



# Análisis no paramétrico:



Spearman, Chi- cuadrado



# Análisis No Paramétrico

## ¿Qué es?

Los métodos no paramétricos no asumen una distribución específica (como la normal).

Se aplican cuando:

- Los datos son ordinales o nominales
- Hay pocos datos o no se cumple normalidad
- Las varianzas son muy distintas entre grupos



# Coeficiente de Correlación de Spearman ( $\rho$ o $R_s$ )

## ★ ¿Qué mide?

La fuerza y dirección de la relación monótona (no necesariamente lineal) entre dos variables ordinales o cuantitativas mal distribuidas.

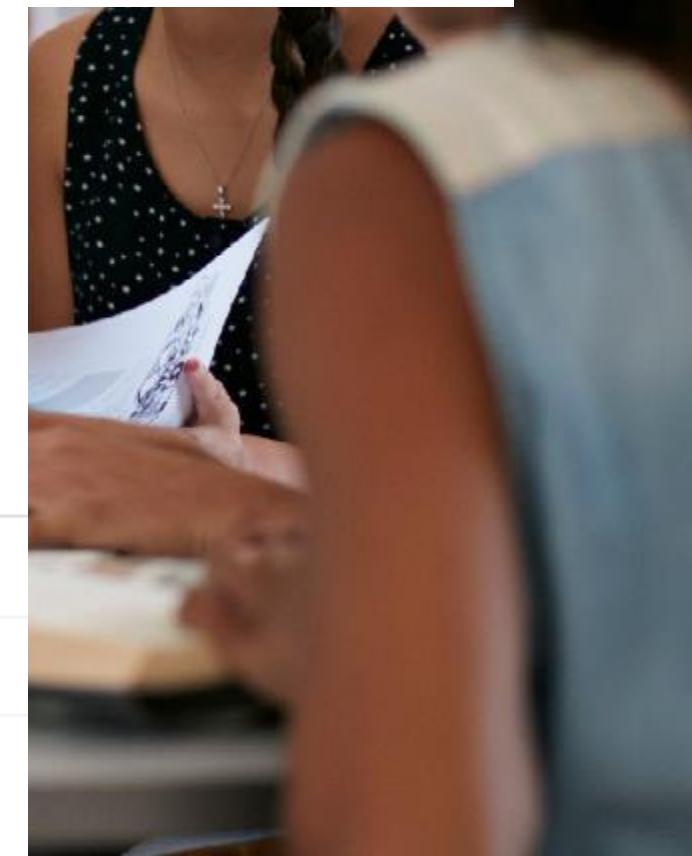
### 1. "La fuerza"

Esto significa qué tan fuerte es la relación entre dos variables.

- Si dos variables suben o bajan juntas consistentemente, la fuerza es alta.
- Si los valores varían sin patrón claro, la fuerza es baja.

El coeficiente de Spearman ( $\rho$ ) va de -1 a +1:

Valor de $\rho$	Interpretación
1	Relación perfectamente creciente
0	Sin relación
-1	Relación perfectamente decreciente



## 2. "Y dirección"

Se refiere a si una variable sube o baja cuando la otra cambia:

- Dirección positiva:

Cuando una variable aumenta, la otra también.

Ejemplo: más horas de estudio → mejores calificaciones

- Dirección negativa:

Cuando una variable aumenta, la otra disminuye.

Ejemplo: más estrés → menor rendimiento académico

### 3. "Relación monótona"

Una relación monótona es aquella donde las variables siempre se mueven en la misma dirección, aunque no sea de forma constante ni perfectamente alineada.

- Si cuando X sube, Y también sube (aunque no de forma lineal) → monótona creciente
- Si cuando X sube, Y baja (aunque de forma irregular) → monótona decreciente

⚠️ No es lo mismo que una línea recta (relación lineal).

Spearman no necesita que los puntos estén en línea recta, solo que vayan en una misma dirección.

