

# **TEST *Chi-Cuadrado***

## **ASOCIACIÓN DE VARIABLES CUALITATIVAS O CATEGÓRICAS**

***Carmen M. Cadarso-Suárez***

[eicadar@usc.es](mailto:eicadar@usc.es)

En ocasiones, el investigador está interesado en el estudio de una posible

### Asociación entre dos variables categóricas $A$ y $B$

$$A = \{A_1, A_2, \dots, A_i, \dots, A_r\} \quad \text{y} \quad B = \{B_1, B_2, \dots, B_j, \dots, B_s\}$$

$A$ , presentando  $r$  niveles de respuesta, y  $B$  con  $s$  niveles de respuesta.

### EJEMPLOS:

1. Asociación entre Sobrepeso (si,no) e Hipertensión (si,no).
2. Bajo peso del niño al nacer ( $\leq 2500$  grs,  $> 2500$  grs) y edad de la madre ( $< 40$  años,  $\geq 40$  años).
3. Tipo de patología (inflamatoria, neoplasia, traumática, otras) y tipo de operación (limpia, potencialmente contaminada, contaminada, sucia).
4. Estado del paciente aquejado de artrosis (peor, igual, mejor) y tipo de tratamiento recibido ( T1, T2).....

En la práctica, contamos con una tabla de contingencia  $r \times s$

		B			
		$B_1$	$B_2$	$B_j$	$B_s$
A	$A_1$	$n_{11}$	$n_{12}$	$n_{1j}$	$n_{1s}$
	$A_2$	$n_{21}$	$n_{22}$	$n_{2j}$	$n_{2s}$
	$A_i$	$n_{i1}$	$n_{i2}$	$n_{ij}$	$n_{is}$
	$A_r$	$n_{r1}$	$n_{r2}$	$n_{rj}$	$n_{rs}$
		$n_{.1}$	$n_{.2}$	$n_{.j}$	$n_{.s}$

- En un estudio de Infecciones post-operatorias, interesa estudiar la asociación entre patología y tipo de operación:

Recuento		TIPO DE OPERACIÓN				
		limpia	pot_cont	contam	sucia	
PATOLOGÍA	inflam	142	352	213	347	1054
	neo	186	268	39	25	518
	traum	21	15	13	16	65
	otras	526	118	29	43	716
		875	753	294	431	2353

Para una mejor interpretación, damos los porcentajes por filas o por columnas:

Tabla de contingencia PATOL \* TIP\_OPER

			TIP OPER				Total
			limpia	pot cont	contam	sucia	
PATOL	inflam	Recuento	142	352	213	347	1054
		% de TIP_OPER	16.2%	46.7%	72.4%	80.5%	44.8%
	neo	Recuento	186	268	39	25	518
		% de TIP_OPER	21.3%	35.6%	13.3%	5.8%	22.0%
	traum	Recuento	21	15	13	16	65
		% de TIP_OPER	2.4%	2.0%	4.4%	3.7%	2.8%
	otras	Recuento	526	118	29	43	716
		% de TIP_OPER	60.1%	15.7%	9.9%	10.0%	30.4%
Total		Recuento	875	753	294	431	2353
		% de TIP_OPER	100.0%	100.0%	100.0%	100.0%	100.0%

- Resultados de una biopsia hepática, ¿se distribuyen por igual en ambos sexos?

Tabla de contingencia SEXO \* Biopsia Hepática

		Biopsia Hepática					
		Esteatosis	Fibrosis	Hepatitis	Cirrosis	Total	
SEXO	mujer	Recuento	3	3	23	15	44
		% de SEXO	6.8%	6.8%	52.3%	34.1%	100.0%
	hombre	Recuento	45	24	47	24	140
		% de SEXO	32.1%	17.1%	33.6%	17.1%	100.0%
Total		Recuento	48	27	70	39	184
		% de SEXO	26.1%	14.7%	38.0%	21.2%	100.0%

- Fumar durante el embarazo ¿es un factor de riesgo para el bajo peso niño al nacer?

madre fumadora y bajo peso del niño al nacer

			PESO_N		
			>=2500 grs	<2500 grs	Total
FUMADORA	no	Recuento	86	29	115
		% de FUMADORA	74.8%	25.2%	100.0%
	si	Recuento	44	30	74
		% de FUMADORA	59.5%	40.5%	100.0%
Total		Recuento	130	59	189
		% de FUMADORA	68.8%	31.2%	100.0%

## CONTRASTE DE ASOCIACIÓN $\chi^2$

Planteamos las siguientes hipótesis de trabajo:

$$\left\{ \begin{array}{ll} \mathbf{H_0 : A \text{ y B son INDEPENDIENTES} \quad (\text{no asociación})} \\ \mathbf{H_1 : A \text{ y B son DEPENDIENTES} \quad (\text{asociación})} \end{array} \right.$$

1. Partimos de un total de  $n$  individuos.
2. La tabla de frecuencias **OBSERVADAS** es:

	$B_1$	$B_2$	$B_j$	$B_s$	
$A_1$	$n_{11}$	$n_{12}$	$n_{1j}$	$n_{1s}$	$n_{1.}$
$A_2$	$n_{21}$	$n_{22}$	$n_{2j}$	$n_{2s}$	$n_{2.}$
$A_i$	$n_{i1}$	$n_{i2}$	$n_{ij}$	$n_{is}$	$n_{i.}$
$A_r$	$n_{r1}$	$n_{r2}$	$n_{rj}$	$n_{rs}$	$n_{r.}$
	$n_{.1}$	$n_{.2}$	$n_{.j}$	$n_{.s}$	$n$

