



La importancia de las Técnicas Cuantitativas. Análisis de tablas de contingencia

Análisis de Tablas de Contingencia

Tabla de contenidos

01 Introducción

02 Tes de homegenidad de varias muestras categoricas

- ✓ Caso general tablas de rxs.
- ✓ Caso particular tablas de 2x2
- ✓ Test de Ji cuadrado (condiciones de validez)
- ✓ Test exacto de Fisher

03 Test de diagnóstico

- ✓ Sensibilidad
- ✓ Especificidad

Introducción



Una tabla de contingencia proporciona una forma de representar los datos que puede facilitar el cálculo de probabilidades. La tabla ayuda a determinar las probabilidades condicionales con bastante facilidad. **La tabla muestra los valores de la muestra en relación con dos variables categóricas diferentes**

Introducción

Cuestiones a tener en cuenta

- ✓ La tabla de contingencia es una técnica de análisis bivariado, relaciona dos variables cualitativas y trata de averiguar si una variable explica la otra.
- ✓ Las tablas de contingencia permiten relacionar variables nominales y ordinales, pero no variables cuantitativas (a no ser que las recodifiquemos por rangos).
- ✓ La lectura de una tabla dependerá del punto de vista de la predicción, de la direccionalidad de la relación que el investigador necesite establecer. En algunos casos esta distinción es más evidente: si relacionamos la categoría profesional con el género, es el hecho de ser varón o mujer lo que determina alcanzar mayores o menores posiciones ocupacionales, y no al revés, que según la ocupación se es varón o mujer.

Test de homogeneidad de varias muestras categóricas

Tablas $r \times s$



Se trata de una tabla de contingencia de doble entrada donde cada una de las dos variables cualitativas implicadas tiene más de 2 categorías

Tablas 2×2



La situación más simple de comparación entre dos variables cualitativas es aquella en la que ambas tienen sólo dos posibles opciones de respuesta (es decir, variables dicotómicas).



- **Tabla de contingencia**: tablas de doble entrada, en cuya entrada se sitúan las diferentes categorías de la variable.
- En las casillas de la tabla se ubica la frecuencia o número de casos.
- Frecuencias:
 - Observadas: número real de casos en nuestro estudio
 - Esperadas: Número de casos que debería haber en la casilla si las categorías fueran independientes.

- Para estimar el grado de asociación entre variables: **chi-cuadrado** (o una razón de verosimilitud).
- Chi-cuadrado parte de la hipótesis de que las variables son independientes (i.e. no existe relación entre ellas y ninguna ejerce influencia sobre la otra).
 - Si $p > 0.05$ se acepta la hipótesis nula (H_0)
- El nivel de significación del χ^2 se calcula mediante 3 métodos:
 - Asintótico → Si la muestra tiene más de 20 sujetos; menos del 20% de las casillas tienen frecuencias esperadas < 5 ; no hay casillas con frecuencia observada < 1 .
 - Monte Carlo → No hay que cumplir los anteriores supuestos. Muestras > 30 sujetos.
 - Exacto: No hay que cumplir los anteriores supuestos. No hay que tener en cuenta el tamaño de la muestra.



