UNIVERSIDAD NACIONAL DE LUJAN
DEPARTAMENTO DE EDUCACION
AREA METODOLOGIA DE LA INVESTIGACION

Ficha N°
Primera revisión 2002
Carlos A. Borsotti

ANÁLISIS DE INFORMACIÓN CUANTIFICADA EN CUADROS DE DOS O MAS VARIABLES

Supongamos que tenemos el siguiente cuadro:

Aprendizajes logrados en matemática por estudiantes de nivel medio según el nivel socioeconómico de la familia de origen y su situación laboral.

Cuadro 1:

Aprendizaje en matemática	Situación laboral del estudiante											
	Trabaja				No trabaja							
	Nivel socioeconómico de la familia de origen			Total	Nivel socioeconómico de la familia de origen			Total	Nivel socioeconómico de la familia de origen			
	Bajo	Medio	Alto		Bajo	Medio	Alto		Bajo	Medio	Alto	Total
0 - 3	50	17	3	70	12	5	2	19	62	22	5	89
4 – 5	40	28	6	74	9	10	6	25	49	38	12	99
6 - 10	10	10	11	31	9	10	22	41	19	20	33	72
Total	100	55	20	175	30	25	30	85	130	80	50	260

Antes de comenzar el análisis, debemos preguntarnos:
¿Cuáles son las variables?
¿Cuáles son los indicadores?
¿Cuál es el nivel de medición original y cuál el utilizado en el cuadro?
¿Cuál es la ubicación cronológica de las variables (antecedente, intermedia, resultado)
¿Cuál será la pregunta a la que se quiso responder con el cuadro?
¿En qué situación problemática pudo haberse originado?
¿Cuál será el objetivo (tipo de conocimiento) que se intentó producir?
¿Cuál será la hipótesis?
¿Cuáles serán los grupos experimentales que se han conformado?
¿Cuál es la situación común en que ha estado la población?
¿Cuál/es la unidad de análisis?
¿Cuál es la dimensión temporal?
¿Cuál/es es/son la/s fuentes de información?
¿Cuál/es es/son la/s unidades informantes?

Después de responder a esas preguntas se está en condiciones de proceder al análisis.

En términos generales, en el texto de interpretación de los cuadros no corresponde repetir los números. En general, los números se saben y se pueden ver. Lo que corresponde es formular una proposición que sintetice lo que dicen los números. Una formulación de las proposiciones que ya muestra una tendencia a la comprensión sería la que dijera, por ejemplo: "uno de cada dos (tres de cada diez, etc.) ... tienen tal propiedad".

Análisis univariado:

En este análisis se procede a determinar la distribución de frecuencias de cada una de las variables.

Conviene comenzar con la variable antecedente o independiente. Tratándose de un cuadro que establece relaciones o incidencias, se trata de un diseño experimental natural o ex post facto. Por lo tanto, cada valor de la variable antecedente indica un grupo experimental. En general, con la variable antecedente, se hace el supuesto de que está todo el universo (en un censo) o que se trata de una muestra al azar, proporcional al universo.

En el caso del cuadro que se está analizando:

Cuadro 2:

Nivel socioeconómico de la familia de origen	Frecuencia	Porcentaje
Bajo	130	50,0
Medio	80	30,8
Alto	50	19,2
Total	260	100,0

¿Cuál es la composición del universo? En el universo, la probabilidad de que un alumno provenga de una familia de nivel socioeconómico bajo es: 50%, 0 .50; medio: 30,8%, o .308; alto: 19,2 % o .192. Pero, de acuerdo con lo antes señalado, cabría formular la proposición como: uno de cada dos estudiantes proviene de una familia con nivel socioeconómico bajo.

A continuación hay que preguntarse: ¿qué significa esta distribución? Y anotar todo lo que se nos ocurra que pueda desprenderse de esa situación.

