

Estadística II

Tema 9. Contrastes no paramétricos

Facultad de Ciencias de la Economía y de la Empresa (FCEE)

Curso 2025–2026 · 4.5 ECTS · 2º cuatrimestre

Francisco Rabadán Pérez, Raquel Ibar-Alonso y Ester Muñoz Céspedes

Departamento de Economía Aplicada I e Historia e Instituciones Económicas

Tema 9 · Objetivo

Al terminar este tema podrás:

- identificar cuándo conviene usar un contraste no paramétrico;
- resolver contrastes χ^2 de bondad del ajuste;
- resolver contrastes χ^2 de independencia y de homogeneidad;
- aplicar de forma guiada el contraste de Kolmogorov–Smirnov en una muestra;
- interpretar correctamente el resultado del contraste en contexto.

Cómo leer estas etiquetas

CORE

PLUS

ANEXO

CORE: ideas y procedimientos que sostienen el curso entero y que se practican de forma recurrente.

PLUS: material para profundizar o reforzar intuición; ayuda a consolidar.

ANEXO: material de referencia/consulta.

Tema 9 · Esquema

Esquema del tema (bloques):

1. Introducción a los contrastes no paramétricos
2. Prueba χ^2 : idea general y estructura común
3. Contrastes χ^2 de bondad del ajuste
4. Contrastes χ^2 de independencia y homogeneidad
5. Contraste de Kolmogorov–Smirnov en una muestra
6. Cierre: mapa de decisión y errores típicos

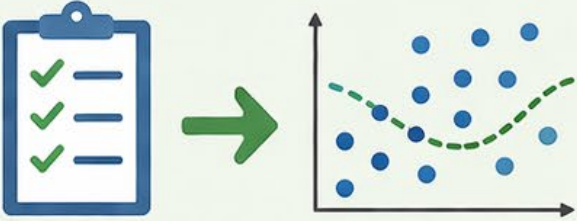
1. Introducción a los contrastes no paramétricos

1.1 ¿Qué significa “no paramétrico”?



Un **contraste no paramétrico** es un contraste de hipótesis en el que la afirmación que se somete a prueba **no se refiere al valor de un parámetro** poblacional, sino a otras **propiedades de la distribución** o de los **datos**.

IDEA CLAVE



Se comprueban afirmaciones sobre la **forma, orden, independencia, simetría**, etc., en lugar de sobre parámetros como μ , σ^2 o π .

EN CONTRASTE CON LOS PARAMÉTRICOS



Paramétricos
(parámetros)



No paramétricos
(propiedades)

¿POR QUÉ USARLOS?



No requieren supuestos estrictos sobre la distribución.



Funcionan bien con datos ordinales, categóricos o sin distribución conocida.



Son más robustos frente a desviaciones de supuestos paramétricos.



Sirven para contrastar hipótesis sobre características de la distribución o de los datos, **no sobre el valor de un parámetro poblacional**.

1.2 ¿Cuándo tiene sentido usarlos?



Los contrastes no paramétricos resultan **especialmente útiles** cuando:



DATOS EN CATEGORÍAS

Trabajamos con datos **clasificados** en categorías.

Categoría	Frecuencia
●	34
▲	21
◆	15
★	10

INFORMACIÓN EN FRECUENCIAS

La información está en forma de **frecuencias**.



NO SE USAN MEDIAS NI VARIANZAS

No es natural trabajar con **medias o varianzas**.



OBJETIVOS TÍPICOS

Interesa contrastar el **ajuste** a una distribución, la **independencia** entre variables o la **homogeneidad** entre poblaciones.



IDEA PRÁCTICA

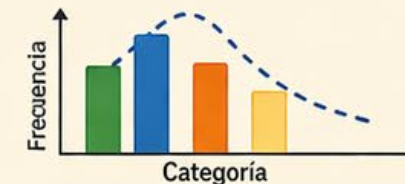
En este tema, su uso principal aparecerá en:

Categoría	Frecuencia
●	
▲	
◆	
★	

Tablas de **frecuencias**

		B			Total
		B ₁	B ₂	B ₃	
A	A ₁				
	A ₂				
	A ₃				
		25	30	20	75

Tablas de **contingencia**



— Distribución empírica
- - - Distribución teórica

Contraste entre **distribución empírica y distribución teórica**

2. Prueba χ^2 : idea general y estructura común

2.1 Idea general de la prueba χ^2



La prueba χ^2 compara las **frecuencias observadas** en la muestra con las **frecuencias esperadas** bajo la hipótesis nula (H_0).

	OBSERVADO (en la muestra)	ESPERADO (bajo H_0)	DIFERENCIA (¿son grandes?)
Categoría 1	●●●● O_1	●●● E_1	$\frac{(O_1 - E_1)^2}{E_1}$
Categoría 2	●●● O_2	●●●● E_2	$\frac{(O_2 - E_2)^2}{E_2}$
Categoría 3	●●●●● O_3	●●●● E_3	$\frac{(O_3 - E_3)^2}{E_3}$
⋮	⋮	⋮	⋮
Categoría k	●● O_k	●●● E_k	$\frac{(O_k - E_k)^2}{E_k}$

= χ^2

Cuantifica cuánto se alejan O de E en conjunto.

Se suman todas las categorías



IDEA CLAVE

- Si las diferencias entre observadas y esperadas son **pequeñas**, la muestra es **compatible con H_0** .
- Si las diferencias son grandes, la muestra aporta **evidencia contra H_0** .
- Por eso la prueba χ^2 es útil en:
 - Bondad del ajuste
 - Independencia
 - Homogeneidad



La prueba χ^2 traduce las diferencias entre lo observado y lo esperado en un único número. Cuanto mayor sea, menos compatible será la muestra con la hipótesis nula.

2.2 Estadístico χ^2



El estadístico de la prueba es:

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

donde:

- O_i : frecuencia observada en la categoría i .
- E_i : frecuencia esperada en la categoría i bajo H_0 .

¿CÓMO SE CALCULA?