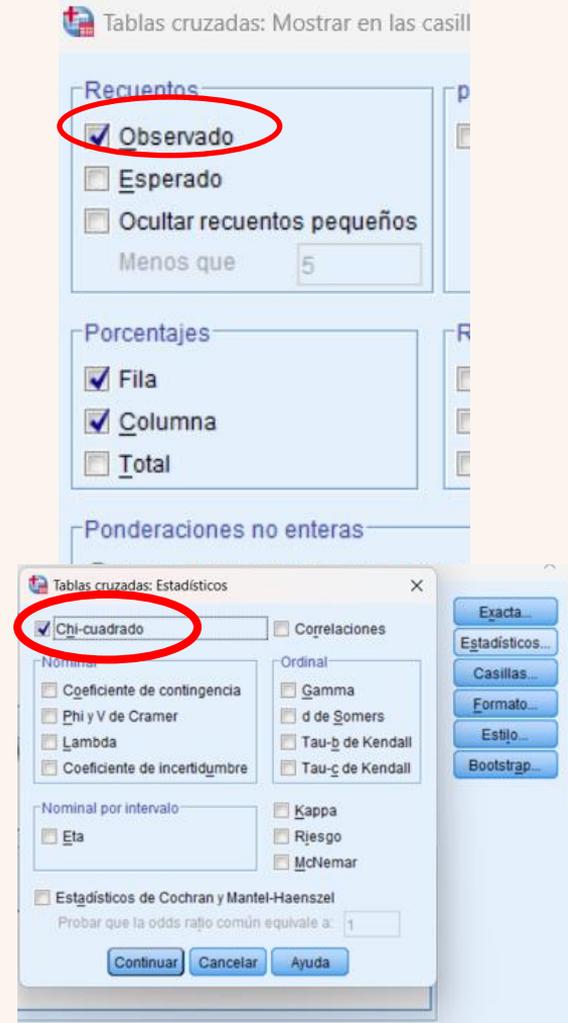
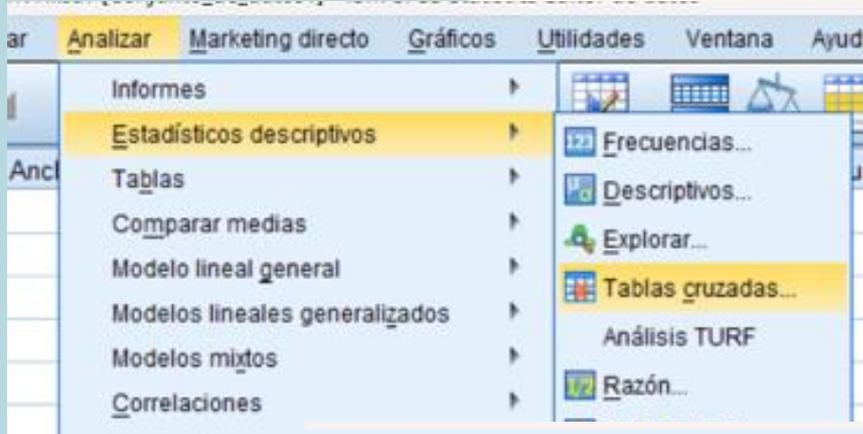
Tablas RxS

Pueden ser:

- r más de dos filas
s dos columnas
- r dos filas
s más de dos columnas
- r más de dos filas
s más de dos columnas

SECUENCIA DE COMANDOS EN SPSS

Analizar → Estadísticos descriptivos → Tabas cruzadas



- ✓ Se utilizará el estadístico de **Chi cuadrado** si todas las frecuencias esperadas son mayores que 5. En la práctica se tolera un máximo del 20% de frecuencias inferiores a 5.
- ✓ Cuando la tabla de contingencia no cumple con las condiciones de validez para utilizar el test de Chi cuadrado, se utilizará el **Test Exacto de Fisher**
- ✓ Por convención, en las filas se representan las categorías de la variable de exposición y en las columnas se representan las categorías de la variable de interés.
- ✓ La interpretabilidad de las Tablas de contingencia puede mejorarse incluyendo porcentajes. Los porcentajes pueden calcularse respecto a la variable **representada en las filas**, respecto a la **variable representada en las columnas**, o **respecto al total**.

Como recomendación, los porcentajes deberían calcularse respecto a la variable de exposición. Por lo tanto, si la variable de exposición se representa en las filas, los porcentajes deberían calcularse por filas.

Tablas rxs

H₀: No existen diferencias significativas entre los porcentajes de las 2 variables a analizar.
No hay relación entre las variables.

H₁: Existen diferencias significativas entre los porcentajes de las dos variables.
Las variables están relacionadas y una explica la otra.

