

A	B	C	D
3	1	1	7
3	1	1	9
3	1	1	8
3	2	1	5
1	1	2	9
1	2	2	6
1	2	2	8
3	1	2	8
3	1	2	7
3	2	2	7
1	1	3	8
1	2	3	5
1	2	3	5
1	2	3	7
3	1	3	5
3	1	3	5
3	1	3	6
3	2	3	4
1	2	4	6
1	2	4	5
2	1	4	7
3	1	4	6
3	1	4	8
3	2	4	5
1	1	5	7
1	2	5	4
2	1	5	8
2	1	5	9
2	2	5	3
3	1	5	9
2	1	6	6
2	2	6	1
3	1	6	8
2	1	7	3
2	1	7	4
2	2	7	4
2	2	7	1
2	1	8	5
2	2	8	5
2	2	9	2

## Investigación en la escuela

En una escuela de primaria acaban de matricularse varios niños y niñas de Guinea Ecuatorial. En el barrio no hay tradición de inmigrantes tan jóvenes y un grupo de estudiantes de psicología decide colaborar en prevenir posibles problemas de rechazo. Para ello, comienzan con un estudio diseñado junto con el AMPA y en el que se pregunta a madres y padres de la escuela algunos asuntos relacionados con la inmigración y la interculturalidad. Cuatro de las preguntas son:

- A. Modelo. ¿Qué es lo que debe hacer un inmigrante que llega aquí? 1. Adquirir la cultura del lugar; 2. Mantener su propia cultura; 3. Mantener y adquirir.
- B. Residencia. ¿En qué parte del barrio vive usted? 1. En la periferia; 2. En el centro.
- C. Incomodidad. De 1 a 9, indique su grado de incomodidad con que los hijos de inmigrantes vengan a nuestra escuela.
- D. Aceptación. De 1 a 9, indique su grado de acuerdo con la afirmación "Me gusta que venga gente de fuera a mi barrio".

Se realizó la entrevista a una muestra aleatoria de 40 madres o padres, obteniendo los siguientes datos. Estudia las relaciones

- A-B: modelo con residencia.
- C-D: incomodidad con aceptación.
- A-C: modelo con incomodidad.

## Soluciones

### Relación A-B. Modelo - Residencia

Se trata de una relación entre dos variables categóricas, por lo que procede poner en marcha una descripción mediante tabla de contingencia, la chi cuadrado de Pearson como estrategia de cuantificación y la V de Cramer como tamaño de efecto.

La tabla de contingencia es:

fo	3	7	12	22
	8	6	4	18
	11	13	16	40

Puede observarse una clara relación a simple vista. En el grupo de personas que vive en la periferia (primera fila), la mayoría opta por la integración, es decir, por que los inmigrantes mantengan su cultura al mismo tiempo que adquieren la del lugar; mientras que la opción menos escogida es la asimilación, es decir, que adquieran la cultura del lugar abandonando la propia. Esta tendencia es la contraria en el grupo de madres y padres que viven en el centro (segunda fila), si bien la diferencia entre las frecuencias esperadas no es tan elevada como en el caso anterior. En general, suelen estar más de acuerdo con la integración (16) que con la asimilación (11), pero esta conclusión "en general", como hemos visto, no es muy fiel a lo que sabemos que ocurre dentro de ambos grupos. En la muestra aleatoria hay más familias de la periferia (22) que del centro (18).

Para el cálculo de la  $\chi^2$  operamos con las frecuencias esperadas:

fe	6,05	7,15	8,8	22
	4,95	5,85	7,2	18
	11	13	16	40

y los componentes de cálculo de la chi cuadrado (residuales al cuadrado, entre la frecuencia esperada):

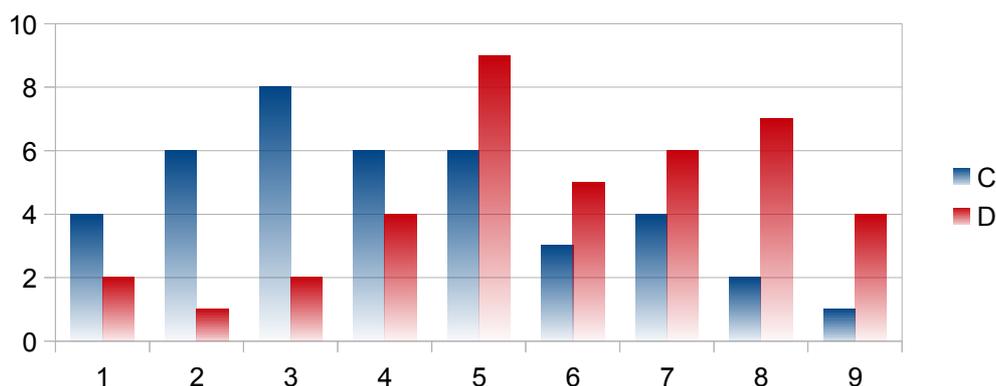
chi	1,54	0,00	1,16	2,70
	1,88	0,00	1,42	3,31
	3,42	0,01	2,59	6,01