Categoría i	Observada (O_i)	Esperada (E_i)	Contribución $c_i = \frac{(O_i - E_i)^2}{E_i}$
1	●●●● O_1	●●● E_1	c_1
2	●●● O_2	●●●●● E_2	c_2
3	●●●●● O_3	●●●●● E_3	c_3
⋮	⋮	⋮	⋮
k	●● O_k	●●● E_k	c_k

Se suman todas las contribuciones: $\chi^2 = \sum_{i=1}^n c_i$

INTERPRETACIÓN BÁSICA



Cuanto más próximas estén O_i y E_i , menor será la **contribución** c_i y, por tanto, el valor de χ^2 .

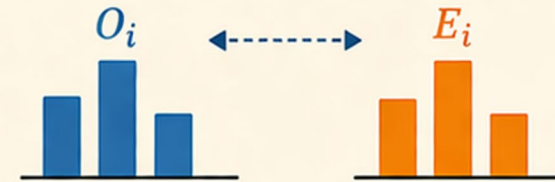


Cuanto mayores sean las discrepancias entre O_i y E_i , **mayor** será la contribución c_i y, por tanto, el valor de χ^2 .



IDEA CLAVE

El estadístico χ^2 mide la distancia global entre las frecuencias **observadas** (O_i) y las esperadas (E_i) bajo H_0 .



Si son parecidas, χ^2 es pequeño.
Si son diferentes, χ^2 es grande.



En resumen: χ^2 cuantifica cuánto se alejan los datos observados de lo que se esperaría bajo la hipótesis nula.

Más diferencia → mayor χ^2 . Menos diferencia → menor χ^2 .

2.3 Regla de decisión e interpretación

CORE

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

donde:

- O_i : frecuencia observada en la categoría i .
- E_i : frecuencia esperada en la categoría i bajo H_0 .

¿CÓMO SE INTERPRETA?



Valores **pequeños** de χ^2 indican **buena concordancia** entre frecuencias observadas y esperadas.

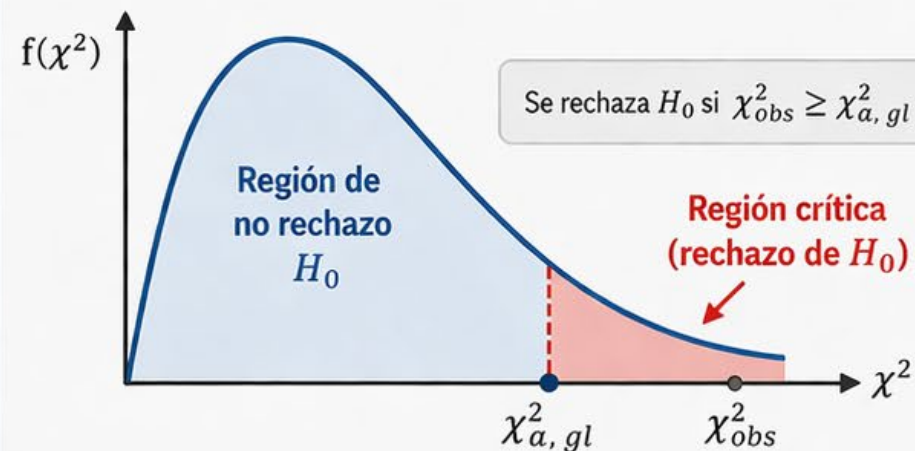


Valores **grandes** de χ^2 indican **discrepancias importantes**.



Se rechaza H_0 cuando el valor observado del estadístico cae en la **región crítica**.

LA PRUEBA χ^2 ES UN CONTRASTE DE COLA DERECHA



IDEA PRÁCTICA

En esta familia de contrastes, la evidencia contra H_0 aparece cuando χ^2 toma **valores altos**.



χ^2 **pequeño**
→ datos compatibles con H_0 .



χ^2 grande, pero **menor que el crítico** → aún no suficiente para rechazar H_0 .



χ^2 **en la región crítica** (valores altos) → se rechaza H_0 .

2.4 Condiciones de uso y cautelas



$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

- O_i : frecuencia observada en la categoría i .
- E_i : frecuencia esperada en la categoría i bajo H_0 .



PARA APLICAR CORRECTAMENTE LA PRUEBA χ^2 DEBE CUMPLIRSE QUE:



MUESTRA ALEATORIA

La muestra debe seleccionarse de forma aleatoria de la población.



CATEGORÍAS MUTUAMENTE EXCLUYENTES

Cada observación sólo puede pertenecer a una categoría.



UNA SOLA CATEGORÍA

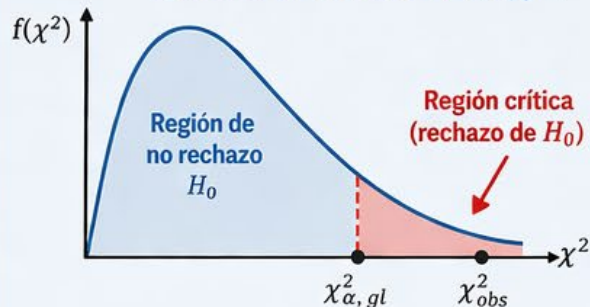
Cada observación contribuye a una sola categoría (nunca a varias).



FRECUENCIAS ESPERADAS NO PEQUEÑAS

Las frecuencias esperadas no deben ser demasiado pequeñas.

¿CÓMO SE INTERPRETA χ^2 ?



- Valores **pequeños** de χ^2 indican buena concordancia entre O_i y E_i .
- Valores **grandes** de χ^2 indican discrepancias importantes.
- **Se rechaza H_0** cuando χ^2_{obs} cae en la **región crítica** (cola derecha).



CAUTELA HABITUAL: CRITERIO PARA REAGRUPAR

Si alguna frecuencia esperada es **menor que 5**, conviene valorar el **reagrupamiento de categorías**.

Criterio práctico recomendado
 $E_i \geq 5$ en todas las categorías



Categoría reagrupada

Este criterio ayuda a que la aproximación asintótica de la prueba χ^2 sea fiable.



