

# Prueba de Independencia

- También es conocida como tablas de contingencia (Dependencia)
- Es una de las aplicaciones más usadas del test de Bondad de Ajuste, que consiste en contrastar si dos maneras de clasificar  $n$  objetos son independientes o no.
- Se tienen dos variables cualitativas:  $X$  e  $Y$
- Sean  $x_1, x_2, \dots, x_m$  los distintos atributos o niveles del criterio  $X$  y  $y_1, y_2, \dots, y_p$  los distintos atributos o niveles para el criterio  $Y$ .

## Prueba de independencia

Se toma una muestra de tamaño  $n$  y se registran las frecuencias observadas que llamaremos  $o_{ij}$ , donde  $o_{ij}$  representa los individuos que tienen los atributos (niveles)  $x_i$  e  $y_j$ , con  $i = 1, 2, \dots, m$  y  $j = 1, 2, \dots, p$

$X \setminus Y$	$y_1$	$y_2$	...	$y_p$	TOTALES
$x_1$	$o_{11}$	$o_{12}$	...	$o_{1p}$	$Tx_1$
$x_2$	$o_{21}$	$o_{22}$	...	$o_{2p}$	$Tx_2$
:	:	:	:	:	:
$x_m$	$o_{m1}$	$o_{12}$	...	$o_{mp}$	$Tx_m$
TOTALES	$Ty_1$	$Ty_2$	...	$Ty_p$	$n$

$Tx_i$ : total de individuos observados que poseen el atributo  $x_i$

$Ty_j$ : total de individuos observados que poseen el atributo  $y_j$

## Prueba de independencia

Recordemos que al ser, la prueba de independencia una prueba de Bondad de Ajuste, se requiere contrastar lo observado con el valor estimado según modelo teórico donde se asuma independencia.

Para el modelo teórico, debemos considerar lo siguiente:

$P(X = x_i \wedge Y = y_j) = \frac{e_{ij}}{n}$ , donde  $e_{ij}$  es la frecuencia real esperada de individuos que poseen simultáneamente los atributos  $x_i$  e  $y_j$ .

Note que en los datos de la tabla anterior lo que ocurre es que  $o_{ij}$  es una estimación de  $e_{ij}$

Cola derecha

TODAS son de Cola derecha

Si se cumple

$\chi^2_{obs} < \chi^2_c$ , NO se rechaza  $H_0$

Si se cumple

$\chi^2_{obs} > \chi^2_c$ , se rechaza  $H_0$

Pruebas de independencia

Son un caso especial de bondad de ajuste

### Ejemplo 4:

Los empleados de una empresa trabajan una jornada de 8 horas distribuidas en dos tandas de horario: el primero corresponde al bloque de la "mañana" en horario 8:00-12:00, luego se da una hora para el almuerzo y posteriormente llega el bloque de la "tarde" en horario 13:00-17:00. Se valoró si su desempeño, en cada uno de los bloques (mañana o tarde), era bueno, regular o malo. Con un nivel de significancia del 5%, ¿se puede asegurar que el desempeño de los empleados en ambos bloques de horarios son independientes?

		$\gamma_1$	$\gamma_{TARDE}$	$\gamma_3$	
		Bueno	Regular	Malo	TOTAL $X$
MAÑANA	Bueno	56 $O_{11}$	71 $O_{12}$	12 $O_{13}$	139 $T_{X1}$
	Regular	47 $O_{21}$	163 $O_{22}$	38 $O_{23}$	248 $T_{X2}$
	Malo	14 $O_{31}$	42 $O_{32}$	85 $O_{33}$	141 $T_{X3}$
TOTAL $Y$		117 $T_{Y1}$	276 $T_{Y2}$	135 $T_{Y3}$	528

→ h

Sea  $X$ : Desempeño de los de la mañana

Sea  $Y$ : Desempeño de los de la tarde

Planteamiento de hipótesis:

H<sub>0</sub>:  $X, Y$  son independientes

H<sub>1</sub>:  $X, Y$  no son independientes

Al igual que en bondad de ajuste, se usan observados y esperados

Se calculan con:

$$e_{ij} = \frac{T_{ij} - T_{i\cdot} \cdot T_{\cdot j}}{h}, \quad T = \text{Total}$$

$h = \text{Total de elementos}$

	$\gamma_1$	$\gamma_{TARDE} \gamma_2$	$\gamma_3$	
	Bueno	Regular	Malo	<b>TOTAL</b>
<b>MAÑANA</b>				
Bueno	56 $\textcolor{green}{O}_{11}$	71 $\textcolor{green}{O}_{12}$	12 $\textcolor{green}{O}_{13}$	139 $\textcolor{red}{T}_{x1}$
Regular	47 $\textcolor{green}{O}_{21}$	163 $\textcolor{green}{O}_{22}$	38 $\textcolor{green}{O}_{23}$	248 $\textcolor{red}{T}_{x2}$
Malo	14 $\textcolor{green}{O}_{31}$	42 $\textcolor{green}{O}_{32}$	85 $\textcolor{green}{O}_{33}$	141 $\textcolor{red}{T}_{x3}$
<b>TOTAL</b>	117 $\textcolor{red}{T}_{y1}$	276 $\textcolor{red}{T}_{y2}$	135 $\textcolor{red}{T}_{y3}$	528