$$\chi^2 = \sum \frac{(f_{oi} - f_{ei})^2}{f_{ei}} = 1,54 + 1,88 + \dots = 6,01 \quad v = \sqrt{\frac{\chi^2}{n(k-1)}} = \sqrt{\frac{6,01}{40(2-1)}} = 0,39$$

Como observamos, la chi cuadrado (6,01) se corresponde con un efecto mediano ( $V = 0,39$ ), por lo que se corrobora la interpretación que hemos realizado a partir de la tabla de contingencia con las frecuencias esperadas: existe relación entre la zona de residencia y la actitud frente a la integración: madres y padres de la periferia prefieren la integración, mientras que quienes residen en el centro apuntan a la asimilación.

### Relación C-D. Incomodidad - Aceptación

Tanto C (grado de incomodidad) como D (grado de aceptación) se miden en una escala cuasicuantitativa, por lo que procede estudiarlas mediante un diagrama de dispersión y una correlación lineal simple de Pearson como estrategias de estudio de la relación. No obstante, antes habrá que conocer cada variable por separado.



En la representación gráfica puede observarse una clara dispersión en ambas variables, si bien en el caso C (incomodidad), los datos se agrupan sobre los valores bajos, mientras que en el caso D (aceptación) ocurre lo contrario. En ninguno de ambos casos esta tendencia es muy pronunciada, aunque se vea con claridad.

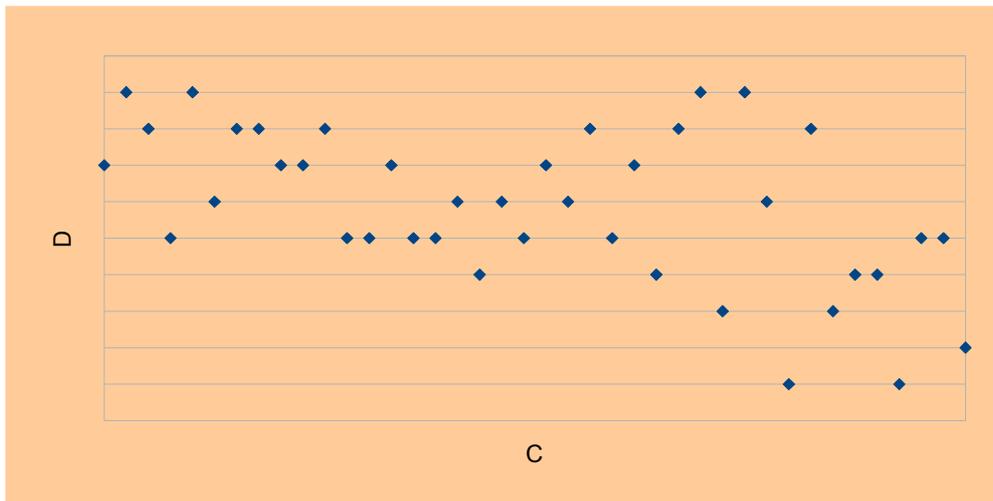
Valor	C			D			Suma	C	D
	f	Cf	C2f	f	Df	D2f			
1	4	4	4	2	2	2			
2	6	12	24	1	2	4			
3	8	24	72	2	6	18			
4	6	24	96	4	16	64			
5	6	30	150	9	45	225			
6	3	18	108	5	30	180			
7	4	28	196	6	42	294			
8	2	16	128	7	56	448			
9	1	9	81	4	36	324			
	40	165	859	40	235	1559			
							Suma	165	235
							S2	859	1559
							n	40	40
							Media	4,13	5,88
							Var.	4,46	4,46
							D.T.	2,11	2,11
							CV	51,19	35,94
							Sa	54	54