En resumen: el estadístico χ^2 mide cuánto se alejan las frecuencias observadas (O_i) de las esperadas (E_i) bajo H_0 .

Más concordancia ($O_i \approx E_i$) \rightarrow χ^2 pequeño.

Más discrepancia (O_i lejos de E_i) \rightarrow χ^2 grande.

2.5 Test rápido

1. **Si el valor observado de χ^2 es grande, eso sugiere que:**
 - a) las frecuencias observadas y esperadas son muy parecidas
 - b) la muestra es muy pequeña
 - c) hay discrepancias importantes entre frecuencias observadas y esperadas
 - d) H_0 es necesariamente verdadera
2. **En la prueba χ^2 , la región crítica está en:**
 - a) la cola izquierda
 - b) la cola derecha
 - c) las dos colas siempre
 - d) el centro de la distribución
3. **Si una frecuencia esperada es demasiado pequeña, una medida habitual es:**
 - a) aumentar el nivel de significación
 - b) cambiar la hipótesis alternativa
 - c) reagrupar categorías
 - d) redondear las frecuencias observadas

2.5 Test rápido

1. **Si el valor observado de χ^2 es grande, eso sugiere que:**
 - a) las frecuencias observadas y esperadas son muy parecidas
 - b) la muestra es muy pequeña
 - c) hay discrepancias importantes entre frecuencias observadas y esperadas
 - d) H_0 es necesariamente verdadera
2. **En la prueba χ^2 , la región crítica está en:**
 - a) la cola izquierda
 - b) la cola derecha
 - c) las dos colas siempre
 - d) el centro de la distribución
3. **Si una frecuencia esperada es demasiado pequeña, una medida habitual es:**
 - a) aumentar el nivel de significación
 - b) cambiar la hipótesis alternativa
 - c) reagrupar categorías
 - d) redondear las frecuencias observadas

3. Contrastes χ^2 de bondad del ajuste

3.1 ¿Qué contrasta la bondad del ajuste?



El **contraste** χ^2 de bondad del ajuste sirve para comprobar si la distribución observada en la muestra es **compatible con una distribución teórica** dada.

HIPÓTESIS HABITUALES



H_0 :

la población sigue la distribución especificada.

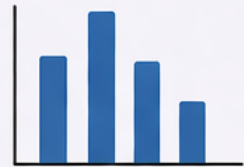


H_1 :

la población **no** sigue esa distribución.

¿QUÉ SE COMPARA?

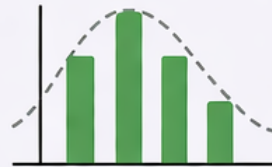
Se comparan las frecuencias observadas con las que cabría esperar si H_0 fuera cierta.



FRECUENCIAS OBSERVADAS (O_i)

Las que obtenemos en la muestra.

VS.



FRECUENCIAS ESPERADAS (E_i)

Las que se esperarían si H_0 fuera cierta.

¿CÓMO FUNCIONA?

1. Se fija una distribución teórica (por ejemplo, binomial, normal, Poisson, uniforme...).



2. Se calculan las frecuencias esperadas (E_i) en cada categoría según esa distribución.



3. Se observa la muestra y se obtienen las **frecuencias observadas** (O_i).



4. Se calcula el estadístico:

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \quad (k = \text{número de categorías})$$

5. Si χ^2 es grande (en la región crítica), **se rechaza H_0** : la muestra no sigue la distribución teórica especificada.

EJEMPLO ILUSTRATIVO

Supongamos que queremos comprobar si el número de caras al lanzar una moneda se ajusta a una distribución Binomial ($n=4, p=0,5$).



Nº de caras (i)	0	1	2	3	4	Total
Observadas (O_i)	3	8	14	9	6	40
Esperadas (E_i)	2,5	10	15	10	2,5	40



Se comparan O_i y E_i en cada categoría, se calcula χ^2 y se toma la decisión.

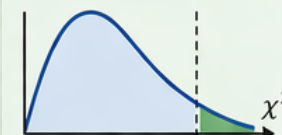


IDEA CLAVE:

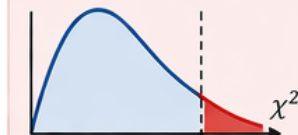
Si las diferencias entre frecuencias observadas y esperadas son pequeñas, los datos son **compatibles con H_0** .
Si las diferencias son grandes, los datos aportan **evidencia contra H_0** .



χ^2 pequeño
No se rechaza H_0
(buen ajuste)



χ^2 grande
Se rechaza H_0
(mal ajuste)



3.2 Tabla de trabajo para bondad del ajuste

Para organizar el cálculo suele construirse una tabla como esta:

Categoría i	Frecuencia observada O_i	Probabilidad teórica p_i	Frecuencia esperada $E_i = n \cdot p_i$	Contribución $c_i = \frac{(O_i - E_i)^2}{E_i}$
1	O_1	p_1	$E_1 = n \cdot p_1$	c_1
2	O_2	p_2	$E_2 = n \cdot p_2$	c_2
...				

Recuerda:

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

$$\chi^2 = \sum_{i=1}^n c_i$$

Utilidad de la tabla:

- reúne en un solo esquema **toda la información necesaria**;
- **permite calcular** de forma ordenada el **estadístico χ^2** ;
- **facilita la comprobación** de si las frecuencias esperadas son razonables.

3.3 Grados de libertad en bondad del ajuste

En el contraste χ^2 de bondad del ajuste, **los grados de libertad dependen del número de categorías (r) y de si se han estimado parámetros (k).**

$$gl = r - 1 - k$$

donde:

- r : número de categorías;
- k : número de parámetros estimados a partir de la muestra.

Nota docente:

- si no se estima ningún parámetro, $gl = r - 1$;
- si se estima algún parámetro, los grados de libertad disminuyen.