Luego debe tomarse la variable intermedia, también llamada de prueba (en inglés 'test'). Cada valor de la variable intermedia o interviniente también indica un grupo experimental. Salvo cuando se ha tomado también a esta variable como criterio para seleccionar la muestra, la proporción de quienes trabajan y quienes no lo hacen resulta de la información que se obtenga.

Cuadro 3:

Situación laboral del estudiante	Frecuencia	Porcentaje
Trabaja	175	67.3
No trabaja	85	32,7
Total	260	100,0

Dado el universo de que se trata, la probabilidad de que un estudiante trabaje es de: 67,3% o una probabilidad de .673 y de que no trabaje es de 32,7% o una probabilidad de .327. De nuevo, la proposición sería que dos de cada tres estudiantes trabajan.

A continuación hay que preguntarse: ¿qué significa esta distribución? Y anotar todo lo que se nos ocurra que pueda desprenderse de esa situación.

Por último, se toma la variable resultado o efecto.

Cuadro 4:

Aprendizajes en matemática	(1) Frecuencia	Porcentaje	Porcentaje acumula.	(2) Punto medio	Amplitud inervalo	(1)x(2)	Amplitud inervalo
0 – 3	89	34,2	34,2	2	4	178	4
4 – 5	99	38,1	72,3	5	2	495	2
6 - 10	72	27,7	100,0	8	4	576	4
Total	260	100,0				1.249	
Media 4,7							

De nuevo, la proposición que correspondería formular, sería: menos de uno de cada tres estudiantes consigue promocionar.

Como esta variable está presentada en el cuadro con el nivel de medición intervalar es posible calcular la media y la mediana. Dada la cantidad de valores de la variable 'aprendizajes en matemática' no tiene sentido calcular otros estadígrafos.

Si el nivel socioeconómico y la situación laboral no tuvieran relación con los aprendizajes en matemática, y si la prueba que se les tomó estuvo bien diseñada, todos los estudiantes tendrían la misma probabilidad de obtener cualquier resultado, por lo que la distribución, dado que se trata de tres valores, daría un resultado de 33,3% para cada valor de la variable (lo que es igual a una probabilidad de .333).

Pero los resultados de la prueba no dieron esa igualdad de probabilidades. Si la prueba estuvo bien diseñada, entonces, la situación socioeconómica y/o la situación laboral tienen relación con los aprendizajes en matemática.

A continuación, hay que ver cómo es esa distribución y preguntarse qué significa.

Terminado el análisis univariado se sigue con el

Análisis bivariado.

Consiste en poner en relación, de a dos, todas las variables del cuadro.

Conviene comenzar con la relación entre la variable antecedente y la variable efecto o resultado o consecuencia. Esta sería la relación que en los textoS se lee como 'xy'.

Cuadro 5:

Aprendizaje en matemática	Nivel socioeconómico			Total
	Bajo (x1)	Medio (x2)	Alto (x3)	
0 - 3	62	22	5	89
4 – 5	49	38	12	99
6 – 10	19	20	33	72
Total	130	80	50	260

Se trata de tres grupos experimentales, uno para cada uno de los valores de la variable antecedente. Como es un experimento natural o ex post facto, los grupos están desigualados y hay que preguntarse por la situación común a la que se vieron expuestos. En este caso sería la evaluación mediante la cual se determinan los aprendizajes en matemática.

Los porcentajes se calculan para cada grupo experimental.

Los porcentajes se interpretan comparando los porcentajes de los distintos grupos experimentales para cada valor de la variable resultado o efecto o dependiente.

De nuevo, si el nivel socioeconómico no tuviera ninguna relación con los aprendizajes en matemáticas y la prueba estuviera bien diseñada, todos los grupos hubieran tenido la misma probabilidad de obtener los distintos resultados, a saber:

Cuadro 6:

Aprendizaje en matemática	Nivel socioeconómico			Total
	Bajo (x1)	Medio (x2)	Alto (x3)	
0 - 3	33.3	33.3	33.3	33.3
4 – 5	33.3	33.3	33.3	33.3
6 – 10	33.3	33.3	33.3	33.3
Total	100.0	100.0	100.0	100.0

En ocasiones, se obtiene el porcentaje para cada celda sobre el total de casos, lo que daría la probabilidad compuesta, es decir, cuál es la probabilidad de obtener un puntaje (entre 0 y 3, 4 y 5, 6 y 10) en la prueba de matemática si se proviene de una familia de nivel socioeconómico bajo, o medio, o alto. Hay que tener cuidado con este procedimiento, porque está muy condicionado por los totales de casos para cada valor de las variables antecedentes e interviniente.