3. Bajo la hipótesis nula de NO ASOCIACIÓN se tendría

$$P(A_i \cap B_j) = P(A_i) \times P(B_j) \quad \forall i, j$$

y entonces la tabla de frecuencias ESPERADAS bajo  $H_0$  sería :

	$B_1$	$B_2$	$B_j$	$B_s$	
$A_1$	$e_{11}$	$e_{12}$	$e_{1j}$	$e_{1s}$	$n_{1.}$
$A_2$	$e_{21}$	$e_{22}$	$e_{2j}$	$e_{2s}$	$n_{2.}$
$A_i$	$e_{i1}$	$e_{i2}$	$e_{ij}$	$e_{is}$	$n_{i.}$
$A_r$	$e_{r1}$	$e_{r2}$	$e_{rj}$	$e_{rs}$	$n_{r.}$
	$n_{.1}$	$n_{.2}$	$n_{.j}$	$n_{.s}$	$n$

donde

$$e_{ij} = (n_{i.} \times n_{.j})/n$$

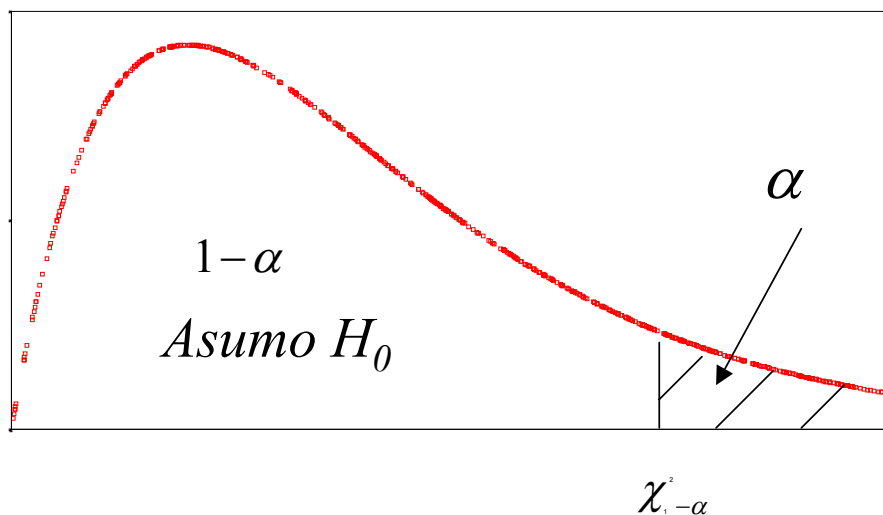
4. Construimos el estadístico “distancia” entre la tabla de frecuencias observadas y frecuencias esperadas:

$$\chi^2 = \frac{(n_{11} - e_{11})^2}{e_{11}} + \frac{(n_{12} - e_{12})^2}{e_{12}} + \dots + \frac{(n_{rs} - e_{rs})^2}{e_{rs}} = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$$

el cual, bajo la hipótesis nula sigue una distribución  $\chi^2$

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \rightarrow \chi_{(r-1)(s-1)}^2$$

5. Fijar un nivel de significación  $\alpha$  ( p.e. , 0.05), y construir el “valor crítico”  $\chi_{1-\alpha}^2$ .



6. Regla de decisión:

- Si  $\chi^2 > \chi_{1-\alpha}^2$  rechazamos  $H_0$  y existe asociación significativa ( $p\text{-value} < \alpha$ )
- Si  $\chi^2 \leq \chi_{1-\alpha}^2$  asumimos  $H_0$  ( $p\text{-value} \geq \alpha$ ). El test es no significativo (n.s.)

$$p\text{-value} = p(\text{rechazar } H_0 \text{ con nuestra muestra particular}) = P(\chi_{(r-1)(s-1)}^2 > \chi^2).$$

**Ejemplo:** ¿Existe asociación significativa entre presencia de hepatomegalia y enfermedad hepática (según resultado de la biopsia)?

Tabla de contingencia hepatomegalia \* Biopsia Hepática

			Biopsia Hepática				Total
			Esteatosis	Fibrosis	Hepatitis	Cirrosis	
hepatomegalia	no	Recuento	12	7	6	4	29
		% de Biopsia Hepática	25.0%	25.9%	8.6%	10.3%	15.8%
	sí	Recuento	36	20	64	35	155
		% de Biopsia Hepática	75.0%	74.1%	91.4%	89.7%	84.2%
Total	Recuento		48	27	70	39	184
	% de Biopsia Hepática		100.0%	100.0%	100.0%	100.0%	100.0%

Aplicamos el test chi-cuadrado y la salida del SPSS es la siguiente:

Tabla de contingencia hepatomegalia \* Biopsia Hepática

			Biopsia Hepática				Total
			Esteatosis	Fibrosis	Hepatitis	Cirrosis	
hepatomegalia	no	Recuento	12	7	6	4	29
		Frecuencia esperada	7.6	4.3	11.0	6.1	29.0
	sí	Recuento	36	20	64	35	155
		Frecuencia esperada	40.4	22.7	59.0	32.9	155.0
Total	Recuento		48	27	70	39	184
	Frecuencia esperada		48.0	27.0	70.0	39.0	184.0