El ejemplo siguiente ha sido tomado del libro BIOESTADISTICA + para las Ciencias de la Salud
A. MARTIN ANDRES
J.D. del CASTILLO

En un estudio sobre tumores cerebrales se desea averiguar si existe asociación entre la localización del tumor y la naturaleza del mismo. A tal efecto se tomaron al azar 141 pacientes afectados de tumor cerebral y se les clasifico según la naturaleza del mismo.

Localización 'Naturaleza del tumor

		Naturaleza del tumor				Total
		T, Benigno	T. Maligno	Otros T		
Localización	Lobulo frontal	Recuento	23	9	6	38
		% dentro de Localización	60,5%	23,7%	15,8%	100,0%
		% dentro de Naturaleza del tumor	29,5%	24,3%	23,1%	27,0%
	% del total	16,3%	6,4%	4,3%	27,0%	
	Lóbulo temporal	Recuento	21	4	3	28
		% dentro de Localización	75,0%	14,3%	10,7%	100,0%
		% dentro de Naturaleza del tumor	26,9%	10,8%	11,5%	19,9%
	% del total	14,9%	2,8%	2,1%	19,9%	
	Otras localizaciones	Recuento	34	24	17	75
% dentro de Localización		45,3%	32,0%	22,7%	100,0%	
% dentro de Naturaleza del tumor		43,6%	64,9%	65,4%	53,2%	
% del total	24,1%	17,0%	12,1%	53,2%		
Total	Recuento	78	37	26	141	
	% dentro de Localización	55,3%	26,2%	18,4%	100,0%	
	% dentro de Naturaleza del tumor	100,0%	100,0%	100,0%	100,0%	
	% del total	55,3%	26,2%	18,4%	100,0%	

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (2 caras)
Chi-cuadrado de Pearson	7,844 ^a	4	,097
Razón de verosimilitud	8,096	4	,088
Asociación lineal por lineal	2,975	1	,085
N de casos válidos	141		

a. 0 casillas (0,0%) han esperado un recuento menor que 5.
El recuento mínimo esperado es 5,16.

Se trata de una tabla de rxs donde $r = 3$ (categorías de localización) y $s = 3$ (categorías de la naturaleza del tumor)

CONCLUSIÓN:

El valor del estadístico Chi Cuadrado es 7,88 con 4 grados de libertad = $(r-1) \times (s-1)$ $p=0,097$ es mayor de $p=0.05$ por tanto no existe asociación entre el tipo de tumor y la localización del mismo.

Aceptamos H_0

Tablas 2x2

Tabla de contingencia general para la comparación de dos variables dicotómicas.

Característica A			
Característica B	Presente	Ausente	Total
Presente	a	b	a + b
Ausente	c	d	c + d
Total	a + c	b + d	n

H₀: No existen diferencias significativas entre los porcentajes de las 2 variables a analizar.
No hay asociación entre las variables.

H₁: Existen diferencias significativas entre los porcentajes de las dos variables.
Las variables están asociadas y una explica la otra.

Se realizó un estudio en 440 mujeres que se dedicaban a la prostitución en una ciudad XX, el interés del estudio fue determinar si existía asociación entre el hecho de consumir drogas por vía parenteral y la infección por VIH

H₀: En la población de mujeres que se dedican a la prostitución en la ciudad XX, no existe una asociación estadísticamente significativa entre el Uso de drogas por vía parenteral y la Infección por VIH.

H₁: En la población de mujeres que se dedican a la prostitución en la ciudad XX existe una asociación estadísticamente significativa entre el Uso de drogas por vía parenteral y la Infección por VIH.

