66$$

Medidas de dispersión

Además de conocer los promedios (media aritmética, mediana y moda) de un grupo de puntajes, es necesario, para su mejor interpretación, que el docente conozca en alguna medida la dispersión de los mismos.

Mientras los promedios son medidas que definen la posición central de la distribución, porque señalan un punto de equilibrio en la distribución, las medidas de dispersión indican cuánto se extienden los puntajes desde la medida de tendencia central, o sea, que son medidas de la variabilidad de los puntajes. Proveen una medida de la homogeneidad de los puntajes y por lo tanto, dan una mejor base para comparar diferentes grupos de estos.

La amplitud o rango (A)

La amplitud se calcula obteniendo la diferencia entre el puntaje más alto y el puntaje más bajo. En el caso de los puntajes de la Tabla 2, la amplitud será: 38 – 10 = 28. La amplitud es útil para tener una idea general de la variabilidad de los puntajes; pero frecuentemente da una descripción no confiable de la misma, así por ejemplo, podemos ver que al comparar dos grupos de puntajes en la tabla 3, en ambos grupos la amplitud es 18; aunque en B los puntajes están dispersos del 19 al 1, en el grupo A hay 7 puntajes juntos y uno muy separado del resto.

TABLA No.3

DOS GRUPOS DE PUNTAJES CON IGUAL AMPLITUD

A	B
19	19
8	17
7	14
6	11
4	7
4	5
3	4
1	1
A = 18	A = 18

La desviación estándar (S)

En el caso de la tabla 3, se puede obtener un mejor cálculo de la dispersión de los puntajes si empleamos otra medida de la variabilidad: la desviación estándar. Su cálculo se basa en la diferencia entre cada puntaje y la media aritmética. Cuanto más grande sean esas diferencias, más dispersos estarán los puntajes y por lo tanto, más grande será la desviación estándar. Esta se calcula por la siguiente fórmula:

$$S = \sqrt{\frac{\Sigma(\bar{X} - \bar{X})^2}{N - 1}}$$

donde:

$(X - \bar{X})^2$;

suma de los cuadrados de la diferencia entre cada puntaje y la media aritmética

$N - 1$;

número de casos menos 1.

El cálculo de la desviación estándar obtenida por la fórmula citada es un trabajo laborioso; por lo que se aconseja el uso de otro procedimiento, sugerido por Mason y Oder (1968) y que da como resultado un valor aproximado bastante preciso.

La fórmula de esta estimación de la desviación estándar es la siguiente:

$$s = \frac{\dfrac{\text{Suma de la 1/6 parte superior de los puntajes} - \text{Suma de la 1/6 parte inferior de los puntajes}}{}}{\dfrac{N-1}{2}} \qquad (5)$$

El primer paso para aplicar esta fórmula es calcular a cuántos puntajes corresponde la 1/6 parte de los mismos. Luego, se suman los puntajes correspondientes a la 1/6 parte superior y se restan de la suma de la 1/6 parte inferior. El resultado de esta operación se divide entre la mitad de los N – 1 puntajes.

En el ejemplo de la tabla 2, hay 32 puntajes; por lo tanto una sexta parte será igual a 32/6 = 5.33; así tendrán que usarse para los cálculos los 5 y 1/3 de los puntajes superior e inferior. De acuerdo con la fórmula se tiene:

$$s = \frac{(38 + 36 + 36 + 35 + 34 + (34 \times 0.33)) - (10 + 11 + 11 + 13 + 15 + (16 \times 0.33))}{\dfrac{32-1}{2}}$$

$$s = \frac{190.22 - 65.28}{15.5} = 8.06$$

Este valor de la desviación solamente se diferencia en un 0.37 % del valor calculado por la fórmula No.4, que es 8.09. Esta pequeña diferencia no es la suficientemente grande para producir errores serios en la interpretación y uso de los datos en medición educativa.

La desviación estándar tiene mucha utilidad en la interpretación de los resultados de las mediciones. En primer lugar, sirve para comparar los resultados en una misma prueba aplicada a dos o más grupos; pues la desviación estándar indica la homogeneidad relativa de los grupos, en la característica que mide la prueba. De manera que cuando el docente administra una prueba a dos grupos y obtiene una desviación estándar de 8 en el primer grupo, y en el segundo, ésta tiene un valor igual a 12; estos resultados indican que en el segundo grupo existen mayores diferencias individuales que en el primero.