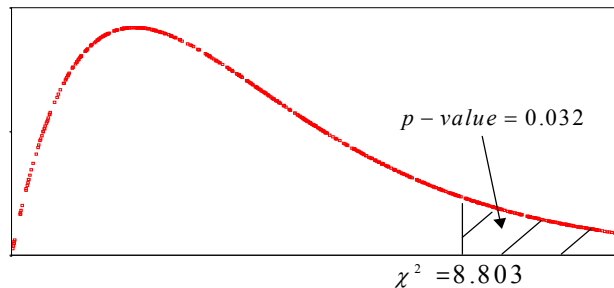
Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	8.803 <sup>a</sup>	3	.032
N de casos válidos	184		

a. 1 casillas (12.5%) tienen una frecuencia esperada inferior a 5.  
La frecuencia mínima esperada es 4.3.

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - e_{ij})^2}{e_{ij}} = \frac{(12 - 7.6)^2}{7.6} + \dots + \frac{(35 - 32.9)^2}{32.9} = 8.803$$

Fijado  $\alpha=0.05$ , el valor crítico es  $\chi^2_{1-\alpha} = \chi^2_{0.95} = 7.81$



Como el  $p\text{-value} = p(\chi^2_3 > 8.803) = 0.032 < 0.05$

“Existe asociación significativa entre presencia de hepatomegalia y tipo de enfermedad hepática ( $p=0.032$ , o también,  $p<0.05$ ) “

## Ejemplo (estudio en pacientes VIH/SIDA):

¿Existe asociación significativa entre la edad del paciente y sus niveles de linfocitos CD4?

Tabla de contingencia CD4\_CO \* EDAD\_CO

		EDAD_CO			Total
		<=30 años	31-40 años	>=41 años	
CD4_CO <200	Recuento	6	30	6	42
	% de EDAD_CO	13.3%	19.9%	15.4%	17.9%
200-500	Recuento	20	72	21	113
	% de EDAD_CO	44.4%	47.7%	53.8%	48.1%
>500	Recuento	19	49	12	80
	% de EDAD_CO	42.2%	32.5%	30.8%	34.0%
Total	Recuento	45	151	39	235
	% de EDAD_CO	100.0%	100.0%	100.0%	100.0%

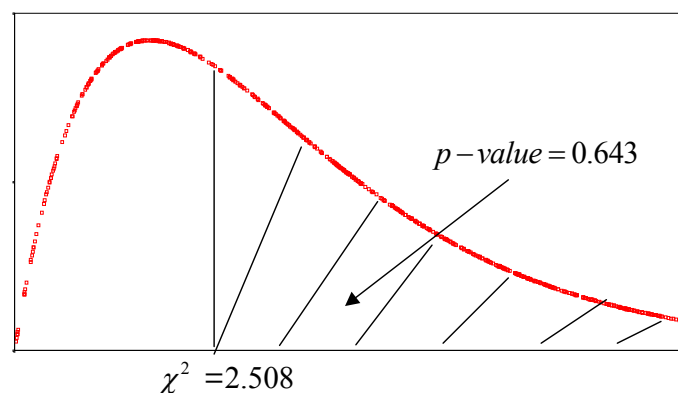
Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	2.508 <sup>a</sup>	4	.643
N de casos válidos	235		

a.

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - e_{ij})^2}{e_{ij}} = 2.508$$

Fijado  $\alpha=0.05$ , el valor crítico es  $\chi^2_{1-\alpha} = \chi^2_{0.95} = 9.49$



Como el  $p\text{-value} = p(\chi^2_4 > 2.508) = 0.643 > 0.05$

“No existe asociación significativa entre edad del paciente y niveles de linfocitos CD4 ( $p=0.643$ , o simplemente, n.s.) “

### CORRECCIÓN POR CONTINUIDAD (Yates)

En aquellos casos en que las frecuencias esperadas  $e_{ij}$  son  $<5$  se recomienda realizar una corrección *por continuidad* del estadístico

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \rightarrow \chi^2_{(r-1)(s-1)}$$

### Corrección de Yates

$$\chi_c^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(|n_{ij} - e_{ij}| - 0.5)^2}{e_{ij}} \rightarrow \chi^2_{(r-1)(s-1)}$$



## INFEC.SAV

### Objetivo del estudio:

“Identificar los factores pronóstico de infección post-quirúrgica”

**EDAD** (años)

**SEXO** 1=varón ; 2=mujer

**PATOL** (Patología) 1=inflamatoria; 2=neoplasia;3=trauma; 4=otras.

**TIP\_OPER** (tipo operación)

1=limpia; 2=potencialente contaminada; 3=contaminada; 4=sucia

**ALB** (albúmina)

**HB** (Hemoglobina)

**HCTO** (Hematocrito)

**LEUCOS** (Leucocitos)

**LINFOPCT** (Linfocitos (%))

**HEMAT** (Hematíes)

**GLUC** (Glucosa)

**OBES** (Obesidad) 1=si ; 2=no.

**DESNUTR** (Desnutrición) 1=si; 2=no.

**DIABETES** 1=si; 2=no.



**INFEC** = Infección post-quirúrgica 0=no; 1=si

## EJEMPLO (infec.sav)

Comprobar si existe asociación entre infección post-quirúrgica (**infec**) y tipo de operación (**tip\_oper**)