Recuerda:

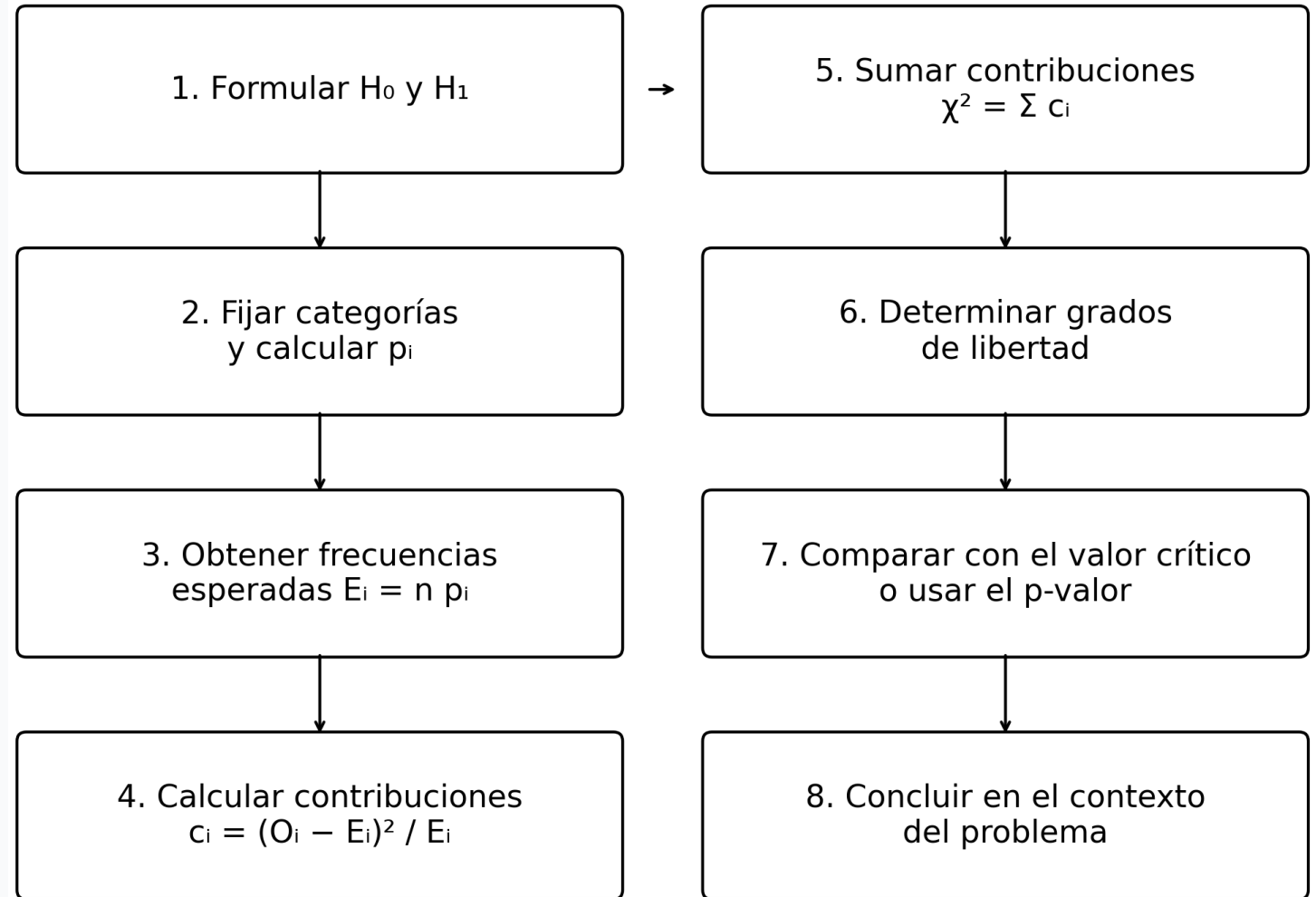
$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

$$\chi^2 = \sum_{i=1}^n c_i$$

3.4 Procedimiento paso a paso

Para aplicar el contraste χ^2 de bondad del ajuste seguimos estos pasos:

1. **Formular H_0 y H_1 .**
2. Fijar las **categorías** y calcular las **probabilidades teóricas p_i .**
3. Obtener las **frecuencias esperadas**
 $E_i = n \cdot p_i$
4. Calcular las **contribuciones $c_i = \frac{(O_i - E_i)^2}{E_i}$.**
5. **Sumar las contribuciones** para obtener el estadístico $\chi^2 = \sum_{i=1}^n c_i$
6. Determinar los **grados de libertad**
 $gl = r - 1 - k$.
7. **Comparar el valor observado con el valor crítico** o interpretar el *p - valor*.
8. **Concluir** en el contexto del problema



3.5 Ejemplo resuelto de bondad del ajuste

Una empresa clasifica 200 compras online según el día de la semana en que se realizaron. Se quiere contrastar si las compras se reparten uniformemente entre los 5 días laborables.

Datos observados:

Día	L	M	X	J	V
O_i	30	46	40	50	34

Hipótesis:

- $H_0: p_L = p_M = p_X = p_J = p_V = 0,2;$
- $H_1:$ la distribución no es uniforme.

Frecuencias esperadas:

$$E_i = 200 \cdot 0,2 = 40$$

para cada día.

Idea del cálculo:

Se comparan las frecuencias observadas con las esperadas y se obtiene el estadístico χ^2

3.6 Ejemplo resuelto de bondad del ajuste (cálculo)

Día i	Frecuencia observada O_i	Frecuencia esperada $E_i = n \cdot p_i$	$c_i = \frac{(O_i - E_i)^2}{E_i}$
L	30	40	2,5
M	46	40	0,9
X	40	40	0
J	50	40	2,5
V	34	40	0,9