Cuadro 7:

Aprendizaje en matemática	Nivel socioeconómico			Total
	Bajo (x1)	Medio (x2)	Alto (x3)	
0 - 3	23,8	8,5	1,9	34,2
4 – 5	18,9	14,6	4,6	38,1
6 – 10	7,3	7,7	12,7	27,7
Total	50,0	30,8	19,2	100,0

Después de ver esto, corresponde volver al análisis y calcular los porcentajes del cuadro 5:

Cuadro 8:

Aprendizaje en matemática	Nivel socioeconómico			Total
	Bajo (x1)	Medio (x2)	Alto (x3)	
0 - 3	47,7 >	27,5 >	10,0	34,2
4 – 5	37,7	< 47,4 >	24,0	38,1
6 – 10	14,6	< 25,1	< 66,0	27,7
Total	100,0	100,0	100,0	100,0

Los signos > y < indican que los porcentajes de una celda son mayores o menores que los de la celda vecina correspondiente al mismo valor de la variable efecto o dependiente. Esta es una ayuda para ver las relaciones.

Esto llevaría a formular una proposición como la siguiente: mientras mayor sea el nivel socioeconómico de la familia de origen, mayores tenderán a ser los aprendizajes de matemática.

Pero esta formulación da cuenta sólo del 51,1% de los casos (cuadro 7: 23,8 + 14,6 + 12,7). Si ha sido obtenido correctamente, la información se refiere a un 100% de los casos y cabe preguntarse: ¿qué pasa con el 49,9%?. Por ejemplo, ¿qué pasa que el 14,6% de los estudiantes provenientes de nivel socioeconómico bajo logra promocionar? ¿qué pasa que el 34% de los estudiantes provenientes de nivel socioeconómico alto no consigue promocionar?. **Es decir, se debe interpretar, formular hipótesis sobre el total de la información obtenida y no sólo sobre aquélla que muestra la tendencia general.**

Un análisis similar al que se acaba de hacer debe realizarse también con las otras dos relaciones que quedan: entre situación laboral y aprendizajes en matemática y entre nivel socioeconómico de la familia de origen y situación laboral.

Cuadro 9:

Aprendiza. en matemát	Absolutos: Situación laboral				Porcentajes: Situación laboral		
	Trabaja	No trabaja	Total		Trabaja	No trabaja	Total
0 – 3	70	19	89		40,0 >	22,4	34,2
4 - 5	74	25	98		42,2 >	29,4	38,1
6 – 10	31	41	72		17,8	< 48,2	27,7
Total	175	85	260		100,0	100,0	100,0

De este cuadro puede formularse la proposición según la cual "de cada diez estudiantes que trabajan apenas tres logran promocionar". **Cabe, de nuevo, preguntarse: ¿qué pasa que el 17,8% de los estudiantes que trabajan pueden promocionar?, ¿qué pasa que el 51,8% de los estudiantes que no trabajan no logran promocionar?**

Falta ver cuál es la relación entre la variable antecedente y la interviniente.

Cuadro 10.1.: En números números absolutos:

Situación laboral	Nivel socioeconómico: Bajo	Medio	Alto	Total
Trabaja	100	55	20	175
No trabaja	30	25	30	85
Total	130	80	50	260

En esta relación entre la variable antecedente y la interviniente, los porcentajes deben calcularse en la dirección de la variable antecedente.