→ **Analizar...**

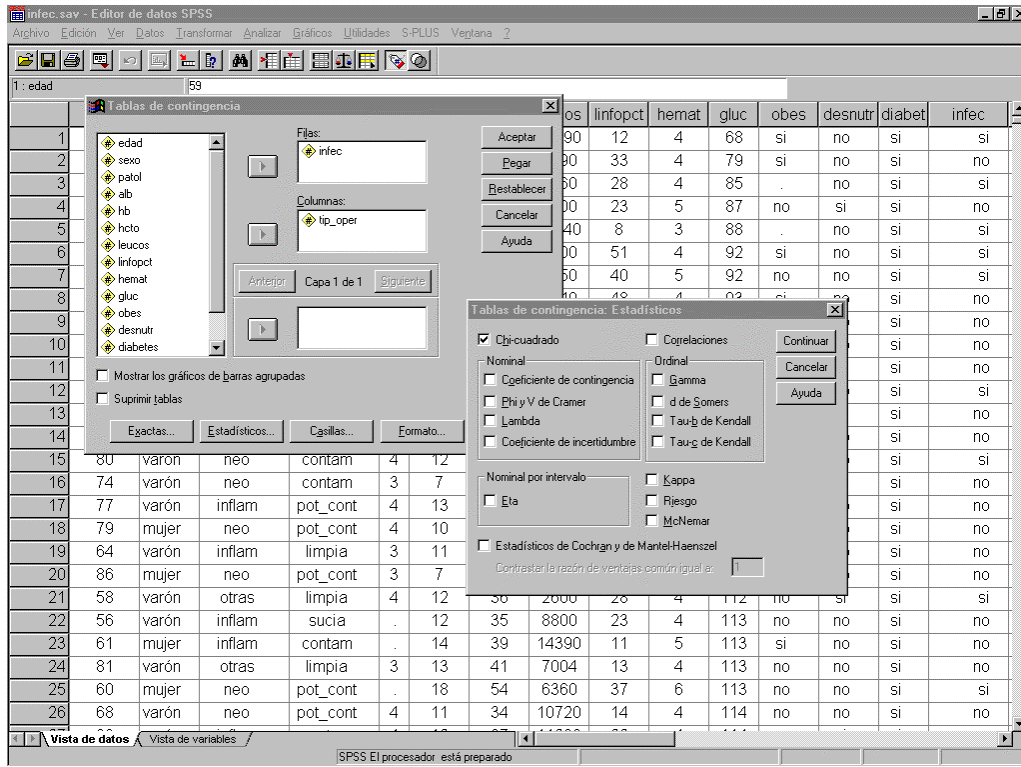
→ **Estadísticos descriptivos.....**

→ **Tablas de contingencia.....**

→ **Filas= infec , Columnas= tip\_oper**

→ **Casillas.....columna** (así tenemos el porcentaje de infección por cada tipo de operación )

→ **Estadísticos.....Chi-cuadrado**



**Tabla de contingencia INFE \* TIP\_OPER**

			TIP_OPER				
			limpia	pot_cont	contam	sucia	Total
INFE	no	Recuento	824	607	211	247	1889
		% de TIP_OPER	94.2%	80.6%	71.8%	57.3%	80.3%
	si	Recuento	51	146	83	184	464
		% de TIP_OPER	5.8%	19.4%	28.2%	42.7%	19.7%
Total		Recuento	875	753	294	431	2353
		% de TIP_OPER	100.0%	100.0%	100.0%	100.0%	100.0%

**Pruebas de chi-cuadrado**

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	263.828 <sup>a</sup>	3	.000
Razón de verosimilitud	268.727	3	.000
Asociación lineal por lineal	262.198	1	.000
N de casos válidos	2353		

a. 0 casillas (.0%) tienen una frecuencia esperada inferior a 5.  
La frecuencia mínima esperada es 57.98.

# REPRESENTACIÓN GRÁFICA

→ Gráficos...

→ Diagramas de barras .....

→ agrupados....

→ Eje de categorías infec

→ Filas= infec , Columnas= tip\_oper

→ Definir grupos por .....tip\_oper

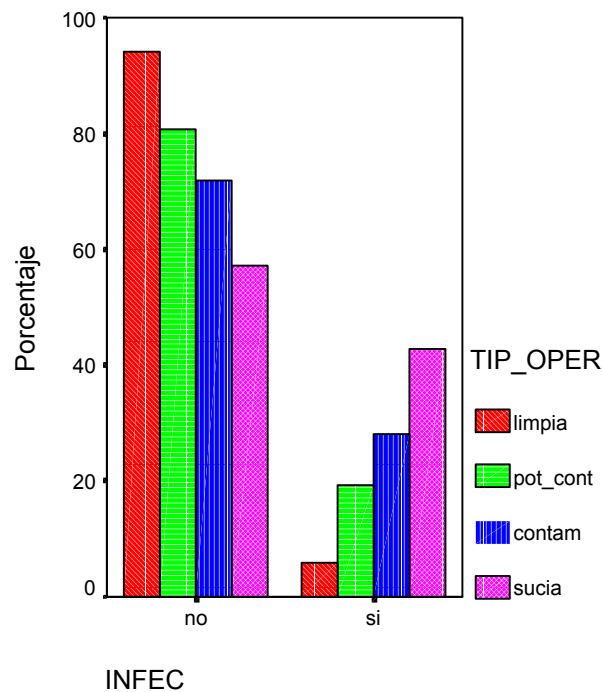
infec.sav - Editor de datos SPSS

Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades SPSS PLUS Ventana ?