#### 4. "Variables ordinales o cuantitativas mal distribuidas"

- **Ordinales:** Son variables que se pueden ordenar pero no tienen distancia fija entre categorías.  
Ejemplo: satisfacción del cliente (baja, media, alta)
- **Cuantitativas mal distribuidas:**  
Son variables numéricas (como edad, nota, tiempo), pero no siguen una distribución normal.  
→ Por ejemplo, si casi todos los valores están concentrados en un extremo

En esos casos, no puedes usar Pearson, pero sí puedes usar Spearman.



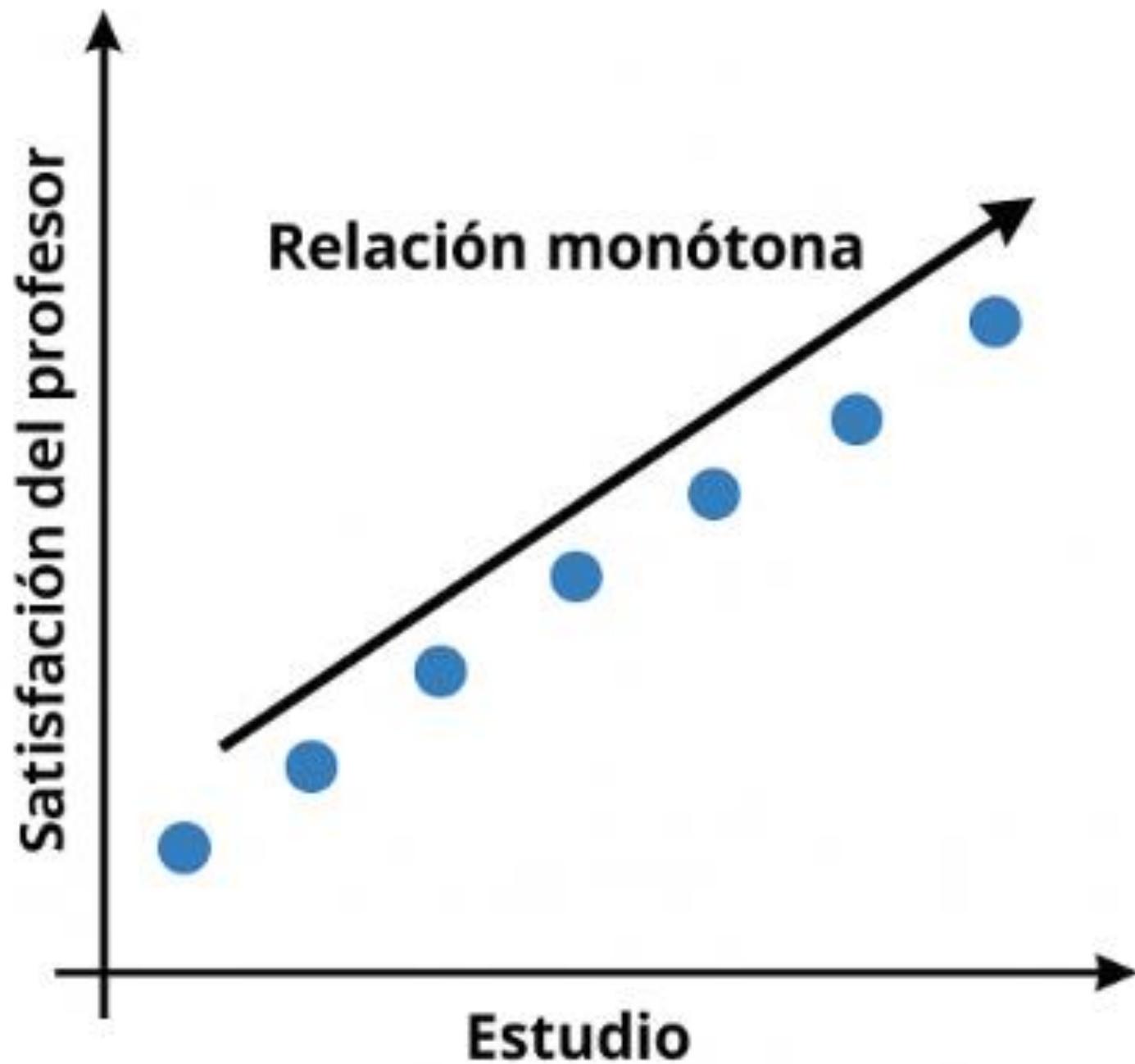
## Ejemplo real:

Imagina estas dos variables:

- Tiempo que pasa un estudiante estudiando (X)
- Nivel de satisfacción del profesor con ese estudiante (Y)

No necesitas que los datos sean perfectos ni numéricamente precisos. Basta con que:  
a más estudio, mayor satisfacción del profesor → relación monótona positiva

*Lorem ipsum dolor sit amet, consectetuer adipiscing elit*



## □ Fórmula de Spearman (rangos sin empate):

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

### ■ Elementos:

- $n$ : número de pares de datos
- $d_i$ : diferencia entre los rangos de cada par
- $\sum d_i^2$ : suma de los cuadrados de esas diferencias
- $\rho$ : valor entre -1 y 1 (igual que Pearson, pero para rangos)



## ¿Qué significa "empate" en rangos?

Cuando ordenas los valores de una variable de menor a mayor para asignar rangos, puede ocurrir que dos o más valores sean iguales. A eso se le llama empate.

---



### Ejemplo sin empate (todo diferente):

Valores: [3, 5, 7]

Rangos: 1, 2, 3

Aquí no hay empates. Cada valor ocupa una posición única.

## Ejemplo con empate:

Valores: [3, 5, 5, 7]

Primero los ordenas:

→ 3 (1.<sup>º</sup>), 5 (2.<sup>º</sup> y 3.<sup>º</sup>), 5 (2.<sup>º</sup> y 3.<sup>º</sup>), 7 (4.<sup>º</sup>)

Los dos valores 5 están empatados en la posición 2 y 3. Entonces, ambos reciben el promedio de sus rangos:

$$\text{Rango de } 5 = \frac{2 + 3}{2} = 2.5$$

Rangos finales:

- 3 → 1
- 5 → 2.5
- 5 → 2.5
- 7 → 4

## ¿Por qué es importante?

La fórmula original de Spearman **sin empates solo sirve si no hay valores repetidos.**

**Cuando hay empates,** se debe usar una fórmula **más general**, o calcular directamente con software estadístico.



## Ejemplo con empates en Spearman



### Datos iniciales:

Persona	X (Horas de estudio)	Y (Nota)
A	3	80
B	6	85
C	6	90
D	9	85
E	12	95

---

## □ Paso 1: Asignar rangos

**Variable X:**

Valores ordenados: 3, 6, 6, 9, 12

- 3 → rango 1
- 6 (empate: posición 2 y 3) →  $(2 + 3)/2 = 2.5$
- 9 → rango 4
- 12 → rango 5

## Rangos X:

Persona	X	Rango X
A	3	1
B	6	2.5
C	6	2.5
D	9	4
E	12	5

### Variable Y:

Valores ordenados: 80, 85, 85, 90, 95

- 80 → rango 1
- 85 (empate: posición 2 y 3) →  $(2 + 3)/2 = 2.5$
- 90 → rango 4
- 95 → rango 5