Cuadro 10.2.: En porcentajes:

Situación laboral	Nivel socioeconómico: Bajo	Medio	Alto	Total
Trabaja	76,9>	68,8>	40,0	67,3
No trabaja	23,1	<31,2	<60,0	32,7
Total	100,0	100,0	100,0	100,0

Aquí también cabe preguntarse acerca de qué pasa que el 23,1% de los estudiantes provenientes de familias de nivel socioeconómico bajo no trabaja y acerca de qué pasa que el 40% de los estudiantes provenientes de familias de nivel socioeconómico alto trabaja.

Con esto se terminaría el análisis bivariado y habría que iniciar el

Análisis trivariado:

En el análisis trivariado, se trata de ver si la relación original (entre la variable antecedente o independiente y la variable resultado o dependiente) (ver cuadro 8) se mantiene, se especifica o es espúrea cuanto se introduce otra variable.

La relación original se mantiene si, para cada valor de la variable interviniente, la relación original permanece relativamente igual; se especifica si en cada valor de la variable interviniente, la relación original se refuerza o de debilita; es espúrea si en algún valor de la variable interviniente, se produce una relación inversa a la relación original.

Hay que ver qué sucede con la relación entre el nivel socioeconómico de la familia de origen y los aprendizajes en matemáticas cuando se mantiene constante la situación laboral.

Por ejemplo, **para quienes trabajan**, la situación sería, en números absolutos:

Cuadro 11.1: (para p1; xy)

Aprendizaje en matemática	Nivel socioeconómico Bajo (x1)	Medio (x2)	Alto (x3)	Total
0 - 3	50	17	3	70
4 – 5	40	28	6	74
6 – 10	10	10	11	31
Total	100	55	20	175

y en números relativos, sería

Cuadro 11.2:

Aprendizaje en matemática	Nivel socioeconómico Bajo (x1)	Medio (x2)	Alto (x3)	Total
0 - 3	50,0 >	30, 9 >	15,0	40,0
4 – 5	40,0	< 50,9 >	30,0	42,3
6 – 10	10,0	< 18,1	< 45,0	17,8
Total	100,0	100,0	100,0	100,0

En principio, la relación inicial se mantiene, pero especificada, en cuanto se presenta como más acentuada para los niveles socioeconómicos bajo y medio y más débil para el nivel socioeconómico alto. **¿Cómo se interpreta esto?**

Luego, puede verse lo que sucede con el otro valor de la variable intermedia o interviniente.

Por ejemplo, **para quienes no trabajan**, la situación sería, en números absolutos:

Cuadro 12.1: (para p2; xy)

Aprendizaje en matemática	Nivel socioeconómico Bajo (x1)	Medio (x2)	Alto (x3)	Total
0 - 3	12	5	2	19
4 – 5	9	10	6	25
6 – 10	9	10	22	41
Total	30	25	30	85

y en números relativos, sería

Cuadro 12.2:

Aprendizaje en matemática	Nivel socioeconómico Bajo (x1)	Medio (x2)	Alto (x3)	Total
0 - 3	40,0 >	20,0 >	6,7	22,4
4 – 5	30,0	< 40,0 >	20,0	29,4
6 – 10	30,0	< 40,0	< 73,3	48,2
Total	100,0	100,0	100,0	100,0

En principio, la relación inicial se mantiene, pero especificada, en cuanto se presenta como más acentuada para el nivel socioeconómico algo y más débil para los niveles bajo y medio. **¿Cómo se interpreta esto?**

Pero, lo más importante, ¿qué proposiciones se pueden obtener de todos estos análisis?, ¿qué nuevas vías se pueden indagar?.

Hasta aquí se ha trabajado con porcentajes. Hay otros estradígrafos, más potentes matemáticamente, pero cuyo potencial interpretativo merece ponerse en duda. Tal vez, es más útil sea el chi cuadrado.

El estadígrafo Chi cuadrado.

El Chi cuadrado es un estadígrafo que proporciona una estimación de la independencia estadística entre dos variables. Esto es, permite estimar si, obtenidas muestras similares, cuál es la probabilidad de que se encuentren resultados similares. No corresponde calcular el Chi cuadrado en caso de tratarse de censos. En este caso no se trataría de muestras, sino de la distribución en el universo.