Consumo droga vía parenteral 'Infección VIH

			Infección VIH		Total
			No infectada VIH	Si infectada VIH	
Consumo droga vía parenteral	No	Recuento	385	6	391
		% dentro de Consumo droga vía parenteral	98,5%	1,5%	100,0%
		% dentro de Infección VIH	92,5%	25,0%	88,9%
		% del total	87,5%	1,4%	88,9%
	Si	Recuento	31	18	49
		% dentro de Consumo droga vía parenteral	63,3%	36,7%	100,0%
		% dentro de Infección VIH	7,5%	75,0%	11,1%
		% del total	7,0%	4,1%	11,1%
Total	Recuento	416	24	440	
	% dentro de Consumo droga vía parenteral	94,5%	5,5%	100,0%	
	% dentro de Infección VIH	100,0%	100,0%	100,0%	
	% del total	94,5%	5,5%	100,0%	

La tabla no cumple con las condiciones de validez para aceptar Chi-cuadrado, (1 casilla 25%) tiene un recuento esperado menor de 5, por ello el estadístico correcto a elegir es la **Prueba exacta de Fisher**, con 1 grado de libertad = Consumo de droga 2 categorías e Infección VIH (2 categorías) $gl = (2-1) \times (2-1)$

Consumo Droga Vía Parenteral 'Infección VIH tabulación cruzada

			Infección VIH		Total
			NO Infecta	SI Infecta	
Consumo Droga Vía Parenteral	SI	Recuento	385	6	391
		Recuento esperado	369,7	21,3	391,0
	NO	Recuento	31	18	49
		Recuento esperado	46,3	2,7	49,0
	Total	Recuento	416	24	440
		Recuento esperado	416,0	24,0	440,0

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (2 caras)	Significación exacta (2 caras)
Chi-cuadrado de Pearson	104,619 ^a	1	,000	
Corrección de continuidad	97,905	1	,000	
Razón de verosimilitud	59,816	1	,000	
Prueba exacta de Fisher				,000
Asociación lineal por lineal	104,381	1	,000	
N de casos válidos	440			

a. 1 casillas (25,0%) han esperado un recuento menor que 5. El recuento mínimo esperado es 2,67.

Una vez que hemos detectado la existencia de una asociación entre el uso de drogas por vía parenteral y la infección por VIH, el siguiente paso consiste en cuantificar la magnitud de la asociación que se mide con la Odds ratio (OR)

Los análisis complejos se basan en una medida de asociación conocida como Odds Ratio (OR).

La Odds se define como la proporción de individuos que presentan la característica de interés entre la proporción de individuos que NO presentan la característica de interés:

Consumo droga vía parenteral*Infección VIH tabulación cruzada

Recuento		Infección VIH		Total
		No infectada VIH	Si infectada VIH	
Consumo droga vía parenteral	No	385	6	391
	Si	31	18	49
Total		416	24	440

Estimación de riesgo

	Valor	Intervalo de confianza de 95 %	
		Inferior	Superior
Odds ratio para Consumo droga vía parenteral (No / Si)	37,258	13,792	100,648
Para cohorte Infección VIH = No infectada VIH	1,556	1,257	1,927
Para cohorte Infección VIH = Si infectada VIH	,042	,017	,100
N de casos válidos	440		

Estimación de riesgo

	Valor	Intervalo de confianza de 95 %	
		Inferior	Superior
Odds ratio para Consumo droga vía parenteral (No / Si)	37,258	13,792	100,648
Para cohorte Infección VIH = No infectada VIH	1,556	1,257	1,927
Para cohorte Infección VIH = Si infectada VIH	,042	,017	,100
N de casos válidos	440		

El valor de la OR en las 440 mujeres de la muestra es **37,25** Es decir, en las mujeres que ejercen la prostitución en la ciudad XX y que consumen drogas por vía parenteral tienen una probabilidad (riesgo) de 37,25 veces más de contraer una infección de VIH que las que no consumen drogas por vía parenteral.