### Rangos Y:

Persona	Y	Rango Y
A	80	1
B	85	2.5
C	90	4
D	85	2.5
E	95	5

□ **Paso 2: Calcular diferencias de rangos  $d_i = R_X - R_Y$**

Persona	Rango X	Rango Y	$d_i$	$d_i^2$
A	1	1	0	0
B	2.5	2.5	0	0
C	2.5	4	-1.5	2.25
D	4	2.5	1.5	2.25
E	5	5	0	0

$$\sum d_i^2 = 2.25 + 2.25 = 4.5$$

## □ Paso 3: Aplicar fórmula de Spearman con rangos

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} = 1 - \frac{6(4.5)}{5(25 - 1)} = 1 - \frac{27}{120} = 1 - 0.225 = 0.775$$



### Interpretación final:

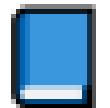
- $\rho = 0.775 \rightarrow$  correlación positiva fuerte
- A mayor número de horas de estudio, tiende a mejorar la nota (aunque no perfectamente).



# Análisis no paramétrico:

• • •

Chi- cuadrado



## Prueba de Chi-cuadrado de independencia ( $\chi^2$ )



### ¿Para qué sirve?

Para determinar si dos variables categóricas están asociadas o son independientes.

• • •



## Ejemplos de uso típico:

- ¿Está relacionada la edad con la preferencia por un producto?
- ¿Está asociado el género con la frecuencia de ejercicio?
- ¿Hay relación entre nivel educativo y uso de redes sociales?

## □ Fórmula de Chi-cuadrado:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

### ○ Elementos:

- $O_i$ : frecuencia observada
- $E_i$ : frecuencia esperada
- $\chi^2$ : estadístico que mide la diferencia acumulada entre observado y esperado

## Ejemplo práctico

### ?

#### Planteamiento:

¿Está relacionada la asistencia a tutorías con aprobar el examen?

---

### □ Paso 1: Tabla de frecuencias observadas

	Aprobó	No aprobó	Total
Asistió	30	10	40
No asistió	10	20	30
Total	40	30	70

## □ Paso 2: Calcular frecuencias esperadas

$$E_{ij} = \frac{(\text{Total fila}_i)(\text{Total columna}_j)}{\text{Total general}}$$

	Aprobó (E)	No aprobó (E)
Asistió	$40 \times 40/70 = 22.86$	$40 \times 30/70 = 17.14$
No asistió	$30 \times 40/70 = 17.14$	$30 \times 30/70 = 12.86$

### □ Paso 3: Calcular $\chi^2$

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

$$\chi^2 = \frac{(30 - 22.86)^2}{22.86} + \frac{(10 - 17.14)^2}{17.14} + \frac{(10 - 17.14)^2}{17.14} + \frac{(20 - 12.86)^2}{12.86}$$

$$\chi^2 \approx \frac{50.59}{22.86} + \frac{50.94}{17.14} + \frac{50.94}{17.14} + \frac{50.59}{12.86} \approx 2.21 + 2.97 + 2.97 + 3.93 \approx \boxed{12.08}$$



## Paso 4: Determinar grados de libertad (gl)

$$gl = (n^{\circ} \text{ filas} - 1)(n^{\circ} \text{ columnas} - 1) = (2 - 1)(2 - 1) = 1$$



### Valor crítico ( $\alpha = 0.05$ , $gl = 1$ )

De la tabla Chi-cuadrado:

$$\chi_{\text{crítico}}^2 = 3.841$$

---



### Conclusión

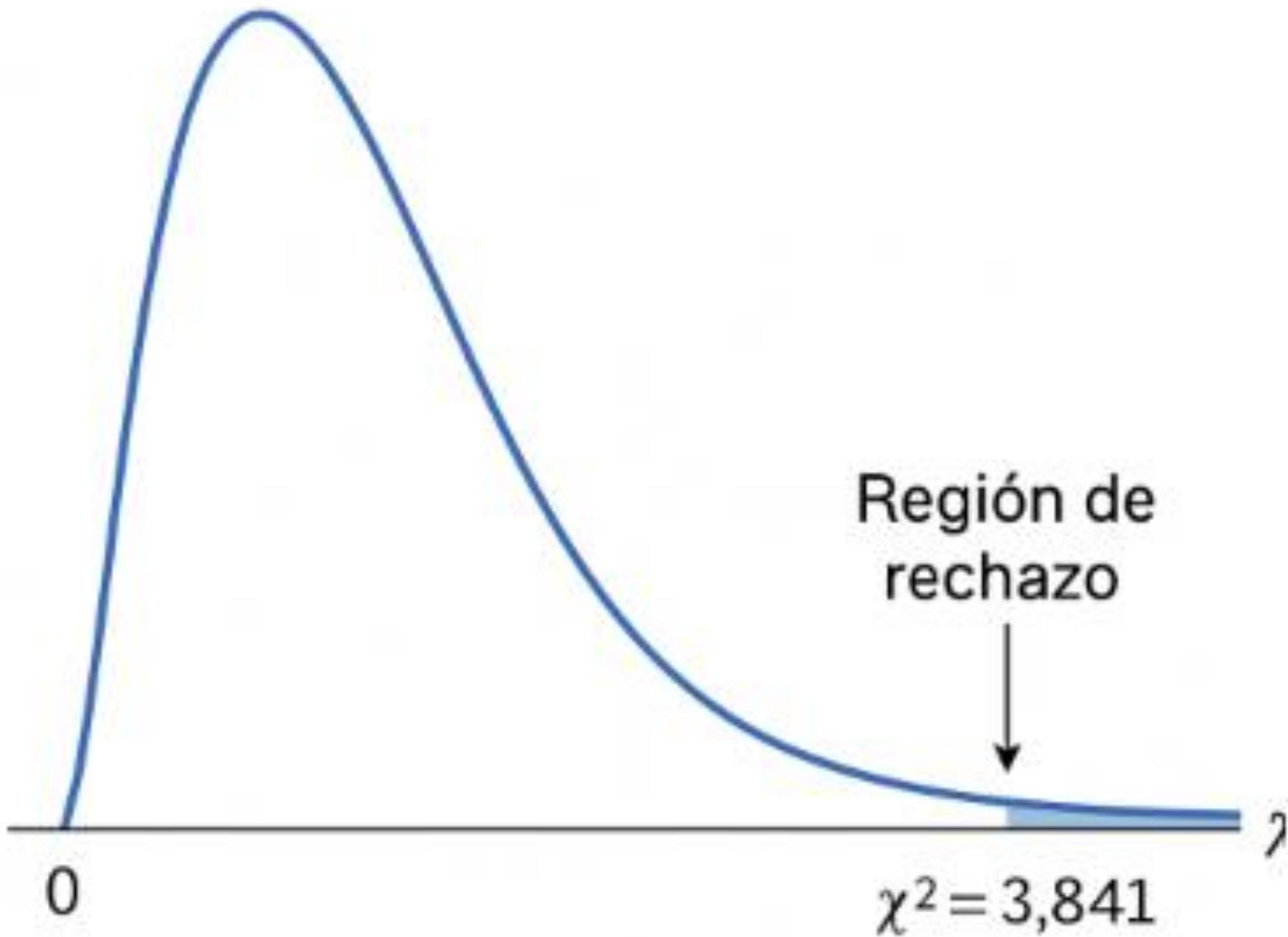
$$\chi^2 = 12.08 > 3.841 \rightarrow \text{Rechazamos } H_0$$

→ Existe una asociación significativa entre asistir a tutorías y aprobar el examen.