Si se tratara de una muestra en la cual no hubiera ninguna incidencia del nivel socioeconómico de la familia de origen, ni de la situación laboral, ni de algún sesgo en la prueba, una distribución matemáticamente perfecta daría que el número de casos por celda sería, en el caso de la relación entre el nivel socioeconómico de la familia de origen y los aprendizajes en matemática (9 celdas sobre 260 casos), en cada celdas habría que obtener el total de casos dividido por el número de celdas, esto es, algo así como 29 casos por celdas.

Pero entre la matemática y la sociedad, se encuentra la información que se ha obtenido.

Entonces, se tienen los casos observados, esto es, la información obtenida (véanse cuadros 5 y 7). A Fisher y Yates se les ocurrió el estadígrafo Chi cuadrado que tiene en cuenta, además de los valores observados, el cálculo de las frecuencias esperadas.

Las frecuencias esperadas se obtienen a partir de un producto entre el porcentaje correspondiente a una determinada fila por el total de casos que se registran en la columna correspondiente.

Cuadro 13.1:

Aprendizaje en matemática	Nivel socioeconómico de la familia de origen			Total
	Bajo (x1)	Medio (x2)	Alto (x3)	
0 - 3	62	22	5	89 (.342)
4 – 5	49	38	12	99 (.381
6 – 10	19	20	33	72 (27,7)
Total	130	80	50	260

Este cuadro tiene las frecuencias observadas (cuadro 5), con el agregado de los porcentajes de los porcentajes obtenidos en el marginal del total de aprendizajes en matemática (cuadro 7).

En este cuadro, como se procede en general, las celdas se enumeran desde el extremo superior izquierdo avanzando hacia la derecha. De esta manera, se tendrían las siguientes celdas:

Cuadro 13.2:

Aprendizaje en matemática	Nivel socioeconómico de la familia de origen		
	Bajo (x1)	Medio (x2)	Alto (x3)
0 - 3	1	2	3
4 – 5	4	5	6
6 – 10	7	8	9

En las diferentes celdas se tendrían las siguientes frecuencias observadas:

Cuadro 13.3

Celda	Frecuencia observada
1	62
2	22
3	5
4	49
5	38
6	12
7	19
8	20
9	33
Total	260

¿Cuál sería la frecuencia esperada?. Resultara del producto del número absoluto que corresponde al total de la columna donde está la celda por el porcentaje que le corresponde a la fila donde está la celda.

Según los cuadro 5 y 7 se tendrían las siguientes frecuencias esperadas:

Cuadro 13.4

Celda	Total de la columna 'n'	Probabili-da de la fila	Producto
1	130	.342	44,46
2	80	.342	27.36
3	50	.342	17,1
4	130	.381	49,53
5	80	.381	30,48
6	50	.381	19,05
7	130	.277.	36,01
8	80	.277	22.16
9	50	.277	13,85
Total			260

Debe notarse que el total de las frecuencias esperadas es igual al total de las frecuencias observadas.

A partir de ahí, se elevan al cuadrado las frecuencias observadas y se las divide por las frecuencias esperadas. A la sumatoria de ese producto se le resta el total de casos (en el método corto) y se obtiene un valor del Chi cuadrado.

En este caso, se tendría:

Cuadro 13.5:

Celda	(1) Frecuencia observada	Frecuencia observada al cuadrado	Frecuencia esperada	(2) Frecuencia observada al cuadrado dividida por frecuencia esperada		Diferencia (2 - 1)
1	62	3.844	44,46	86,4597		24,4597
2	22	484	27.36	17,6901		- 4,3199
3	5	25	17,1	1,4619		- 3,5381
4	49	2.401	49,53	48,4756		- 0,5244
5	38	1.444	30,48	47,3753		9, 3753
6	12	244	19,05	12,8083		0,8083
7	19	361	36,01	10,0249		- 8,9751
8	20	400	22.16	18,0505		- 1,9495
9	33	1.089	13,85	78,6281		45,6281
Total	260			321,1985		

Se resta el total obtenido (321,1985) al total de casos (260), lo que da un valor del Chi cuadrado de 61.1985.