La desviación estándar también puede ser usada para dar un valor relativo a varias pruebas, al hacer un promedio y al finalizar un curso o un período lectivo determinado; de manera tal que las pruebas con más altos valores de la desviación estándar serán las pruebas que más cuenten al hacer los promedios.

La desviación estándar sirve también, para comparar los resultados obtenidos por un estudiante en pruebas sucesivas. Aunque las desviaciones estándar de dos pruebas no se pueden comparar directamente, sí podemos usar para esa comparación puntajes derivados, en los que se emplea la desviación estándar, la media aritmética y el puntaje del estudiante en la prueba.

Puntajes derivados

No se puede conocer el rendimiento relativo de un estudiante con respecto al rendimiento del grupo por el puntaje obtenido en la prueba. Tampoco podemos conocer el rendimiento relativo de un estudiante en dos pruebas, pues los puntajes de las pruebas están dados en diferentes escalas. Para remediar esta falta de comparabilidad, se usa información del rendimiento del grupo, combinada con los puntajes obtenidos por los estudiantes y así,se obtienen puntajes derivados.

Una clase de puntaje derivado es el puntaje estándar. Este se obtiene mediantela siguiente fórmula:

$$z = \frac{\text{puntaje } (X) - \text{media aritmética } (\bar{X})}{\text{desviación estándar } (S)} \qquad (6)$$

En la primera prueba, con X = 20 y S = 5, tres estudiantes obtuvieron los siguientes puntajes:

Estudiante 1, X = 33; $Z = \frac{33 - 20}{5} = \frac{13}{5} = 2.60$

Estudiante 2, X = 20; $Z = \frac{20 - 20}{5} = \frac{0}{5} = 0.00$

Estudiante 3, X = 15; $Z = \frac{15 - 20}{5} = \frac{-5}{5} = 1.00$

En la segunda prueba, con X = 35 y S = 9, esos mismos tres estudiantes obtuvieron los siguientes puntajes:

Estudiante 1, X = 39; $Z = \frac{39 - 35}{9} = 0.44$

Estudiante 2, X = 35; $Z = \frac{35 - 35}{9} = 0.00$

Estudiante 3, X = 25; $Z = \frac{24 - 35}{9} = 1.22$

Del primer estudiante, podemos deducir que tuvo en la primera prueba, un rendimiento mejor que en la segunda. El segundo estudiante obtuvo igual rendimiento en ambas pruebas. Por último, el tercer estudiante está por debajo del promedio en ambas pruebas, aunque su rendimiento es más bajo en la segunda. Cuando se convierte un grupo de puntajes a puntajes estándar, se obtiene una distribución con una media aritmética igual a 0 y una desviación estándar de 1.

Otra clase de puntaje derivado es el rango percentil. Este se basa en el porcentaje de estudiantes que obtuvieron puntaje menor que el de un determinado estudiante. Para calcularlo se usa la siguiente fórmula:

$$RP = \frac{\text{número de puntajes debajo del puntaje X}}{\text{número total de puntajes}} \times 100 \quad (7)$$

Usando la Tabla 2, podemos calcular el rango percentil del puntaje 24. Si contamos el número de puntajes por debajo de 24, encontramos que hay 12 puntajes, aplicando la fórmula, tenemos que:

$$RP = \frac{12}{32} \times 100 = 37.5$$

Para el puntaje 32, el número de puntajes por debajo del 32 es 24; por lo tanto el rango percentil es:

$$RP = \frac{24}{32} \times 100 = 75$$

Este método simplificado es un poco diferente de los que se presentan en textos de estadística; aunque es adecuado para usarlo en el análisis de los resultados obtenidos por la aplicación de los instrumentos de medición hechos por el docente, según Scanell y Tracy (1975).

Medidas de relación

Las medidas de correlación señalan cuándo dos grupos de puntajes se relacionan, esto es, hasta dónde los estudiantes tienen las mismas posiciones relativas en los dos grupos de puntajes. El concepto de correlación es de muy variadas aplicaciones en Medición Educativa. Entre esas aplicaciones tenemos cálculos de confiabilidad y de validez de las pruebas.

Cuandos dos grupos de puntajes se relacionan en forma directa y perfecta el índice numérico o coeficiente es de 1.00; cuando no hay relación, el coeficiente es 0; por último, cuando la relación es perfecta pero inversa, el coeficiente es −1.00. En la mayoría de las ocasiones, las relaciones que se dan entre mediciones de interés para el docente tendrán valores entre 0 y 1.00. Raramente se obtienen valores negativos; un ejemplo de esto es: la relación entre la edad de un grupo de estudiantes, con el nivel de aprovechamiento escolar.