Otro ejemplo

Translation missing: es.problem

Supongamos que un estudio sobre infracciones de velocidad y conductores que utilizan teléfonos móviles arroja los siguientes datos ficticios:

	Infracción por exceso de velocidad durante el año anterior	Ninguna infracción por exceso de velocidad durante el año anterior	Total
Utiliza el teléfono móvil mientras conduce	25	280	305
No utiliza el teléfono móvil mientras conduce	45	405	450
Total	70	685	755

Con los datos de la tabla anterior calcule las siguientes probabilidades:

- Calcule $P(\text{El conductor es un usuario de teléfono móvil})$.
- Calcule $P(\text{el conductor no tuvo ninguna infracción durante el año pasado})$.
- Calcule $P(\text{El conductor no tuvo ninguna infracción durante el año pasado Y era usuario de teléfono móvil})$.
- Calcule $P(\text{El conductor es un usuario de teléfono móvil O el conductor no tuvo ninguna infracción durante el año pasado})$.
- Calcule $P(\text{El conductor es un usuario de teléfono móvil DADO que el conductor tuvo una infracción durante el año pasado})$.
- Calcule $P(\text{El conductor no tuvo ninguna infracción el año pasado DADO que el conductor no usaba el teléfono móvil})$.

Solución:

$$\text{a. } \frac{\text{número de usuarios de teléfonos móviles}}{\text{número total en el estudio}} = \frac{305}{755}$$

$$\text{b. } \frac{\text{número que no tenía ninguna infracción}}{\text{número total en el estudio}} = \frac{685}{755}$$

$$\text{c. } \frac{280}{755}$$

$$\text{d. } \left(\frac{305}{755} + \frac{685}{755} \right) - \frac{280}{755} = \frac{710}{755}$$

Solución

e. $\frac{25}{70}$ (El espacio de la muestra se reduce al número de conductores que tuvieron una infracción).

f. $\frac{405}{450}$ (El espacio muestral se reduce al número de conductores que no eran usuarios de teléfonos móviles).

Test de diagnóstico (solo para tablas 2x2)



	Sensibilidad
	Especificidad

Test de diagnóstico (solo para tablas 2x2)

La validez de una prueba diagnóstica: Sensibilidad y especificidad

El caso más sencillo que se nos puede plantear es el de una prueba dicotómica, que clasifica a cada paciente como **sano o enfermo** en función de que el resultado de la prueba sea **positivo o negativo**.

En casos como éste, generalmente un **resultado positivo se asocia con la presencia de enfermedad** y **un resultado negativo con la ausencia de la misma**. Cuando se estudia una muestra de pacientes, los datos obtenidos permiten clasificar a los sujetos en cuatro grupos en una tabla 2x2

Test de diagnóstico (solo para tablas 2x2)

En la tabla 2x2 las filas de dicha tabla contienen el resultado de la prueba diagnóstica y las columnas contienen el estado real de los pacientes o, en su defecto, el resultado de la prueba de referencia o **“gold standard”** que vayamos a utilizar. El resultado de la prueba puede ser correcto (verdadero positivo y verdadero negativo) o incorrecto (falso positivo y falso negativo). El análisis de su validez puede obtenerse calculando los valores de

SENSIBILIDAD Y ESPECIFICIDAD:

Test de diagnóstico (solo para tablas 2x2)

Todas las pruebas diagnósticas son imperfectas. Todas fallan en alguna ocasión, en un sentido y en otro.

- ✓ Cuando realizamos una prueba a un enfermo esperamos obtener un resultado **positivo verdadero (VP)**, pero siempre habrá algunos casos en los que obtendremos negativos, aunque sean enfermos: **son los falsos negativos (FN)**.
- ✓ Y lo mismo ocurre con los sanos, en los que tendremos verdaderos negativos (VN) y falsos positivos (FP). Por eso necesitamos conocer bien las características de las pruebas diagnósticas para apoyar en ellas nuestras decisiones clínicas.
- ✓ **La Sensibilidad (S)** es la probabilidad de que la prueba identifique como enfermo a aquél que efectivamente lo está.
- ✓ **La especificidad (E)** indica la capacidad que tiene la prueba de identificar como sanos (no enfermos) a los que efectivamente lo son.