Como hay $r = 5$ categorías y no se estima ningún parámetro:

$$gl = r - 1 = 4$$

Si trabajamos con nivel de significación $\alpha = 0,05$:

$$\chi^2_{1-\alpha,4} = \chi^2_{0,95;4} = 9,49$$

Como

$$\chi^2 < \chi^2_{0,95;4} \quad ; \quad 6,8 < 9,49$$

No se rechaza H_0

$$\chi^2 = \sum c_i = 2,5 + 0,9 + 0 + 2,5 + 0,9 = 6,8$$

3.7 Test rápido de bondad del ajuste

1. En un contraste χ^2 de bondad del ajuste, la hipótesis nula suele afirmar que:
 - a) la muestra no es aleatoria
 - b) la distribución observada coincide con una distribución teórica
 - c) todas las frecuencias observadas son iguales
 - d) el estadístico χ^2 es pequeño.
2. Si en una categoría se observa $O_i = 18$ y se espera $E_i = 20$, la contribución de esa categoría es:
 - a) 0,1
 - b) 2,0
 - c) 0,4
 - d) 0,2
3. En un contraste χ^2 de bondad del ajuste con $r = 6$ categorías, si no se estima ningún parámetro, los grados de libertad son:
 - a) 6
 - b) 5
 - c) 4
 - d) 3

3.7 Test rápido de bondad del ajuste

1. En un contraste χ^2 de bondad del ajuste, la hipótesis nula suele afirmar que:
 - a) la muestra no es aleatoria
 - b) la distribución observada coincide con una distribución teórica
 - c) todas las frecuencias observadas son iguales
 - d) el estadístico χ^2 es pequeño.
2. Si en una categoría se observa $O_i = 18$ y se espera $E_i = 20$, la contribución de esa categoría es:
 - a) 0,1
 - b) 2,0
 - c) 0,4
 - d) 0,2
3. En un contraste χ^2 de bondad del ajuste con $r = 6$ categorías, si no se estima ningún parámetro, los grados de libertad son:
 - a) 6
 - b) 5
 - c) 4
 - d) 3

4. Contrastes χ^2 de independencia e homogeneidad

4.1 ¿Qué significa independencia?

Dos variables aleatorias son independientes cuando **la distribución condicionada de una coincide con su distribución marginal, cualquiera que sea el valor o categoría fijado de la otra.**

En términos prácticos:

- conocer el valor de **una variable no aporta información sobre la otra** (no modifica la distribución);
- **si existe relación** entre ellas, hablamos de **asociación**;
- **el contraste χ^2 de independencia** permite estudiar esa **posible asociación en una tabla de contingencia.**

¿Existe relación entre el tipo de cliente y el canal preferido?



Tipo de cliente (X)



Canal preferido (Y)

	Web	Tienda	App
Joven	40	20	30
Adulto	25	35	20
Senior	10	30	10



Si las variables son **independientes**, los porcentajes por fila serían parecidos.

Si hay asociación, los porcentajes cambian de una fila a otra.




INDEPENDENCIA (NO HAY ASOCIACIÓN)

Conocer una variable **NO** aporta información sobre la otra.



→


La distribución de Y es la misma en cada categoría de X.


X = A




X = B



X = C




No se observan patrones en los datos.




ASOCIACIÓN (HAY DEPENDENCIA)

Conocer una variable **SÍ** aporta información sobre la otra.



→


La distribución de Y cambia según la categoría de X.


X = A




X = B



X = C



Se observan patrones: las categorías no se distribuyen igual.



4.2 Hipótesis del contraste de independencia

En una tabla de contingencia, el contraste χ^2 de independencia suele plantearse así:

H_0 : las dos variables son independientes

H_1 : las dos variables no son independientes

Interpretación:

- si no se rechaza H_0 , la muestra es compatible con independencia;
- si se rechaza H_0 , la muestra aporta evidencia de asociación entre las variables.

INDEPENDENCIA (NO HAY ASOCIACIÓN)	ASOCIACIÓN (HAY DEPENDENCIA)
<p>Conocer una variable NO aporta información sobre la otra.</p>  <p>La distribución de Y es la misma en cada categoría de X.</p> <div style="display: flex; justify-content: space-around;"> <div style="background-color: #d9ead3; padding: 5px; border: 1px solid #ccc;">X = A </div> <div style="background-color: #d9ead3; padding: 5px; border: 1px solid #ccc;">X = B </div> <div style="background-color: #d9ead3; padding: 5px; border: 1px solid #ccc;">X = C </div> </div> <p>No se observan patrones en los datos.</p> 	<p>Conocer una variable SÍ aporta información sobre la otra.</p>  <p>La distribución de Y cambia según la categoría de X.</p> <div style="display: flex; justify-content: space-around;"> <div style="background-color: #f4cccc; padding: 5px; border: 1px solid #ccc;">X = A </div> <div style="background-color: #f4cccc; padding: 5px; border: 1px solid #ccc;">X = B </div> <div style="background-color: #f4cccc; padding: 5px; border: 1px solid #ccc;">X = C </div> </div> <p>Se observan patrones: las categorías no se distribuyen igual.</p> 

4.3 Frecuencias esperadas en una tabla de contingencia

Si H_0 afirma independencia, la frecuencia esperada de la celda (i, j) es:

$$E_{ij} = \frac{n_{i.} \cdot n_{.j}}{n}$$

Donde:

- $n_{i.}$: total de la fila i ;
- $n_{.j}$: total de la columna j ;
- n : total de la muestra.

Idea clave:

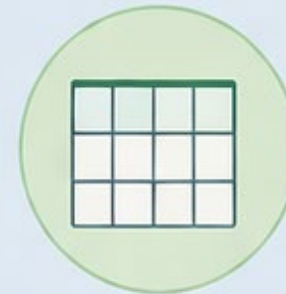
Si las variables fueran independientes, la frecuencia esperada en cada celda vendría dada por los pesos marginales de su fila y de su columna.

CONTRASTE CHI-CUADRADO DE INDEPENDENCIA

Se usa una tabla de contingencia para comprobar si las variables son independientes.



1. Se calculan las frecuencias esperadas bajo independencia.



2. Se compara lo observado con lo esperado.



3. Si la diferencia es grande, hay evidencia de asociación.

Idea clave: si conocer una variable no cambia la distribución de la otra, no hay asociación. Si la cambia, sí la hay.