Existen diferentes maneras de calcular coeficiente de correlación. En este trabajo se ofrece un procedimiento sencillo y fácil de aplicar, (Wolfe, 1971) que da como resultado un índice aproximado. En la tabla 4, aparecen los resultados de dos pruebas aplicadas a un mismo grupo de estudiantes. Usando esos datos, vamos a calcular el coeficiente de correlación.

Para calcular el coeficiente, se debe seguir el siguiente procedimiento:

1.– Clasificar en cada grupo, los puntajes altos y bajos. El número de puntajes que se clasifican debe decidirlo el docente, aunque una cuarta parte es suficiente. En nuestro ejemplo se han clasificado los ocho puntajes más altos de cada prueba, señalados con una "A", y los puntajes más bajos, señalados con una "B".

TABLA No.4

PUNTAJES DE DOS PRUEBAS USADOS PARA EL CALCULO DEL COEFICIENTE DE CORRELACION

	PUNTAJES		CLASIFICACION	
ESTUDIANTE	PRUEBA 1	PRUEBA 2	PRUEBA 1	PRUEBA 2
1	25	51		A
2	27	35		
3	10	22	B	B
4	14	20	B	B
5	37	60	A	A
6	40	55	A	A
7	38	57	A	A
8	35	51	A	A
9	22	48		
10	18	56	B	A
11	17	26	B	B
12	20	30		
13	25	29		
14	29	38		
15	33	28	A	B
16	30	37		
17	12	24	B	B

18	16	27	B	B
19	22	42		
20	26	35		
21	35	49	A	
22	18	41		
23	15	49		
24	37	50	A	A
25	31	54	A	A
26	25	40		
27	15	25	B	B
28	17	27	B	B
29	23	30		
30	28	33		

2.– Una vez clasificados los puntajes más altos y más bajos, se debe hacer una tabla como la que aparece en la No.5. Esta tabla muestra los cuatro tipos de clasificación y el número de casos que caen dentro de cada una de las celdas. Así, por ejemplo, en la celda AA están los estudiantes 5, 6, 7, 8, 24 y 25 ya que tienen clasificación "A" en ambas pruebas; en la celda AB está únicamente el estudiante 15, quien es clasificado con "A" en la primera prueba y "B" en la segunda, y así sucesivamente, como se puede ver en la tabla No.5.

TABLA No.5
TABLA DE CLASIFICACION PARA EL CALCULO DEL COEFICIENTE DE CORRELACION

		PRUEBA 2 A	PRUEBA 2 B
PRUEBA 1	A	AA 6	AB 1
	B	BA 1	BB 7

Se puede notar que los estudiantes 1 y 21 no son colocados en ninguna de las celdas, pues únicamente son clasificados en una de las pruebas.

3.– Se aplica la fórmula para la correlación:

$$r = \frac{(\text{suma de AA y BB}) - (\text{suma de AB y BA})}{2\,n} \qquad (8)$$

En la fórmula, "n" es el número de altos o bajos clasificados en uno de los grupos; en nuestro ejemplo n = 8, aplicando la fórmula a los datos de la Tabla No.5, se tiene:

$$r = \frac{(6+7)-(1+1)}{2 \times 8} = \frac{11}{16} = 0.69$$

Este resultado es una estimación del coeficiente de correlación. Este coeficiente no se puede interpretar como una proporción; es decir, un coeficiente de 0.75 no significa, que hay un 75% de la correlación perfecta. El coeficiente muestra cuántos puntajes bajos y altos de una primera prueba se asocian con puntajes bajos y altos de la segunda prueba. Cuanto más alto sea el índice, más fuerte será la correlación. De esta manera, un "r" mayor de $\pm$ 90 significa una relación muy fuerte, y una "r" menor de $\pm$ 30 significa una relación bastante débil. No se debe olvidar que la correlación, cuando tiene valores diferentes de 1, 0 y –1 es un descriptor de tendencias generales. En nuestro ejemplo, una r = 0.69 entre la prueba 1 y la prueba 2, significa que los estudiantes que logran puntajes altos en la primera prueba, tienden a obtener puntajes altos en la segunda y los que tienen puntajes bajos, a continuar con estos.