Tabla de contingencia dos pruebas diagnósticas

Criterio Verdadero (gold estándar)

Prueba Diagnóstica	Enfermos (+)	No enfermos (-) Sanos	Total
Enfermos (+)	VP	FP	VP + FP
No enfermos (-) Sanos	FN	VN	FN + VN
Total	VP + FN	FP + VN	n

$$\text{SENSIBILIDAD} = \frac{VP}{VP+FN}$$

$$\text{ESPECIFICIDAD} = \frac{VN}{VN+FP}$$

Supóngase que se tiene una situación donde se quiere estimar **la sensibilidad y la especificidad de cierto cuestionario XXX** para diagnosticar la presencia de un desorden alimentario en adolescentes.

El cuestionario tiene 5 preguntas que se responden con Sí o No y se considera positivo si la respuesta es Sí en al menos dos de las preguntas.

Se toman 341 mujeres entre 15 y 25 años que acuden a una consulta de Psiquiatría durante 1 año. El equipo investigador les aplica el cuestionario en cuestión. Un año después, en el Servicio se tienen elementos suficientes (sin conocer el resultado de la aplicación del cuestionario XXX) para clasificarlas a todas como **“enfermas con un desorden alimentario de cualquier tipo”** o **“no enfermas de esa dolencia”** y se obtiene la siguiente tabla de 2x2

Tabla de contingencia dos pruebas diagnósticas

Criterio Verdadero (gold estándar)

Prueba Diagnóstica			
Cuestionario XXX	Enfermos	No enfermos (sanos)	Total
Positivos	11	34	45
Negativos	2	294	296
Total	23	128	341

Los resultados del análisis de la tabla anterior se han realizado utilizando el programa informático estadístico: Epidat 4.2

<https://www.google.com/search?q=epidat+descargar&oq=epidat&aqs=chrome.3.69i57j0i512l9.5646j0j15&sourceid=chrome&ie=UTF-8>

Pruebas diagnósticas simples			
Nivel de confianza:		95,0%	
Prueba diagnóstica	Prueba de referencia		Total
	Enfermos	Sanos	
Positivo	11	34	45
Negativo	2	294	296
Total	13	328	341
	Valor	IC (95%)	
Sensibilidad (%)	84,62	61,16	100,00
Especificidad (%)	89,63	86,18	93,09
Indice de Validez (%)	89,44	86,03	92,85
Valor predictivo + (%)	24,44	10,78	38,11
Valor predictivo - (%)	99,32	98,22	100,00
Prevalencia (%)	3,81	1,63	5,99
Indice de Youden	0,74	0,54	0,94
Razón de verosimilitud +	8,16	5,51	12,10

Pruebas diagnósticas simples

Nivel de confianza: 95,0%

Prueba diagnóstica	Prueba de referencia		Total
	Enfermos	Sanos	
Positivo	11	34	45
Negativo	2	294	296
Total	13	328	341

	Valor	IC (95%)	
Sensibilidad (%)	84,62	61,16	100,00
Especificidad (%)	89,63	86,18	93,09
Indice de Validez (%)	89,44	86,03	92,85
Valor predictivo + (%)	24,44	10,78	38,11
Valor predictivo - (%)	99,32	98,22	100,00
Prevalencia (%)	3,81	1,63	5,99
Indice de Youden	0,74	0,54	0,94
Razón de verosimilitud +	8,16	5,51	12,10

Sensibilidad:

La prueba diagnóstica (nuevo cuestionario a probar) clasifica correctamente al 84,62% de las mujeres enfermas. Es decir, siendo clasificadas como enfermas realmente son enfermas

$$(11/13) \times 100 = 84,62\%$$

Especificidad

La prueba diagnóstica (prueba nueva a probar) clasifica correctamente al 89,63% de las pacientes sanas. Es decir, siendo clasificadas como sanas, realmente son sanas.

$$(294/328) \times 100 = 89,63\%$$



GRACIAS

Delfín rosado. Amazonia de Colombia