La tabla de frecuencias es innecesaria al contar ya con una representación gráfica suficiente para una impresión de la generalidad. No obstante, nos sirve como vehículo intermedio para el cálculo de los índices habituales. Así, observamos que las medias aritméticas se sitúan algo por debajo (C) o por encima (D) del centro del recorrido (5), lo que muestra grados aproximadamente intermedios de incomodidad y aceptación, con la tendencia ya apuntada en la representación gráfica. Las varianzas coinciden en ambos casos, así como la desviación acotada (aproximadamente la mitad de la máxima posible). El coeficiente de variación de Pearson indica que la desviación tipo en ambos casos es apreciable (51% de la media para C, y 36% para D), sin llegar a invalidar su capacidad de representación.

Así pues, en términos generales, sabemos que madres y padres de la escuela cuentan con un grado de incomodidad muy variable, con cierta tendencia a valores bajos, y un grado de aceptación más coincidente, pero también variable, con cierta tendencia a valores altos. Para el estudio de la relación comenzamos con un diagrama de dispersión.

El diagrama de dispersión muestra variabilidad de relación entre ambas variables, si bien con una inclinación negativa, es decir, parece que conforme la incomodidad con la asistencia de pequeños inmigrantes a la escuela es mayor, es menor también la aceptación de personas inmigrantes en el barrio, un resultado que resulta coherente con el sentido común. Al observar

esta tendencia en la representación gráfica, es pertinente seguir con el proceso, cuantificando esta relación mediante el coeficiente de correlación lineal simple de Pearson.



La tabla que sigue permite realizar los cálculos que necesitamos para encontrar el valor de la correlación:

CD	C^2	D^2	CD	C^2	D^2	CD	C^2	D^2
7	1	49	15	9	25	45	25	81
9	1	81	15	9	25	15	25	9
8	1	64	18	9	36	45	25	81
5	1	25	12	9	16	36	36	36
18	4	81	24	16	36	6	36	1
12	4	36	20	16	25	48	36	64
16	4	64	28	16	49	21	49	9
16	4	64	24	16	36	28	49	16
14	4	49	32	16	64	28	49	16
14	4	49	20	16	25	7	49	1
24	9	64	35	25	49	40	64	25
15	9	25	20	25	16	40	64	25
15	9	25	40	25	64	18	81	4
21	9	49				<b>874</b>	<b>859</b>	<b>1559</b>

	C	D	CD
S(X)	165	235	874
S(X <sup>2</sup> )	859	1559	
Media	4,13	5,88	21,85
Var.	4,46	4,46	
D.T.	2,11	2,11	

Aplicando la expresión de cálculo:

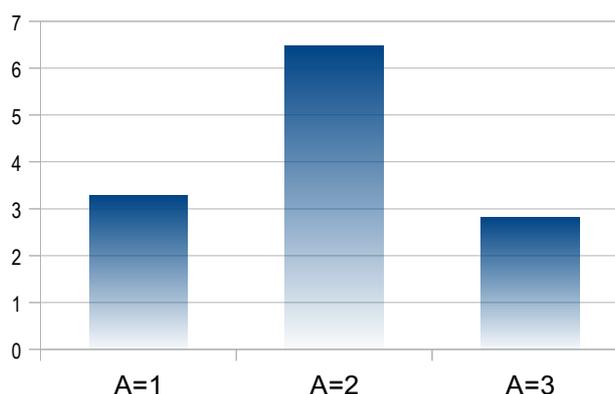
$$r = \frac{\overline{X_A X_B} - \bar{X}_A \bar{X}_B}{S_A S_B} = \frac{21,85 - 4,13 \cdot 5,88}{2,11 \cdot 2,11} = -0,535$$

El valor de la correlación es muy elevado. En términos absolutos, hablamos de un tamaño de efecto grande. La representación gráfica mostraba una inclinación o tendencia clara y negativa pero tal vez no con la apariencia de la cuantía o efecto que hemos encontrado. Ya sabemos que la  $r$  de Pearson es muy sensible a la existencia de relación, de tal forma que nubes de puntos aparentemente muy dispersas y sin ninguna tendencia clara, dan lugar a cuantías que muestran relación lineal.

Así pues, se corrobora la impresión que hemos elaborado a partir de la representación gráfica: conforme es mayor la incomodidad ante la matriculación de niños y niñas inmigrantes, disminuye el grado de aceptación a la presencia inmigrante en el barrio.