Confiabilidad

Una prueba confiable es aquella que mide con consistencia. La confiabilidaddde una prueba se obtiene indirectamente, pues en su definición se encuentran términos muy difíciles de obtener. La definición de confiabilidad se deduce de la teoría de medición, cuya ecuación básica establece que el puntaje observado es el resultado del puntaje verdadero de la prueba, más el error producido en la medición.

$$X_{ob} = X_v + E \qquad (9)$$

Esta ecuación, cuando se aplica a un grupo de puntajes, puede ser escrita en términos de la varianza. O sea de la desviación estándar al cuadrado.

$$S^2_{ob} = S^2_v + S^2_E \qquad (10)$$

O sea, que la "varianza" de los puntajes observados es igual a la suma de las "varianzas" de los puntajes verdaderos y de la "varianza" del error. Si dividimos ambos lados de la ecuación entre S^2_{ob}, resulta:

$$\frac{S^2_{ob}}{S^2_{ob}} + \frac{S^2_v}{S^2_{ob}} + \frac{S^{2E}}{S^2_{ob}} = 1 \qquad (11)$$

Se define confiabilidad como la relación que existe entre la varianza verdadera y la varianza observada.

$$r_{tt} = \frac{S^2_v}{S^2_{ob}} \qquad (12)$$

Sustituyendo en la ecuación 11, $\frac{S^2_v}{S^2_{ob}}$ por su valor, se obtiene

$$r_{tt} = 1 - \frac{S^2_E}{S^2_{ob}} \qquad (13)$$

$$r_{tt} = \frac{S^2_{ob} - S^2_E}{S^2_{ob}} \qquad (14)$$

Para aplicar esta última fórmula, se puede calcular la S^2_{ob} de la distribución de puntajes y el $S^2 E$ de la distribución de discrepancias entre pares de puntajes de una prueba aplicada dos veces.

Ya que hay interés en la distribución de pares de puntajes para un grupo de individuos, la medida apropiada de consistencia (confiabilidad) es el coeficiente de correlación. Las medidas de confiabilidad, en su mayoría, se expresan como coeficientes de correlación, que se llaman, entonces, coeficientes de confiabilidad. Comúnmente se usan tres variedades de coeficientes de confiabilidad:

1.– Coeficiente de estabilidad. Se aplica cuando se asume que es estable la característica medible, por lo que la misma prueba, se administra dos veces a un grupo y se calcula el coeficiente de correlación a los puntajes.

2.– Coeficiente de equivalencia. Se aplica para conocer la consistencia de los puntajes de dos formas de una misma prueba. Al usar el coeficiente de correlación a los puntajes obtenidos, después de la administración de las dos pruebas al mismo grupo, se obtiene el coeficiente de equivalencia.

3.– Por último las medidas de consistencia interna, indican si los ítemes de una prueba miden una misma característica (prueba homogénea), o si la prueba mide varias características (heterogénea).

Las situaciones de medición comunes del docente le impiden aplicar una prueba dos veces o tener dos pruebas paralelas para estimar la confiabilidad. Afortunadamente existen variosmétodos para estimar la confiabilidad.Uno de los más comunes es dividir la prueba en dos mitades. Esta división se hace después de aplicar la prueba. Para satisfacer las condiciones de un coeficiente de equivalencia, las dos mitades deben ser iguales en contenido, dificultad, media aritmética y desviación estándar. Comúnmente, estas mitades se pueden hacer usando los ítemes pares para una mitad y los impares para la otra mitad. Este método se puede aplicar para pruebas con ítemes del mismo tipo. Si la prueba tiene, por ejemplo, ítemes de selección e ítemes de desarrollo, primero se debe partir la prueba en dos mitades, tomando en cuenta únicamente los ítemes de selección, y luego se calcula la confiabilidad. Después, se sigue el mismo procedimiento con los ítemes de desarrollo (únicamente). En este segundo caso, se debe tener todavía más cuidado de no irrespetar las condiciones antes señaladas.

Siguiendo el procedimiento explicado para calcular el coeficiente de correlación, se presenta la Tabla 6 donde aparecen los puntajes correspondientes de las mitades de una prueba.