## Visualización de la distribución $\chi^2$ :

- Área sombreada a la derecha del valor crítico = región de rechazo
- Si tu estadístico cae allí, rechazas la independencia



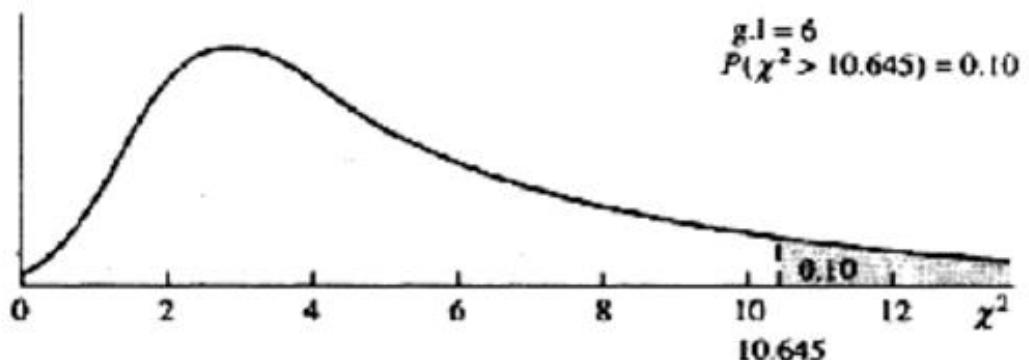
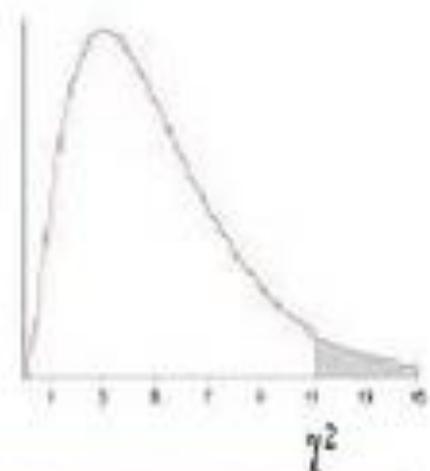


Tabla H Distribución chi-cuadrado

g.l.	$\chi^2_{0.995}$	$\chi^2_{0.990}$	$\chi^2_{0.975}$	$\chi^2_{0.950}$	$\chi^2_{0.900}$	$\chi^2_{0.700}$	$\chi^2_{0.500}$	$\chi^2_{0.300}$	$\chi^2_{0.200}$	$\chi^2_{0.100}$	$\chi^2_{0.050}$	$\chi^2_{0.025}$	$\chi^2_{0.010}$	$\chi^2_{0.001}$	
1	0.000	0.000	0.001	0.004	0.016	0.148	0.455	1.074	1.642	2.706	3.841	5.024	5.412	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	0.713	1.386	2.408	3.219	4.605	5.991	7.378	7.824	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	1.424	2.366	3.665	4.642	6.251	7.815	9.348	9.837	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	2.195	3.357	4.878	5.989	7.779	9.488	11.143	11.668	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	3.000	4.351	6.064	7.289	9.236	11.070	12.833	13.388	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	3.828	5.348	7.231	8.558	10.645	12.592	14.449	15.033	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	4.671	6.346	8.383	9.803	12.017	14.067	16.013	16.622	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	5.527	7.344	9.524	11.030	13.362	15.507	17.535	18.168	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	6.393	8.343	10.656	12.242	14.684	16.919	19.023	19.679	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	7.267	9.342	11.781	13.442	15.987	18.307	20.483	21.161	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	8.148	10.341	12.899	14.631	17.275	19.675	21.920	22.618	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	9.034	11.340	14.011	15.812	18.549	21.026	23.337	24.054	26.217	28.299
13	3.565	4.107	5.009	5.892	7.042	9.926	12.340	15.119	16.985	19.812	22.362	24.736	25.472	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	10.821	13.339	16.222	18.151	21.064	23.685	26.119	26.873	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	11.721	14.339	17.322	19.311	22.307	24.996	27.488	28.259	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	12.624	15.338	18.418	20.465	23.542	26.296	28.845	29.633	32.000	34.267
17	5.697	6.406	7.564	8.672	10.085	13.531	16.338	19.511	21.615	24.769	27.587	30.191	30.995	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	14.440	17.338	20.601	22.760	25.989	28.869	31.526	32.346	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	15.352	18.338	21.689	23.900	27.204	30.144	32.852	33.687	36.191	38.582