### Relación A-C. Modelo - Incomodidad

Este objetivo se inscribe perfectamente en el esquema “relación entre una variable categórica y una cuantitativa”. En ella, la categórica forma grupos de datos cuyas mediciones se corresponden con la variable cuantitativa. El procedimiento implica un diagrama de medias o de caja y patillas combinado como estrategia gráfica inicial, a lo que hay que añadir un análisis de la varianza como procedimiento de cuantificación de la relación. No es necesario que nos impliquemos en estudiar cada variable por separado porque esta tarea está ya realizada en los apartados anteriores.



El gráfico de medias muestra con claridad que la mayor incomodidad con la presencia de niños inmigrantes en la escuela tiene lugar en el grupo de personas que defienden la asimilación, es decir, que las personas inmigrantes han de prescindir de sus costumbres y adquirir las del lugar de recepción.

Para llevar a cabo un análisis de la varianza, es necesario operar con los datos de Incomodidad dentro de cada grupo de Modelo, lo que se lleva a cabo en la siguiente tabla. En ella se ha incluido, para uno de los tres grupos de Modelo, dos columnas: las puntuaciones directas y las distancias cuadráticas a la media de cada grupo. Así, por ejemplo, para el grupo A=1, el valor más bajo es 2 y la media aritmética 3,273. Esto implica una distancia cuadrática de valor

$$(2 - 3,273)^2 = 1,62$$

La suma de cuadrados de este primer grupo es 12,182.

Datos de C para					
A=1		A=2		A=3	
C	d2	C	d2	C	d2
2	1,620	4	6,059	1	3,285
2	1,620	5	2,136	1	3,285
2	1,620	5	2,136	1	3,285
3	0,074	5	2,136	1	3,285
3	0,074	6	0,213	2	0,660
3	0,074	6	0,213	2	0,660
3	0,074	7	0,290	2	0,660
4	0,529	7	0,290	3	0,035
4	0,529	7	0,290	3	0,035
5	2,983	7	0,290	3	0,035
5	2,983	8	2,367	3	0,035
		8	2,367	4	1,410
		9	6,444	4	1,410
				4	1,410
				5	4,785
				6	10,160

Suma	36	12,182	84	25,231	45	34,438	=	71,85
n	11		13		16			
Media	3,273		6,462		2,813			

La suma de las tres sumas de cuadrados de cada grupo se corresponde con la suma de cuadrados *dentro* grupos.

$$SC_D = \sum \sum (X - \bar{X}_{Di})^2 = SC_{D1} + SC_{D2} + SC_{D3} = 12,182 + 25,231 + 34,438 = 71,85$$

Para la suma de cuadrados *entre* grupos, operamos con las medias:

Grupo	Suma	n	Media	SC	d2n
A=1	36	11	3,27	12,18	7,99
A=2	84	13	6,46	25,23	70,97
A=3	45	16	2,81	34,44	27,56
Total	165	40	4,13		106,52

$$SC_E = \sum (\bar{X}_i - \bar{X}_T)^2 n_i = (3,27 - 4,13)^2 \cdot 11 + \dots = 7,99 + 70,97 + \dots = 106,52$$

La suma de cuadrados total se obtiene sumando las dos anteriores, si bien se puede calcular directamente. Ya sabemos de un estudio unitario anterior, que la varianza de C tiene el valor 4,46 (con más precisión: 4,459375). La suma de cuadrados total es el numerador de la varianza, por lo que:

$$S^2 = \frac{\sum (X - \bar{X})^2}{n} = \frac{SC_T}{n} \rightarrow SC_T = S^2 n = 4,46 \cdot 40 = 178,4$$

178,4 es la suma de SC *entre* y SC *dentro*.

De este modo, la tabla completa del análisis de la varianza, incluyendo el cálculo del tamaño de efecto  $\eta^2$ , es como sigue:

FV	SC	gl	V	F	Eta2
Entre	106,52	2	53,26	27,43	0,60
Dentro	71,85	37	1,94		
Total	178,38	39			

El valor de F es sensiblemente elevado, circunstancia que se ve con mucha mayor claridad a través del tamaño de efecto  $\eta^2$ , cuyo resultado es 0,60 (según nuestro baremo, se trata de un efecto grande).

Por tanto, podemos concluir en el mismo sentido que hicimos al trabajar con la gráfica de medias: existe relación entre ambas variables en el contexto de los datos de nuestro estudio.