Se tiene:

$$e_{11} = \frac{T_{x1} \cdot T_{y1}}{n} = \frac{139 \cdot 117}{528} = 30,80$$

$$e_{12} = \frac{T_{x1} \cdot T_{y2}}{n} = \frac{139 \cdot 276}{528} = 72,66$$

$$e_{13} = \frac{T_{x1} \cdot T_{y3}}{n} = \frac{139 \cdot 135}{528} = 38,54$$

$$e_{21} = \frac{T_{x2} \cdot T_{y1}}{n} = \frac{248 \cdot 117}{528} = 54,95$$

$$e_{22} = \frac{T_{x2} \cdot T_{y2}}{n} = \frac{248 \cdot 276}{528} = 129,64$$

$$e_{23} = \frac{T_{x2} \cdot T_{y3}}{n} = \frac{248 \cdot 135}{528} = 63,41$$

$$e_{31} = \frac{T_{x3} \cdot T_{y1}}{n} = \frac{141 \cdot 117}{528} = 31,24$$

$$e_{32} = \frac{T_{x3} \cdot T_{y2}}{n} = \frac{141 \cdot 276}{528} = 73,70$$

$$e_{33} = \frac{T_{x3} \cdot T_{y3}}{n} = \frac{141 \cdot 135}{528} = 36,05$$

En base a esos resultados debe hacerse una nueva tabla similar pero con los nuevos datos

		$\gamma_1$	TARDE $\gamma_2$	$\gamma_3$	TOTAL	
		Bueno	Regular	Malo		
$\chi^2$ MAÑANA	Bueno	56 $O_{11}$	71 $O_{12}$	12 $O_{13}$	139 $T_{x1}$	
	Regular	47 $O_{21}$	163 $O_{22}$	38 $O_{23}$	248 $T_{x2}$	
	Malo	14 $O_{31}$	42 $O_{32}$	85 $O_{33}$	141 $T_{x3}$	
	TOTAL	117 $T_{y1}$	276 $T_{y2}$	135 $T_{y3}$	528	

} Tabla  
} Vieja i

Tarde

	Bueno	Regular	Malo	Total	Tabla nueva j
Bueno	30,80	72,66	35,54	139	
Regular	57,95	129,64	63,41	248	
Malo	31,24	73,70	36,05	141	
Total	117	276	135	528	

Para probar independencia cuando cada  $e_{ij} \geq 5$

Estadístico de prueba:  $\chi^2 = \sum_{i=1}^f \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$   
con  $\nu = (f-1)(c-1)$ , si  $\nu \geq 2$ .

Para el  $\chi^2_{0.05}$ ,  
se suman todos

$$\frac{(56-30,80)^2}{30,80} + \frac{(71-72,66)^2}{72,66} + \frac{(12-35,54)^2}{35,54} +$$

$$\frac{(47-57,95)^2}{57,95} + \frac{(163-129,64)^2}{129,64} + \frac{(38-63,41)^2}{63,41} +$$

$$\frac{(14-31,24)^2}{31,24} + \frac{(42-73,70)^2}{73,70} + \frac{(85-36,05)^2}{36,05}$$

$$\chi^2_{0.05} = 145,7812 \quad \alpha = 0,05 \quad \nu = (f-1)(c-1),$$

$$\nu = (3-1)(3-1) = 4$$

$$\chi^2_c = \chi^2_{0.05, 4} = 9,48773 \quad 3 \times 3 \rightarrow 2 \times 2$$

$$9 \times 4 \rightarrow 3 \times 3$$

II  $\chi^2_{0.05} = 145,7812 > 9,48773$   
se rechaza  $H_0$ , NO es independiente

Con valor P

$$\chi^2_{\text{obs}} = 145,7812 \quad v=7 \quad \alpha=0,05$$

$$P(\chi^2 > 145,7812) = 0,00000$$

$$P=0,00000 \quad \alpha=0,05$$

Se rechaza  $H_0$ , No es independiente

## Ejemplo 3:

En un estudio de una vacuna de hepatitis participan 1083 voluntarios. De éstos, se eligen aleatoriamente 549 y son vacunados. Los otros, 534, no son vacunados. Después de un cierto tiempo, se observa que 70 de los 534 no vacunados han contraído la hepatitis, mientras que sólo 11 de los 549 vacunados la han contraído.