2: hcto 36

	edad	sexo	patol	tip_oper	alb	hb	hcto	leucos	linfopct	hemat	gluc	obes	desnutr	diabet	infec
1	59	mujer	inflam	sucia	4	13	38	19090	12	4	68	si	no	si	si
2	65	mujer	otras	pot_cont	4	12	36	6190	33	4	79	si	no	si	no
3	69	varón	neo	contam	4	13	38	6360	28	4	85	.	no	si	si
4	70	mujer	neo	contam	.	14	39	8200	23	5	87	no	si	si	no
5	79	mujer	inflam	contam	.	8	23	16940	8	3	88	.	no	si	si
6	55	mujer	otras	pot_cont	4	12	.	.	.	.	.	.	.	.	si
7	56	mujer	traum	contam	.	13	.	.	.	.	.	.	.	.	si
8	65	mujer	otras	pot_cont	3	11	.	.	.	.	.	.	.	.	no
9	70	mujer	neo	pot_cont	4	8	.	.	.	.	.	.	.	.	no
10	80	mujer	otras	pot_cont	3	12	.	.	.	.	.	.	.	.	no
11	53	mujer	neo	pot_cont	4	10	.	.	.	.	.	.	.	.	no
12	77	mujer	neo	pot_cont	3	12	.	.	.	.	.	.	.	.	si
13	72	mujer	inflam	limpia	4	14	.	.	.	.	.	.	.	.	no
14	85	mujer	inflam	pot_cont	3	12	.	.	.	.	.	.	.	.	no
15	80	varón	neo	contam	4	12	.	.	.	.	.	.	.	.	si
16	74	varón	neo	contam	3	7	.	.	.	.	.	.	.	.	no
17	77	varón	inflam	pot_cont	4	13	.	.	.	.	.	.	.	.	no
18	79	mujer	neo	pot_cont	4	10	.	.	.	.	.	.	.	.	no
19	64	varón	inflam	limpia	3	11	.	.	.	.	.	.	.	.	no
20	86	mujer	neo	pot_cont	3	7	25	13590	13	4	112	si	no	si	no
21	58	varón	otras	limpia	4	12	36	2600	28	4	112	no	si	si	si
22	56	varón	inflam	sucia	.	12	35	8800	23	4	113	no	no	si	no
23	61	mujer	inflam	contam	.	14	39	14390	11	5	113	si	no	si	no
24	81	varón	otras	limpia	3	13	41	7004	13	4	113	no	no	si	no
25	60	mujer	neo	pot_cont	.	18	54	6360	37	6	113	no	no	si	si
26	68	varón	neo	pot_cont	4	11	34	10720	14	4	114	no	no	si	no

SPSS El procesador está preparado



## EJEMPLO (*infec.sav*)

Comprobar si existe asociación entre infección post-quirúrgica (**infec**) y tipo de patología (**patol**):

Tabla de contingencia INFEC \* PATOL

			PATOL				
			inflam	neo	traum	otras	Total
INFEC	no	Recuento	829	379	45	636	1889
		% de PATOL	78.7%	73.2%	69.2%	88.8%	80.3%
	si	Recuento	225	139	20	80	464
		% de PATOL	21.3%	26.8%	30.8%	11.2%	19.7%
Total		Recuento	1054	518	65	716	2353
		% de PATOL	100.0%	100.0%	100.0%	100.0%	100.0%

## EJEMPLO (*infec.sav*)

Comprobar si existe asociación entre infección post-quirúrgica (**infec**) y edad del paciente (Nota: Trabajos previos indican que las personas con  $\geq 65$  años tienen más riesgo de infección ....)

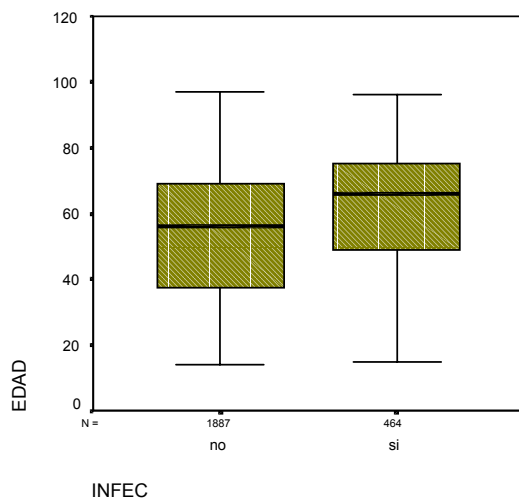
La variable **infec** es **binaria** y la edad es **continua**.