4.4 Estadístico y grados de libertad

El estadístico del contraste es:

$$\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Donde:

- O_{ij} : frecuencia observada en la celda (i, j) ;
- E_{ij} : frecuencia esperada en la celda (i, j) bajo H_0 ;

Recuerda:

$$E_{ij} = \frac{n_{i.} \cdot n_{.j}}{n}$$

Si la tabla tiene f filas y c columnas, los grados de libertad son:

$$gl = (f - 1)(c - 1)$$

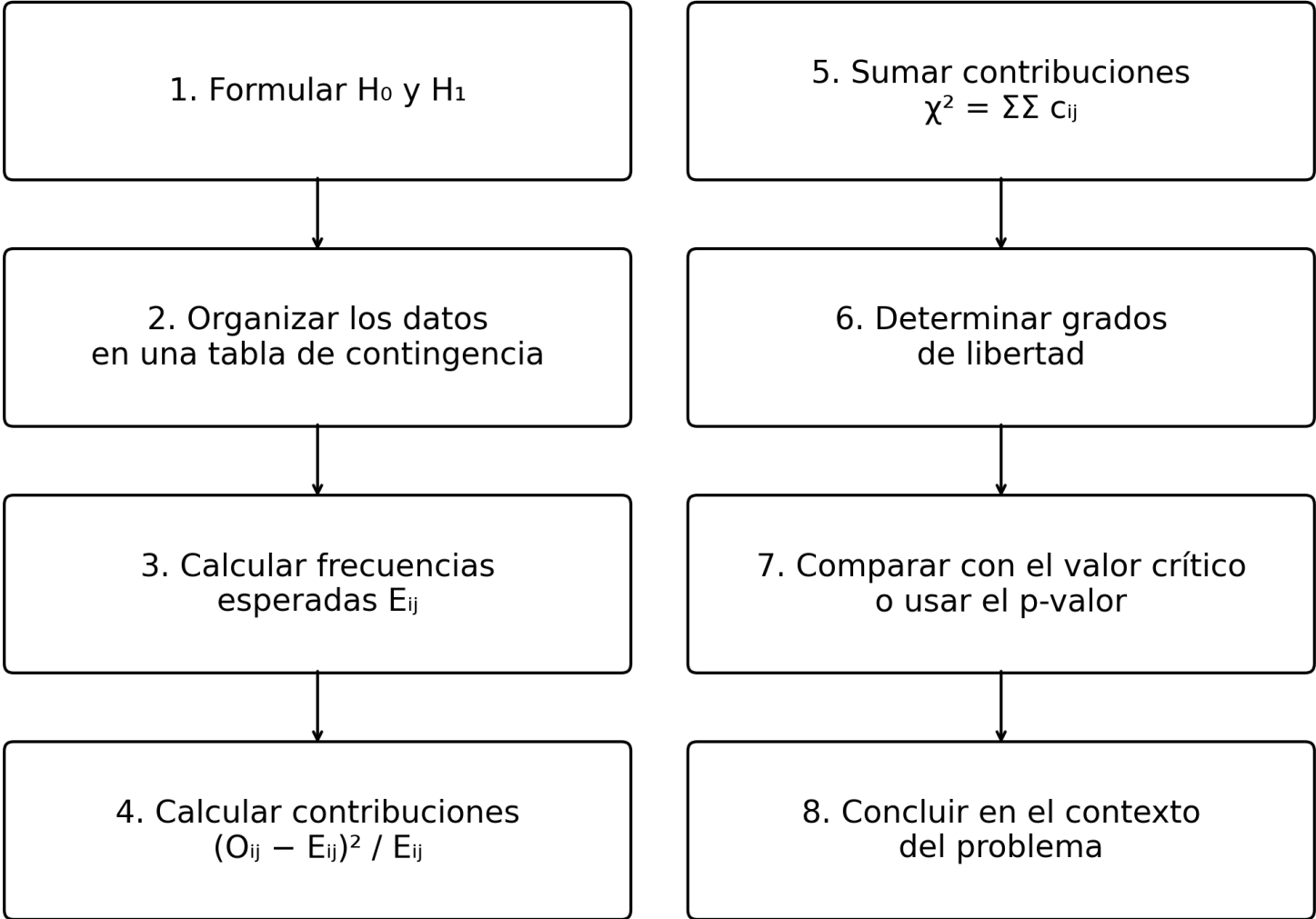
Interpretación básica::

- valores altos de χ^2 indican discrepancias importantes entre frecuencias observadas y esperadas;
- en ese caso, la evidencia contra H_0 es mayor.

4.5 Procedimiento paso a paso

Para aplicar el **contraste χ^2 de homogeneidad** seguimos estos pasos:

1. **Formular H_0 y H_1 .**
2. Organizar los datos en una **tabla de contingencia**.
3. Calcular las **frecuencias esperadas E_{ij}** .
4. Calcular las **contribuciones $c_i = \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$** de cada celda
5. **Sumar las contribuciones** para obtener el estadístico $\chi^2 = \sum_{i=1}^n c_i$
6. Determinar los **grados de libertad**
 $gl = (f - 1)(c - 1)$
7. **Comparar el valor observado con el valor crítico** o interpretar el *p - valor*.
8. **Concluir** en el contexto del problema



4.6 Ejemplo resuelto de independencia

Se estudia si existe relación entre el tipo de jornada de una persona trabajadora y su uso habitual del transporte público.

	Usa transporte público	No usa transporte público	Total
Jornada completa	36	24	60
Jornada parcial	14	26	40
Total	50	50	100

Hipótesis:

H_0 : tipo de jornada y uso del transporte público son independientes

H_1 : tipo de jornada y uso del transporte público no son independientes

Idea del cálculo:

Si H_0 fuera cierta, las frecuencias esperadas de cada celda vendrían dadas por los marginales de fila y de columna.