¿Es el hecho de contraer hepatitis independiente de haber sido vacunado contra la enfermedad?

$$\begin{array}{ll} V = \text{Vacunados} & NV = \text{No Vacunados} \\ H = \text{Hepatitis} & NH = \text{No Hepatitis} \end{array}$$

	H	NH	Total
V	11	538	549
NV	70	964	534
Total	81	1002	1083 → n

Calcular esperados  $\frac{T_{xi} \cdot T_{yj}}{n}$ ,  $n = 1083$

$$E_{11} = \frac{T_{x1} \cdot T_{y1}}{n} = \frac{549 \cdot 81}{1083} \approx 41,06$$

$$E_{12} = \frac{T_{x1} \cdot T_{y2}}{n} = \frac{549 \cdot 1002}{1083} \approx 507,97$$

$$E_{21} = \frac{T_{x2} \cdot T_{y1}}{n} = \frac{534 \cdot 81}{1083} \approx 39,97$$

$$E_{22} = \frac{T_{x2} \cdot T_{y2}}{n} = \frac{534 \cdot 1002}{1083} \approx 499,06$$

Tabla vieja

	H	NH	Total x
V	11	538	549
NV	70	964	534
Total y	81	1002	1083

Tabla nueva

	H	NH	Total x
V	71,06	507,97	549
NV	39,97	797,06	534
Total y	81	1002	1083

Para probar independencia cuando cada  $e_{ij} \geq 5$ 

Estadístico de prueba:  $\chi^2 = \sum_{i=1}^f \sum_{j=1}^c \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$   
con  $\nu = (f-1)(c-1)$ , si  $\nu \geq 2$ .

$$\begin{aligned}\chi^2_{0.05} &= \frac{(11 - 71,06)^2}{71,06} + \frac{(538 - 507,97)^2}{507,97} + \\ &\quad \frac{(70 - 39,97)^2}{39,97} + \frac{(964 - 797,06)^2}{797,06}\end{aligned}$$

$$\chi^2_{0.05} = 48,2388$$

$$\nu = (2-1)(2-1) = 1$$

$$\alpha = 0,05$$

$$\chi^2_C = \chi^2_{0.05}, 1 = 3,84146$$

R/ Como  $\chi^2_{0.05} = 48,2388 > \chi^2_C = 3,84146$ ,  
se rechaza  $H_0$ , NO son independientes

Con valor P

$$P(\chi^2 > 48,2388) = 0,0000 \quad \alpha = 0,05$$

P = 0,0000 ( $\alpha = 0,05$ ), se rechaza  $H_0$   
NO son independientes

## Ejemplo

En una universidad se realiza un estudio para verificar si el tipo de trabajo (administrativo y docente) se relaciona con el grado de estrés (I, II y III) de los trabajadores. Para lo cual se elige una muestra aleatoria de 300 trabajadores y se clasifican en la tabla siguiente. (5 puntos)

	I	II	III
Administrativos	42	24	30
Docentes	54	78	72

Pruebe la hipótesis de que el tipo de trabajo afecta el grado de estrés del trabajador.

$H_0$ : El trabajo y estreses son independientes  
 $H_1$ : El trabajo y estreses no son independientes

Tabla inicial  $A = \text{Administrativos}$   $D = \text{Docentes}$

	I	II	III	Total x
A	96	24	30	96
D	54	78	72	207
Total y	96	102	102	300 $\rightarrow n$

Calcular esperados  $E_{ij} = \frac{T_{xi} \cdot T_{yj}}{n}$ ,  $n = 300$

$$E_{11} = \frac{96 \cdot 96}{300} = 30,72 \quad E_{21} = \frac{207 \cdot 96}{300} = 65,28$$

$$E_{12} = \frac{96 \cdot 102}{300} = 32,64 \quad E_{22} = \frac{207 \cdot 102}{300} = 69,36$$

$$E_{13} = \frac{96 \cdot 102}{300} = 32,64 \quad E_{23} = \frac{207 \cdot 102}{300} = 69,36$$

Tabla vieja

	I	II	III	Total x
A	92	24	30	96
D	58	78	72	209
Total y	96	102	102	300

Para probar independencia cuando cada  $e_{ij} \geq 5$

Estadístico de prueba:

$$\chi^2 = \sum_{i=1}^f \sum_{j=1}^c \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

con  $\nu = (f-1)(c-1)$ , si  $\nu \geq 2$ .

Tabla Nueva

	I	II	III	Total x
A	30,72	32,69	32,69	96
D	65,28	69,36	69,36	209
Total y	96	102	102	300

$$\begin{aligned} \chi^2_{\text{obs}} &= \frac{(92-30,72)^2}{30,72} + \frac{(24-32,69)^2}{32,69} + \frac{(30-32,69)^2}{32,69} + \\ &\quad \frac{(58-65,28)^2}{65,28} + \frac{(78-69,36)^2}{69,36} + \frac{(72-69,36)^2}{69,36} \end{aligned}$$

$$\chi^2_{\text{obs}} = 9,7683$$

$$v = (2-1) \cdot (3-1) = 2$$

$$\alpha = 0,05$$

$$\chi^2_c = \chi^2_{0,05,2} = 5,99196$$

R/ Como  $\chi^2_{\text{obs}} = 9,7683 > \chi^2_c = 5,99196$   
se rechaza  $H_0$ , NO son independientes

Con valor P

$$P(\chi^2 > 9,7683) = 0,007566 \quad \alpha = 0,05$$

Como  $P = 0,007566 < 0,05$  se rechaza  $H_0$ ,  
NO son independientes