Podemos actuar de dos formas:

1) Utilizar el test de **Mann-Whitney**:

Informe

EDAD					
INFEC	Media	Desv. típ.	Mínimo	Máximo	N
no	53.01	20.039	14	97	1887
si	61.21	18.509	15	96	464
Total	54.63	20.011	14	97	2351



Estadísticos de contraste<sup>a</sup>

EDAD	
U de Mann-Whitney	332239.5
W de Wilcoxon	2113568
Z	-8.058
Sig. asintót. (bilateral)	.000

a. Variable de agrupación: INFEC

con lo que,

*“La infección post-quirúrgica está significativamente asociada a la edad ( $p < 0.001$ )”.*

2) Categorizar edad en **edad\_co**: “ $\geq 65$  años”, “ $< 65$  años”,

**Edad\_co** = 1 si edad  $\geq 65$ .

**Edad\_co** = 2 si edad  $< 65$ .

y contrastar la posible asociación entre **infec** y **edad\_co** con el test  $\chi^2$  .....

## ¿Cómo categorizar variables en el SPSS?

→ **Transformar..**

→ **Recodificar.....**

→ **en distintas variables....**

→ **Variable numérica** edad

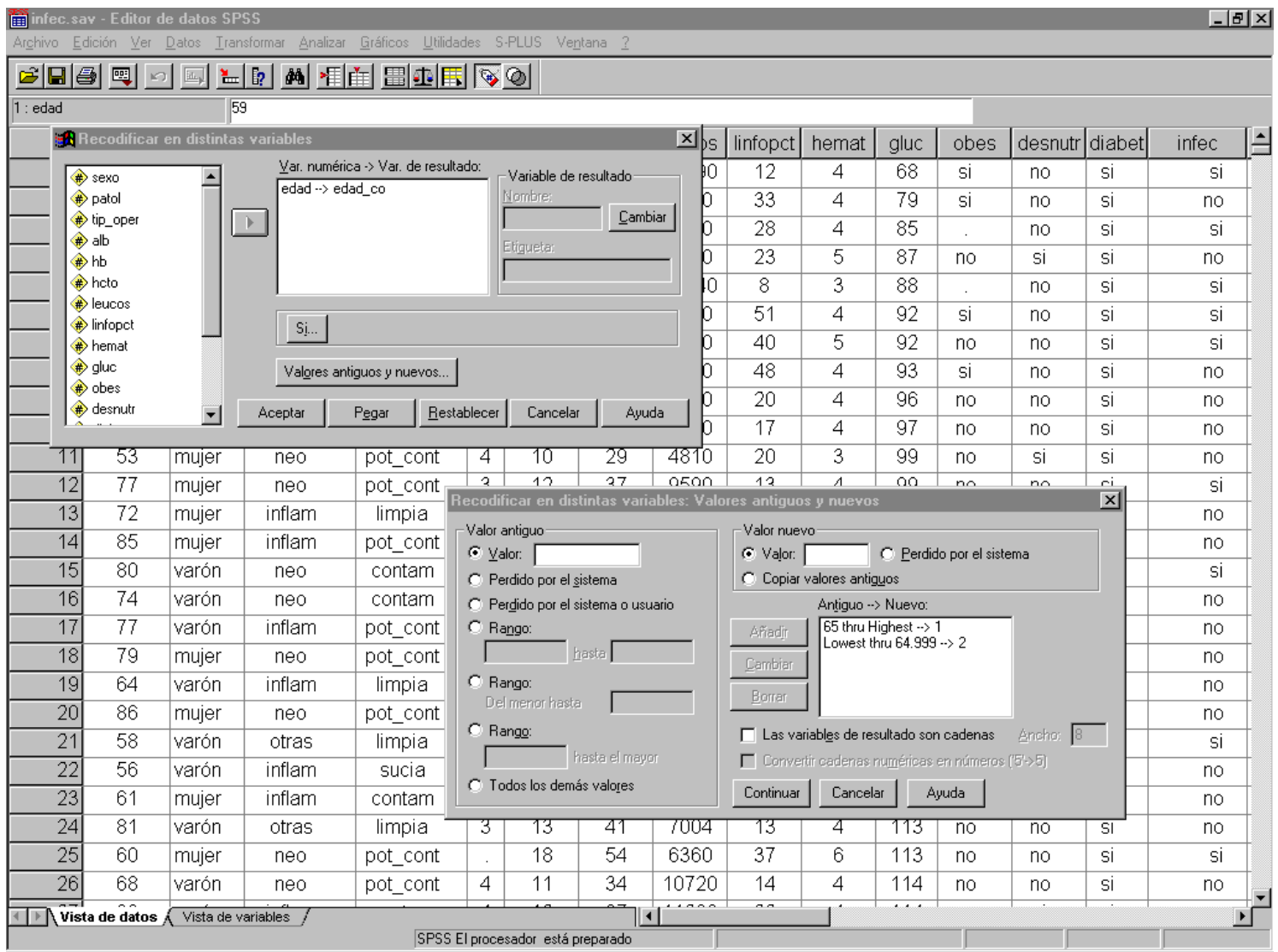
→ **Variable de resultado** edad\_co

→ **Valores antiguos y nuevos .....**

**Antiguo → nuevo**

**65 thru Highest → 1**

**Lowest thru 64.999 → 1**



Entonces se crea la variable **edad\_co** que el SPSS incorpora automáticamente en la base de datos.....

Antes de continuar, podemos ponerle etiquetas (labels) a los valores de **edad\_co**

infec.sav - Editor de datos SPSS

Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades S-PLUS Ventana ?