4.7 Ejemplo resuelto de independencia (cálculo)

	Usa transporte público	No usa transporte público	Total
Jornada completa	36	24	60
Jornada parcial	14	26	40
Total	50	50	100

Las frecuencias esperadas son:

$$E_{11} = \frac{60 \cdot 50}{100} = 30, \quad E_{12} = \frac{60 \cdot 50}{100} = 30,$$

$$E_{21} = \frac{40 \cdot 50}{100} = 20, \quad E_{22} = \frac{40 \cdot 50}{100} = 20$$

	O_{ij}	E_{ij}	$\frac{(O_{ij} - E_{ij})^2}{E_{ij}}$
Completa / Usa	36	30	1,20
Completa / No usa	24	30	1,20
Parcial / Usa	14	20	1,80
Parcial / No usa	26	20	1,80

$$\chi^2 = 1,20 + 1,20 + 1,8 + 1,8 = 6,00$$

Como la tabla es 2x2,

$$gl = (2 - 1)(2 - 1) = 1$$

Con nivel de significación $\alpha = 0,05$, el valor crítico es

$$\chi^2_{1-\alpha;1} = \chi^2_{0,95;1} = 3,84$$

Como

$$\chi^2 > \chi^2_{0,95;1} ; 6,00 > 3,84 \quad \text{se rechaza } H_0.$$

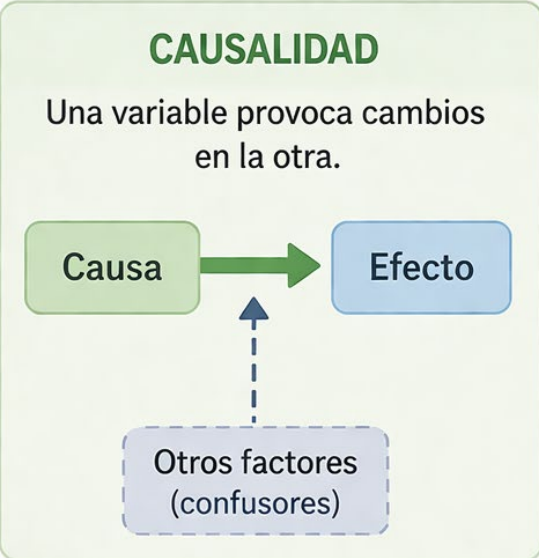
Conclusión: La muestra aporta evidencia de asociación entre el tipo de jornada y el uso habitual del transporte público.

4.8 Interpretación económica correcta

El contraste χ^2 de independencia permite detectar **relación estadística, no demostrar una relación causal**

RELACIÓN ESTADÍSTICA
Las variables varían juntas más de lo que cabría por azar.

	B1	B2
A1	30	20
A2	15	35



Rechazar la independencia significa que la muestra aporta evidencia de asociación entre las variables.

Pero **atención**:

- asociación no significa causalidad;
- el contraste no mide la intensidad de la relación;
- la interpretación debe hacerse en el contexto del problema.

IDEA CLAVE
El χ^2 de independencia solo dice si hay asociación. No explica por qué existe.

EJEMPLO
Puede observarse que quienes hacen más ejercicio tienen mejor salud (hay asociación), pero la causa puede ser otra: dieta, edad, ingresos, etc.

Asociación ≠ Causalidad.
Para hablar de causa y efecto se necesitan argumentos adicionales y un buen diseño de estudio.

4.9 Test rápido de independencia

1. **En un contraste χ^2 de independencia, la hipótesis nula afirma que:**
 - a) las frecuencias observadas y esperadas son iguales en todas las celdas
 - b) las dos variables son independientes
 - c) las dos variables tienen la misma media
 - d) no hay error muestral.
2. **Si H_0 es cierta, la frecuencia esperada de una celda se obtiene a partir de:**
 - a) el total de la fila; el total de la columna; y el total muestral
 - b) solo el total de la fila
 - c) solo el total de la columna
 - d) la media de las cuatro celdas
3. **En el contraste χ^2 de independencia, rechazar H_0 significa que:**
 - a) las variables son necesariamente causales
 - b) las variables son idénticas
 - c) la muestra aporta evidencia de asociación entre las variables
 - d) las frecuencias esperadas eran incorrectas.

4.9 Test rápido de independencia

1. **En un contraste χ^2 de independencia, la hipótesis nula afirma que:**
 - a) las frecuencias observadas y esperadas son iguales en todas las celdas
 - b) las dos variables son independientes
 - c) las dos variables tienen la misma media
 - d) no hay error muestral.
2. **Si H_0 es cierta, la frecuencia esperada de una celda se obtiene a partir de:**
 - a) el total de la fila; el total de la columna; y el total muestral
 - b) solo el total de la fila
 - c) solo el total de la columna
 - d) la media de las cuatro celdas
3. **En el contraste χ^2 de independencia, rechazar H_0 significa que:**
 - a) las variables son necesariamente causales
 - b) las variables son idénticas
 - c) la muestra aporta evidencia de asociación entre las variables
 - d) las frecuencias esperadas eran incorrectas.

5. Contrastes χ^2 de homogeneidad

5.1 ¿Qué contrasta la homogeneidad?

El contraste χ^2 de HOMOGENEIDAD



Permite comprobar si varias poblaciones presentan la **misma distribución** respecto de una **variable cualitativa**.



EJEMPLO Una empresa quiere saber si la preferencia de canal de compra es la misma en tres ciudades.

Variable cualitativa:
Canal preferido
(Tienda, Web, App)



¿La distribución de preferencias es la misma en las tres ciudades?

Canal	Ciudad A		Ciudad B		Ciudad C		Total
	Observado	%	Observado	%	Observado	%	
Tienda	120	40%	90	30%	60	20%	270
Web	100	33%	120	40%	80	26%	300
App	80	27%	90	30%	160	54%	330
Total	300	100%	300	100%	300	100%	900

El contraste χ^2 de homogeneidad analiza si las diferencias observadas entre ciudades son **mayores de lo que cabría esperar por azar**.

Hipótesis habituales:

H_0 : las poblaciones son homogéneas

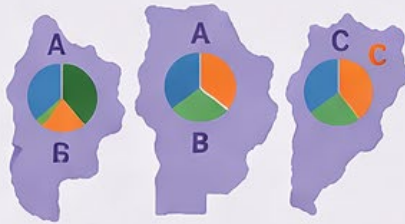
H_1 : al menos una distribución difiere

Idea clave: Se compara si las proporciones observadas en las distintas poblaciones son **compatibles con una distribución común**.