Tabla D.7: VALORES CRÍTICOS DE LA DISTRIBUCIÓN JI CUADRADA



g.d.l	0,041	0,005	0,01	0,02	0,025	0,03	0,04	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	g.d.J
	1	10,823	7,879	6,635	5,412	5,024	4,709	4,218	3,841	2,705	2,072	1,642	1,323	1,074	0,873	0,708
2	13,816	10,597	9,210	7,824	7,378	7,013	6,438	5,991	4,605	3,794	3,219	2,773	2,408	2,100	1,833	2
3	16,266	12,838	11,345	9,837	9,348	8,947	8,311	7,815	6,251	5,317	4,642	4,108	3,665	3,283	2,946	3
4	18,467	14,860	13,277	11,668	11,143	10,712	10,026	9,488	7,779	6,745	5,989	5,385	4,878	4,438	4,045	4
5	20,515	16,750	15,086	13,388	12,833	12,375	11,644	11,070	9,236	8,115	7,289	6,626	6,064	5,573	5,132	5



## **Ejemplo: ¿Está relacionado el género con la preferencia por un tipo de película?**



### **Planteamiento:**

Una encuesta a 100 personas preguntó su género y si prefieren acción o romance. Se quiere saber si existe una asociación significativa entre género y preferencia de película.

## □ Paso 1: Tabla de frecuencias observadas ( $O$ )

Género	Acción	Romance	Total
Hombre	30	20	50
Mujer	10	40	50
Total	40	60	100

## □ Paso 2: Calcular frecuencias esperadas (E)

$$E_{ij} = \frac{(\text{Total fila}_i)(\text{Total columna}_j)}{\text{Total general}}$$

Género	Acción (E)	Romance (E)
Hombre	$50 \cdot 40/100 = 20$	$50 \cdot 60/100 = 30$
Mujer	$50 \cdot 40/100 = 20$	$50 \cdot 60/100 = 30$

## □ Paso 3: Calcular $\chi^2$

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

$$\chi^2 = \frac{(30 - 20)^2}{20} + \frac{(20 - 30)^2}{30} + \frac{(10 - 20)^2}{20} + \frac{(40 - 30)^2}{30}$$

$$\chi^2 = \frac{100}{20} + \frac{100}{30} + \frac{100}{20} + \frac{100}{30} = 5 + 3.33 + 5 + 3.33 = \boxed{16.66}$$



## Paso 4: Grados de libertad

$$gl = (\text{filas} - 1)(\text{columnas} - 1) = (2 - 1)(2 - 1) = 1$$



Valor crítico ( $\alpha = 0.05$ ,  $gl = 1$ ):

$$\chi^2_{crítico} = 3.841$$

## Conclusión

$16.66 > 3.841 \Rightarrow$  Rechazamos  $H_0$

- Hay evidencia estadísticamente significativa de que el género influye en la preferencia por tipo de película.

## **Paso 1: Localiza los grados de libertad**

Busca la fila de  $gl = 1$ , que está en la primera fila de la tabla.

---

## **Paso 2: Busca la columna correspondiente a $\alpha = 0.05$**

En la tabla, la columna con título  $\chi^2_{0.050}$  (o "p = 0.05") está justo debajo del encabezado que indica ese nivel de significancia.



## Resultado:

Para  $gl = 1$  y  $\alpha = 0.05$ , el valor crítico es:

3.841

Ese es el número que usamos para comparar con el valor calculado del estadístico  $\chi^2$  en cualquier prueba.

---

## Interpretación:

- Si tu  $\chi^2_{calculado} > 3.841 \rightarrow$  Rechazas la  $H_0$  (hay relación entre variables)
- Si tu  $\chi^2_{calculado} < 3.841 \rightarrow$  No hay evidencia suficiente para rechazar  $H_0$

**FIN**