	IMPARES A	IMPARES B
PARES	AA 8	AB 0
	BA 0	BB 6

TABLA No.6

CALCULO DEL COEFICIENTE DE CORRELACION

PARA LAS DOS MITADES DE UNA PRUEBA

ESTUDIANTES	PUNTAJES PARES	PUNTAJES IMPARES	CLASIFICACION PARES	CLASIFICACION IMPARES
1	8	8		
2	8	7		
3	7	8		
4	10	11	A	A
5	9	10	A	A
6	6	2		B
7	11	9	A	A
8	2	1	B	B
9	5	7		
10	12	12	A	A
11	0	3	B	B
12	1	2	B	B
13	6	7		
14	4	5		
15	3	8	B	
16	10	9	A	A
17	1	0	B	B

18	2	1	B	B
19	8	4		
20	7	8		
21	6	5		
22	5	7		
23	9	11	A	A
24	3	1	B	B
25	2	2	B	B
26	11	10	A	A
27	7	7		
28	4	8		
29	10	9	A	A
30	8	6		

Aplicando la fórmula 8 a los datos de la Tabla 6, se obtiene:

$$r_{pi} = \frac{(8 + 6) - (0 + 0)}{2 \times 8} = \frac{14}{16} = 0.88$$

El valor 0.88 es una estimación del coeficiente de correlación entre las mitades de la prueba. Como la confiabilidad depende de la longitud de la prueba, este valor es más bajo que el obtenido de la prueba completa. Para estimar el coeficiente de confiabilidad de la prueba completa, a partir de la correlación de las mitades, se usa la fórmula de Spearman-Brown (Brown, p.73, 1976).

$$r_{tt} = \frac{2r_{pi}}{1 + r_{pi}} \qquad (15)$$

La confiabilidad será, entonces:

$$r_{tt} = \frac{2\ (0.88)}{1 + 0.88} = 0.94$$

El valor encontrado indica que la prueba es muy confiable, ya que el valor es mayor que 0.70, (mínimo aceptable para pruebas de rendimiento académico).

También hay métodos de fácil aplicación para estimar la consistencia interna de pruebas hechas por el docente. La primera es la fórmula 21 de Kuder y Richarson (KR-21) (1967), quien usa como elementos para su cálculo el número de ítemes, la desviación estándar y la media aritmética.

$$r_{tt} = \frac{K}{K - 1} \left[1 - \frac{\bar{X}\ (K - \bar{X})}{K\ s^2} \right] \qquad (16)$$

En esta fórmula:

K = número de ítemes

$\bar{X}$ = media aritmética

s^2 = varianza de los puntajes

r_{tt} = consistencia interna de la prueba

Los datos de la prueba, cuya distribución aparece en la Tabla 2, son los siguientes:

K = 40

$\bar{X}$ = 24.66

s^2 = $(8.06)^2$ = 64.96

Aplicando la fórmula 16, resulta

$$r_{tt} = \frac{40}{40 - 1} \left[1 - \frac{24.66\ (40 - 24.66)}{40\ (64.96)} \right] = 0.877$$

También se puede emplear otra fórmula, aún más simple, para calcular el coeficiente de consistencia interna. Esta fórmula es una aproximación a KR-21, presentada por J. L. Saupe (1961). Es la siguiente:

$$r_{tt} = 1 - \frac{0.19\ K}{s^2} \qquad (17)$$

Aplicándola a los datos anteriores, se observa que:

$$r_{tt} = 1 - \frac{0.19\ (40)}{64.95} = 0.883$$

Este valor es un poco más alto que el obtenido por la fórmula KR–21. Estas fórmulas sólo se pueden aplicar a pruebas que tengan únicamente ítemes objetivos (o sea ítemes que se califican con 0 ó 1). Esto implica que si el docente tiene una prueba que está compuesta por ítemes objetivos e ítemes de desarrollo, debe considerar la prueba como compuesta de dossub-pruebasy además, calcular la consistencia interna para la sub-prueba que tiene ítemes objetivos usando KR-21 o Saupe. Para calcular la consistencia interna de la sub-prueba compuesta por ítemes de desarrollo, deberá usar la fórmula dada por Cronbach (1967) Esta fórmula es:

$$\alpha = \frac{K}{K-1} \quad 1 - \frac{s^2{}_i}{s^2{}_t} \qquad (18)$$

En ella:

K = número de ítemes

$s^2{}_i$ = suma de las "varianzas" de los ítemes

$s^2{}_t$ = "varianza" total de la prueba

= coeficiente de consistencia interna

TABLA No.7
VARIANZAS DE LOS ITEMES DE DESARROLLO
Y VARIANZA DE LA SUB-PRUEBA

"ITEMES" Nª	1	2	3	4	5	6	7	
"VARIANZA"	0.3969	0.6561	1.96	1.764	0.5625	1.5625	0.3035	$s^2{}_t$ = 36.845 $s^2{}_i$ = 7.205

Usando la fórmula de Cronbach Alfa a los datos de la Tabla 7, resulta:

$$\alpha = \frac{7}{7-1}\left(1 - \frac{7.205}{36.845}\right) = 0.94$$

Este valor indica una alta consistencia interna de la sub-prueba compuesta por los siete ítemes de desarrollo.