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida
1	edad	Numérico	5	0		Ninguno	Ninguno	7	Centrado	Escala
2	sexo	Numérico	2	0		{1, varón}...	Ninguno	6	Centrado	Ordinal
3	patol	Numérico	11	0		{1, inflam}...	Ninguno	9	Centrado	Ordinal
4	tip_oper	Numérico	2	0		{1, limpió}	Ninguno	0	Centrado	Ordinal
5	alb	Numérico	7	0					Centrado	Ordinal
6	hb	Numérico	7	0					Centrado	Ordinal
7	hcto	Numérico	7	0					Centrado	Escala
8	leucos	Numérico	7	0					Centrado	Escala
9	linfopct	Numérico	7	0					Centrado	Escala
10	hemat	Numérico	7	0					Centrado	Ordinal
11	gluc	Numérico	7	0					Centrado	Escala
12	obes	Numérico	2	0					Centrado	Ordinal
13	desnutr	Numérico	2	0		{1, si}...	Ninguno	6	Centrado	Ordinal
14	diabetes	Numérico	2	0		{1, si}...	Ninguno	5	Centrado	Ordinal
15	infec	Numérico	8	0		{0, no}...	Ninguno	8	Derecha	Escala
16	infec1	Numérico	8	0		{1, si}...	Ninguno	8	Derecha	Escala
17	edad_co	Numérico	8	2		Ninguno	Ninguno	8	Derecha	Escala
18										
19										
20										
21										
22										
23										
24										
25										
26										
27										

Etiquetas de valor

Etiquetas de valor:

Valor: 2.00

Etiqueta de valor: <65 años

1.00 = ">=65 años"

2.00 = "<65 años"

Aceptar Cancelar Ayuda

Vista de datos Vista de variables

SPSS El procesador está preparado

Vista de variables

## Asociación entre infec y edad\_co....

1 : edad

59

edad sexo patol tip\_oper alb hb hcto leucos linfopct hemat gluc obes desnutr diabet infec

1 59 mujer inflam sucia 4 13 38 19090 12 4 68 si no si si

2 65 mujer otras pot\_cont 4 12 36 6190 33 4 79 si no si no

3 69 varón neo contam 4 13 38 6360 28 4 85 . no si si si

4 70 mujer neo . no si si no

5 79 mujer inflam . no si si

6 55 mujer otras si no si si

7 56 mujer traum no no si si

8 65 mujer otras si no si no

9 70 mujer neo no no si no

10 80 mujer otras no no si no

11 53 mujer neo no si si no

12 77 mujer neo no no si si

13 72 mujer inflam si si si no

14 85 mujer inflam si no si no

15 80 varón neo si no si si

16 74 varón neo no si si no

17 77 varón inflam no no si no

18 79 mujer neo si no si no

19 64 varón inflam no no si no

20 86 mujer neo pot\_cont 3 7 25 13590 13 4 112 si no si no

21 58 varón otras limpia 4 12 36 2600 28 4 112 no si si si

22 56 varón inflam sucia . 12 35 8800 23 4 113 no no si no

23 61 mujer inflam contam . 14 39 14390 11 5 113 si no si no

24 81 varón otras limpia 3 13 41 7004 13 4 113 no no si no

25 60 mujer neo pot\_cont . 18 54 6360 37 6 113 no no si si

26 68 varón neo pot\_cont 4 11 34 10720 14 4 114 no no si no

SPSS El procesador está preparado

**Tabla de contingencia EDAD\_CO \* INFEC**

			INFEC		Total
			no	si	
EDAD_CO	>=65 años	Recuento	651	245	896
		% de EDAD_CO	72.7%	27.3%	100.0%
	<65 años	Recuento	1236	219	1455
		% de EDAD_CO	84.9%	15.1%	100.0%
Total		Recuento	1887	464	2351
		% de EDAD_CO	80.3%	19.7%	100.0%

**Pruebas de chi-cuadrado**

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	52.892 <sup>b</sup>	1	.000
Corrección por continuidad	52.119	1	.000
N de casos válidos	2351		

a.

b.

*“La edad esta significativamente asociada a la infección post-quirúrgica (p-value<0.001)”:  
La presencia de infección es de un 27% en pacientes ≥65 años, frente a un 15.1% en  
pacientes <65 años.*



## EJEMPLO (*infec.sav*)

Comprobar si existe asociación entre Linfocitos% (**linfopct**) e infección post-quirúrgica (**infec**), pero considerando la variable **Linfopct** (que es continua) como categórica.

1. Categorizamos la variable Linfopct en tres niveles:

1= "Normal" = " $25\% \leq \text{linfopct} \leq 33\%$ "

2="linfocitosis" = " $\text{linfopct} > 33\%$ "

3="linfocitopenia" = " $\text{linfopct} < 25\%$ ".

creando la variable categórica **linf\_co**.

2. Obtenemos la siguiente tabla de contingencia de **infec** y **linf\_co**:

Tabla de contingencia INFEC * LINF_CO						
			LINF_CO			
			normal	linfocitosis	linfocitopenia	Total
INFEC	no	Recuento	443	460	959	1862
		% de LINF_CO	88.2%	90.7%	72.9%	80.1%
	si	Recuento	59	47	356	462
		% de LINF_CO	11.8%	9.3%	27.1%	19.9%
Total		Recuento	502	507	1315	2324
		% de LINF_CO	100.0%	100.0%	100.0%	100.0%

3. Aplicamos el test chi-cuadrado obteniendo un ***p-value* < 0.001**.

## ANÁLISIS ESTRATIFICADOS

En ocasiones dos variables categóricas están (o no) asociadas. Sin embargo, puede existir una tercera variable, C, que “modifica”, “confunde” o “interacciona” en la relación entre A y B.