Para determinar finalmente la probabilidad de obtener resultados similares con muestras similares se recurre a la tabla correspondiente.

Para ello es necesario determinar los grados de libertad (en inglés, degrees of freedom, por eso, en la tabla aparecen como 'df'). Los grados de libertad están determinados por el número de columnas menos uno multiplica por el número de hileras menos uno.

Los grados de libertad se encuentran en la columna de la derecha de la tabla y los porcentajes de probabilidad en la fila de encabezamiento. Se busca la celda correspondiente al menor valor que el encontrado del Chi cuadrado y, volviendo a la fila de encabezamiento, se encuentra la probabilidad de encontrar resultados similares en muestras similares.

En este caso, en un cuadro con 4 grados de libertad (3-1 de nivel socioeconómico de la familia de origen multiplicado por 3-1 de los aprendizajes en matemática) da una probabilidad de que en el 99,9% de los casos en que se tomen muestras similares se obtengan resultados semejantes.

En una de sus publicaciones, María Antonia Gallart sostiene que la magnitud y el signo de la diferencia entre las frecuencias esperadas y las frecuencias observadas, también proporciona elementos para la interpretación de la información contenida en la totalidad del cuadro. Según la última columna de la derecha del cuadro 13.5, podría concluirse que los resultados están claramente vinculados con lo que sucede en las celdas 1 y 9.

ANALISIS DE UN CUADRO CON INFORMACION CUANTIFICADA.

1. La situación problemática que pudo haberlo originado.
2. Las disciplinas en que puede encontrarse conocimiento disponible vinculado con el/los sistema/s de problemas.
3. Las preguntas científicas que se formularon.
4. Los objetivos a que puede responder el cuadro (descripción, explicación, etc.). Si se ha formulado para construir nuevas variables (tipología, índice sumatorio).
5. Las hipótesis que pueden haberlo originado y de qué tipo.
6. Los conceptos que se utilizan y su nivel de medición original.
7. Tipo/s de variable/s (antecedente-consecuente; independiente, de prueba, dependiente).
8. Los indicadores. ¿Reemplazan válidamente a los conceptos?
9. Los niveles de medición utilizados en el cuadro.
10. La unidad de análisis (universo) y el número de casos (censo, muestra).
11. La/s fuente/s de información.
12. Las unidades informantes.
13. Las técnicas y los instrumentos.
14. Posibilidad o necesidad de reducir el cuadro.

Análisis univariado.

1. La distribución de cada variable (información desagregada o agrupada en intervalos, intervalos iguales o desiguales, abiertos o cerrados, amplitud de los intervalos, puntos medios, valores absolutos, relativos, gráficos, estadígrafos).
2. Análisis de los resultados.

Análisis bivariado de un cuadro que responde a un objetivo explicativo o de relaciones.

1. Distinción de los grupos experimentales y de los criterios para constituirlos.
2. Estímulo o situación común a la que están sometidos los grupos.
3. Efecto considerado (variable dependiente).
4. Dirección en que están dispuestos lo valores.
5. Cálculo de los porcentajes.
6. Análisis de los porcentajes calculados.
7. Conclusiones en relación con la hipótesis.
8. Obtención de una medida de síntesis. Dependencia estadística (Chi cuadrado). Asociación según tipo de hipótesis, configuración del cuadro y nivel de medición utilizado.
9. Análisis de los resultados.

Análisis trivariado de un cuadro con objetivo explicativo o relacional.

1. Efectuar el análisis bivariado para las relaciones entre todas las variables (independiente - dependiente; de prueba -dependiente; independiente - de prueba; independiente - dependiente, manteniendo constante cada valor de la variable de prueba): replicación, especificación, espuriedad.
2. Análisis de los resultados.