Análisis de ítemes

El análisis de ítemes incluye por un lado, el cálculo de dos índices: nivel de dificultad e índice de discriminación, por otro, la interpretación de los resultados de éstos. El análisis de ítemes tiene tres propósitos:

1– evaluar la efectividad de la instrucción;

2– estimar lo que han aprovechado los estudiantes y el grupo por medio de su ejecución en cada "item", y

3– mejorar la calidad de las pruebas futuras, pues por medio del análisis, se identifican ítemes valiosos para ser almacenados y ser usados nuevamente.

Nivel de dificultad para ítemes objetivos

El nivel de dificultad indica la proporción de individuos que contestan un "ítem" correctamente. Para el análisis, por razones teóricas, se recomienda tomar el 27% de las mejores pruebas y el 27% de las pruebas de más bajo puntaje; pero el docente puede usar el 25% de los más altos y el 25% de los más bajos, siempre que el número de pruebas sea mayor de 40. Cuando el número de pruebas es menor de 20, se deben usar todas para el análisis, después de dividir el grupo en altos y bajos. Con este fin, se usa la mediana de los puntajes.

Cuando el docente ordena las pruebas de mayor a menor y separa el 27% más alto y el 27% más bajo, se puede dar el caso de empates de puntajes. Esto se resuelve escogiendo al azar los que deben ser analizados.

TABLA No.8
RESULTADOS DE UNA PRUEBA PRACTICADA A 38 ESTUDIANTES

X	f	27% Superior	27% Inferior
32	1	32	
31	2	31 , 31	
30	1	30	
28	3	28 , 28 , 28	
27	2	27 , 27	
25	3	25 , 25 , 25 (Dos de los tres, escogidos al azar)	
24	2		
23	5		
21	1		
20	1		
19	2		
17	2		
15	2		
14	1		14
13	3		13 , 13 , 13
12	2		12 , 12
10	3		10 , 10 , 10
9	1		9
8	1		8
	N = 38		

Usando los resultados de la prueba que aparecen en la Tabla 8, se puede escoger el 27% superior y el 27% inferior. Como el grupo consta de 38 estudiantes, 10.26 son el 27%; pero, para usar números enteros, se escogen los 11 mejores puntajes y los 11 peores.

En este caso, los 12 mejores incluyeron del puntaje 32 al 25; pero sólo se necesitan 11, por lo que al azar se escogen dos pruebas con puntaje 25 entre los tres que hay en el grupo. Los seleccionados pueden observarse en la Tabla 8.

Una vez escogidas las pruebas de puntajes altos y bajos, se cuentan las respuestas correctas que se han obtenido para cada pregunta, tanto para los estudiantes del grupo alto como para los del grupo bajo. Para el cálculo del nivel de dificultad se emplea la siguiente fórmula:

$$N.D. = \frac{A + B}{N} \times 100 \qquad (19)$$

En ella:

A = número de respuestas correctas del grupo alto

B = número de respuestas correctas del grupo bajo

N = número de exámenes analizados

Así, por ejemplo, en la Tabla 9, están los resultados del análisis de ocho ítemes de selección, tomados de la prueba anterior.

TABLA NO.9
RESULTADOS DEL ANALISIS DE 8 ITEMES DE SELECCION

ITEM	A	B	ND	I.D.
1	11	10	95.5%	0.09
2	5	2	31.8%	0.27
3	8	4	54.5%	0.36
4	7	9	72.7%	-0.18
5	10	5	68.1%	0.46
6	0	4	18.2%	-0.36
7	10	3	59.1%	0.64
8	6	4	45.5%	0.18

Los ítemes presentan dificultades muy variadas. Por ejemplo, el ítem 1 es fácil, mientras que el 6 es difícil. Los demás pueden clasificarse como ítemes de dificultad intermedia, pues sus niveles de dificultad están en el rango del 30% al 70%. Aunque se ha considerado que la dificultad ideal es la del 50% lo cual significa que los "ítemes tienen la máxima varianza" y por lo tanto, el máximo potencial para correlacionar con otros ítemes. Para uso en la clase, se recomiendan ítemes con dificultades de alrededor de 0.70, pues estos influyen positivamente en variables importantes del proceso de enseñanza aprendizaje. (Marzo, 1969).