### a) RELACIÓN ENTRE INFECCIÓN POST-QUIRÚRGICA Y DIABETES.

Tabla de contingencia DIABETES \* INFEC

			INFEC		
			no	si	Total
DIABETES	si	Recuento	97	45	142
		% de DIABETES	68.3%	31.7%	100.0%
	no	Recuento	1792	419	2211
		% de DIABETES	81.0%	19.0%	100.0%
Total		Recuento	1889	464	2353
		% de DIABETES	80.3%	19.7%	100.0%

- En la muestra global, existe una asociación estadísticamente significativa entre diabetes e infección post-quirúrgica ( $p < 0.001$ ).
- Como se puede observar en la tabla de contingencia, la presencia de infección post-operatoria en diabéticos es del 31.7% y en los no diabéticos de un 19%.

### b) RELACIÓN ENTRE INFECCIÓN POST-QUIRÚRGICA Y DIABETES, ESTRATIFICADA POR EDAD.

- Como ya sabemos la edad\_co y la diabetes, por separado, están significativamente asociadas a la infección post-operatoria.
- En la siguiente tabla podemos comprobar que edad\_co y diabetes están asociadas:

Tabla de contingencia EDAD\_CO \* DIABETES

			DIABETES		
			si	no	Total
EDAD_CO	>=65 años	Recuento	95	801	896
		% de EDAD_CO	10.6%	89.4%	100.0%
	<65 años	Recuento	47	1408	1455
		% de EDAD_CO	3.2%	96.8%	100.0%
Total	Recuento	142	2209	2351	
	% de EDAD_CO	6.0%	94.0%	100.0%	

$p < 0.0001$

- Observamos que la presencia de diabetes es mayor en pacientes  $\geq 65$  años (10.6%) que en pacientes  $< 65$  años (3.2%).
- Si “eliminamos” el efecto de la edad ¿podemos decir que la diabetes está asociada a la infección post-quirúrgica? ¿en qué sentido?
- Respondemos a estas preguntas, “estratificando” por edades ( $\geq 65$ ,  $< 65$ ) y viendo en cada subgrupo de edades la asociación entre Infección post-quirúrgica y diabetes.

## ANÁLISIS ESTRATIFICADOS EN EL SPSS

→ **Analizar...**

→ **Estadísticos descriptivos.....**

→ **Tablas de contingencia.....**

→ **Filas= diabetes** , **Columnas= infec**

→ **Casillas.....filas** (porcentaje de infección por diabetes si,no )

→ **Capa 1 de 1=edad\_co**

→ **Estadísticos.....Chi-cuadrado**

The screenshot shows the SPSS 'Editor de datos' window with the 'Tablas de contingencia' dialog box open. The 'Filas' box contains 'diabetes' and the 'Columnas' box contains 'infec'. The 'Capa 1 de 1' button is highlighted. The 'Estadísticos' sub-dialog box is also open, showing the 'Chi-cuadrado' option checked. The data grid in the background shows columns for various variables including 'os', 'linfopct', 'hemat', 'gluc', 'obes', 'desnutr', 'diabet', and 'infec'.

obteniendo como resultados.....

Tabla de contingencia DIABETES \* INFEC \* EDAD\_CO

				INFEC		
EDAD_CO				no	si	Total
>=65 años	DIABETES	si	Recuento	66	29	95
			% de DIABETES	69.5%	30.5%	100.0%
		no	Recuento	585	216	801
			% de DIABETES	73.0%	27.0%	100.0%
	Total		Recuento	651	245	896
			% de DIABETES	72.7%	27.3%	100.0%
<65 años	DIABETES	si	Recuento	31	16	47
			% de DIABETES	66.0%	34.0%	100.0%
		no	Recuento	1205	203	1408
			% de DIABETES	85.6%	14.4%	100.0%
	Total		Recuento	1236	219	1455
			% de DIABETES	84.9%	15.1%	100.0%

#### Pruebas de chi-cuadrado

EDAD_CO		Valor	gl	Sig. asintótica (bilateral)
>=65 años	Chi-cuadrado de Pearson	.542	1	.462
	Corrección por continuidad	.377	1	.539
	N de casos válidos	896		
<65 años	Chi-cuadrado de Pearson	13.700	1	.000
	Corrección por continuidad	12.208	1	.000
	N de casos válidos	1455		

1) **Para pacientes con ≥65 años:**

No existe asociación significativa entre diabetes e infección post-quirúrgica ( $p=0.539$ ): Como se puede observar en la tabla de contingencia, la presencia de infección post-operatoria en diabéticos es del 30.5% y en los no diabéticos de un 27%.

2) **Para pacientes con <65 años:**

La relación entre diabetes e infección post-quirúrgica es significativa ( $p<0.001$ ): Como se puede observar en la tabla de contingencia, la presencia de infección post-operatoria en diabéticos es del 34% y en los no diabéticos de un 14%.