Indice de discriminación para ítemes objetivos

Un ítem que es contestado correctamente por los alumnos del grupo alto e incorrectamente por los estudiantes del grupo bajo, es un ítem que discrimina muy bien. Una prueba que está constituida por ítemes que discriminan, es una prueba que distinguirá estudiantes con diferentes niveles de aprovechamiento. La fórmula del índice de discriminación es la siguiente:

$$I.D. = \frac{A - B}{N/2} \qquad (20)$$

En esta fórmula, A, B y N tienen el mismo significado que en la fórmula del nivel de dificultad. La fórmula se aplicó a los datos de la tabla 9. Como se observa, hay índices positivos y negativos. El índice de discriminación puede tener valores de −1.00 hasta + 1.00. Un índice de + 1.00 indicaría que todos los estudiantes del grupo alto respondieron correctamente mientras todos los del grupo bajo respondieron incorrectamente. Si el índice es negativo implica que mayor número de estudiantes del grupo bajo respondieron correctamente que el grupo alto. De los 8 ítemes de la tabla 9, únicamente son aceptables para ser usados en una próxima prueba los que tienen un índice mayor de 0.35; éstos son los números 3, 5 y 7. El índice de discriminación es el criterio que más debe pesar en la decisión que toma el docente al seleccionar ítemes para un banco de ítemes.

Nivel de dificultad para ítemes de desarrollo

Para calcular el nivel de dificultad de ítemes que se califican con puntajes diferentes a 0 y 1, c sea ítemes de desarrollo, se debe usar procedimientos diferentes a los anteriores, tales como los propuestos por Whithey y Sabers(1970) y modificados por Scannell y Tracy (1975). Igual que en el procedimiento para el análisis de preguntas objetivas, se escogen altos y bajos en la misma proporción (27%

de cada grupo). Es necesario insistir en el hecho de que la prueba que tenga ítemes de selección y desarrollo debe considerarse en el análisis como dos sub-pruebas. Por lo tanto, los estudiantes que constituyen los grupos alto y bajo, para el análisis hecho a las preguntas objetivas, pueden no ser los mismos que constituyen los grupos alto y bajo al analizar las preguntas de desarrollo.

Como ejemplo, se va a analizar un ítem de desarrollo cuyo puntaje es de 6 puntos. Para esto, se sigue un procedimiento que se resume en la tabla 10.

Primero se observa la frecuencia en que el ítem fue calificado por los diferentes puntajes posibles. Así, en el grupo alto, 4 estudiantes obtuvieron 6 puntos, 3 cinco puntos, 1 cuatro puntos, 2 tres puntos, 1 dos puntos y ninguno obtuvo uno y cero puntos. A continuación, se multiplican los puntajes del ítem por la frecuencia. Los resultados de esta operación se anotan en la columna FX. Esta columna se suma y se denota con S_A. Igual procedimiento se sigue con el grupo bajo.

TABLA No.10

RESUMEN DEL RENDIMIENTO OBTENIDO EN UN ITEM DE DESARROLLO

	GRUPO ALTO			GRUPO BAJO	
Puntaje del Item	F	FX	Puntaje del Item	F	FX
0	0	0	0	5	0
1	0	0	1	3	3
2	1	2	2	1	2
3	2	6	3	1	3
4	1	4	4	0	0
5	3	15	5	0	0
6	4	24	6	0	0
TOTAL	10	51 = S_A		10	8 = S_b

Con los datos obtenidos en la Tabla 10, se aplica la siguiente fórmula:

$$N.D. = \frac{S_A + S_B - N_t\,(X_{min})}{N_t\,(X_{max} - X_{min})} \times 100 \qquad (21)$$

S_A = suma de la columna FX para el grupo alto

S_B = suma de la columna FX para el grupo bajo

X_{max} = puntaje máximo posible del ítem

X_{min} = puntaje mínimo posible del ítem

N_t = número de pruebas en ambos grupos

Para nuestro "ítem", el nivel de dificultad es:

$$N.D = \frac{51 + 8 - \left[(20)\ (0)\right]}{20\ (6 - 0)} \times 100 = 49.17$$

Los niveles de dificultad se interpretan de igual forma que en los de ítemes objetivos.

Indice de discriminación para ítemes de desarrollo

Con los datos de la tabla 10 usados para encontrar el nivel de dificultad, se aplica la fórmula 22, la cual permite calcular el índice de discriminación:

$$I.D. = \frac{S_A - S_B}{N_A\,(X_{max} - X_{min})} \qquad (22)$$

En esa fórmula X_{max}, X_{min}, S_A y S_B tienen igual significado que en la fórmula 21, N_A = número de pruebas en el grupo alto.

El índice de discriminación para el ítem de desarrollo de la tabla 10 es:

$$I.D. = \frac{51 - 8}{10\ (6 - 0)} = 0,72$$

El nivel de dificultad y el índice de discriminación de los ítemes de desarrollo se interpretan como se hace con esos mismos parámetros de los ítemes objetivos.

Conclusiones

Se han presentado en este trabajo técnicas estadísticas para el análisis de los datos obtenidos por la aplicación de instrumentos de medición educativa. En primer lugar, se examinaron las técnicas del cálculo de medidas de tendencia central, de variabilidad y de relación, empleando métodos de cálculo de fácil aplicación para simplificar la labor de maestros y profesores. La bibliografía consultada señala que estos métodos dan como resultado, valores que no difieren más allá de un 3% a un 5%, de aquellos que se obtendrán al aplicar los métodos tradicionales enunciados en los textos estadísticos.

En segundo lugar, se desarrolla el fundamento teórico de la confiabilidad y se ofrecen los métodos más comúnmente usados para su cálculo, haciéndose énfasis en la aplicación apropiada de los distintos índices. Finalmente se revisan las técnicas de análisis de ítemes. En este caso, se separan los métodos de acuerdo con el tipo o clase de ítem, o sea, ítemes objetivos e ítemes de desarrollo.

Se espera que este trabajo indagatorio de análisis y síntesis bibliográfica, sea de utilidad, tanto para los estudiantes universitarios, futuros maestros y profesores, como para aquellos docentes que quieran analizar y mejorar sus pruebas.

BIBLIOGRAFIA

ASHER, J.W. *Educational Research and Evaluation Methods.* Boston: Little, Brown and Co., 1976.

BROWN, F.G. *Principles of Educational and Psychological Testing.* Hindsdale: The Dryden Press, Inc., 1976.

CRONBACH, L.J. Coefficient Alpha and the Internal Structure Of Tests. En N.A. Mehreus y R.L. Ebel. *Principles of Educational and Psychologycal Measurement.* Chicago: Rand McNally and Co., 1967, 123-167.

FERGUSON, G. *Statistical Analysis in Psychology and Education.* New York: McGraw-Hill, 1971.

GAY, L.R. *Educational Evaluation And Measurement.* Columbus: Charles E. Meril Publisking, Co., 1980.

GLASS, G.V. y STANLEY, J.G. *Statistical Methods Education and Psychology.* Englewood Cliffs: Prentice-Hall, Inc., 1970.

HABER, A. y RUNYON, R. P. *Estadística General.* México: Fondo Educativo Interamericano, 1973.

KERLINGER, F.N. *Foundations of Behavioral Research.* New York: Holt, Reinhart and Winston, 1973.

KUDER, F.G. y RICHARSON, M.W. The theory of the Estimation of Test Realibility. En W.A. Mehreus y R.L. Ebel, *Principles of Educational and Psychological Measurement.* Chicago: Rand McNally and Co., 1967, 95-104

LAFOURCADE, P.D. *Evaluación de los Aprendizajes.* Buenos Aires: Editorial Kapelusz, 1973.

MARSO, R.N. The influence of Test Difficulty upon Study Efforts and Achievement. *American Educational Research Journal*, 6 (1969): 621–623.

MASON, G.P. y ODER, R.E. A short – cut Formula for Standard Desviation. *Journal of Educational Measurement 5 (1968): 319 – 320.*

MEHRENS, W.A.y EBEL, R. L.*Principles of Educational and Psychological Measurement.* Chicago: Rand McNally y Co.,1967.

MINIUN, E.W. *Statistical Reasoning in Psychology and Education.* New York: John Weley and Sons, Inc., 1970.

SAUPE, J.L. Some Useful Estimates of The Kuder Richarson Formula Number 20 Realiability Coefficient. *Educational and Psuchological Measurement,* 2 (1961) 63-72.

SCANNEL, D.P. y TRACY, D.B. *Testing and Measurement in the Classron* Boston: Houghton Mifflin Co., 1975.

THORNDIKE, R. y HAGEN, E. *Tests y Técnicas de Medición en Psicología y Educación.* México: Trillas, 1977.

WHITNEY, D.R. y SABERS. D.L. Improving Essay Examinations III. Use of Item Analysus. *Technical Bulletin, II* (IoWa City. University Evaluation and Examination Service, 1970) Mimeografiado.

WOLFE, I.M. A Simple Method of Calculating the coefficient of Correlation. *Journal of Educational Measurement*, 8 (1971): 221-222.