HIPÓTESIS

H_0 : Todas las poblaciones tienen la misma distribución.

H_1 : Al menos una población tiene una distribución diferente.



IDEA

Si H_0 es cierta, los porcentajes de cada categoría deben ser similares en todas las poblaciones.



PROCEDIMIENTO

1. Se calculan las frecuencias esperadas bajo H_0 .
2. Se compara lo observado con lo esperado mediante el estadístico χ^2 .
3. Si χ^2 es grande (p -valor pequeño), se rechaza H_0 .



IDEA CLAVE: Si rechazamos H_0 , concluimos que las poblaciones no son homogéneas: al menos una tiene una distribución distinta.

5.2 Relación con independencia

El contraste χ^2 de **homogeneidad** y el contraste χ^2 de **independencia** comparten la misma forma del estadístico:

$$\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

y también la misma expresión para los grados de libertad:

$$gl = (f - 1)(c - 1)$$

Pero **no responden a la misma pregunta:**

- en **independencia**, se estudia la relación entre **dos variables** en una **misma población**;
- en **homogeneidad**, se compara **si varias poblaciones** tienen la misma distribución respecto de **una variable**.

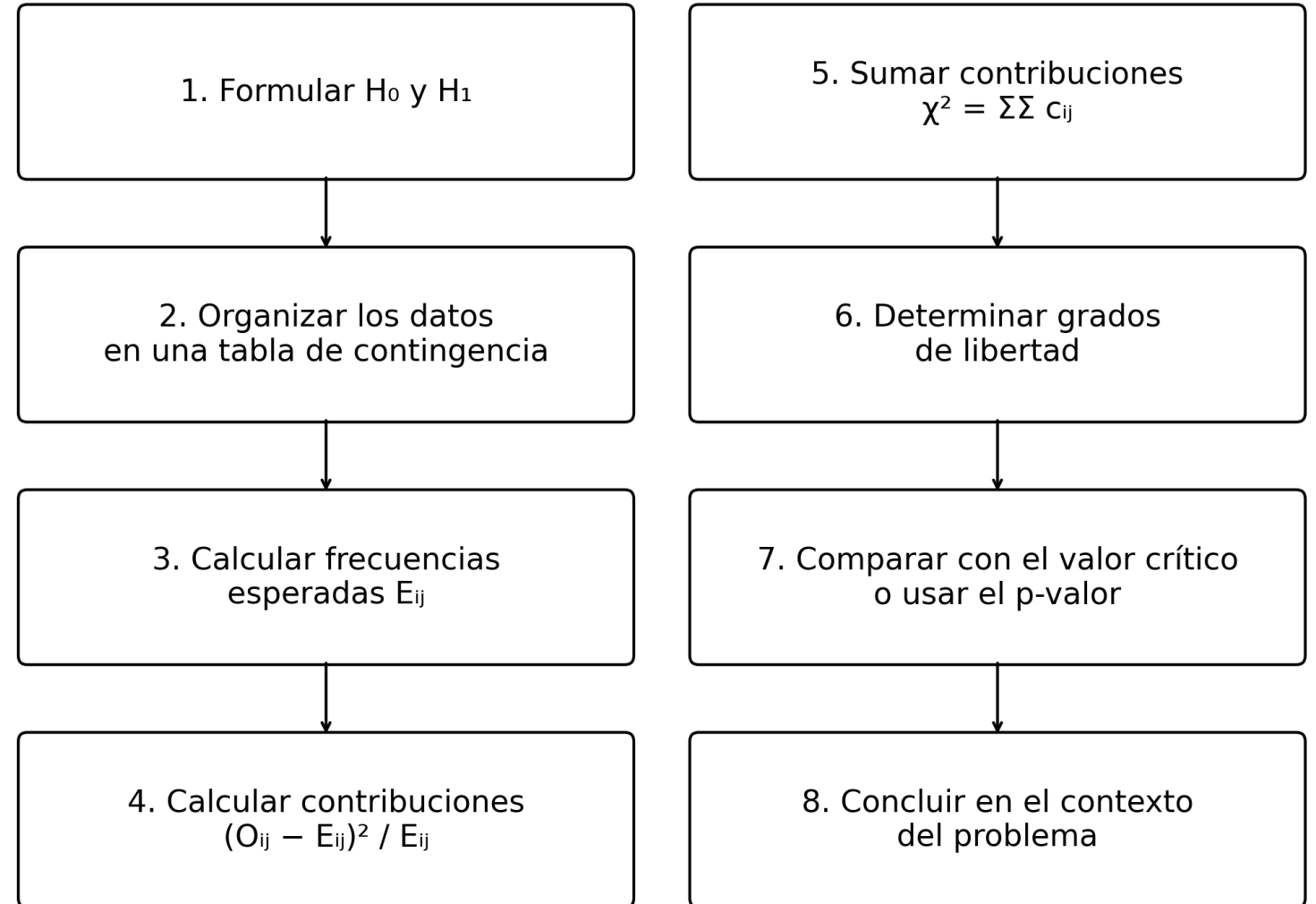
No responden a la misma pregunta:



5.3 Procedimiento y grados de libertad

Para aplicar el **contraste χ^2 de homogeneidad** seguimos estos pasos:

1. **Formular H_0 y H_1 .**
2. Organizar los datos en una **tabla de contingencia**.
3. Calcular las **frecuencias esperadas E_{ij}** .
4. Calcular las **contribuciones $c_i = \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$** de cada celda
5. **Sumar las contribuciones** para obtener el estadístico $\chi^2 = \sum_{i=1}^n c_i$
6. Determinar los **grados de libertad**
 $gl = (f - 1)(c - 1)$
7. **Comparar el valor observado con el valor crítico** o interpretar el *p - valor*.
8. **Concluir** en el contexto del problema



5.4 Ejemplo resuelto de homogeneidad

Se comparan tres grupos de consumidores según el canal en el que realizaron su última compra: tienda física, web o app.

Los datos observados son:

Canal de compra	Jovenes	Adultos	Mayores	Total
Tienda física	18	30	42	90
Web	20	24	12	56
App	22	16	6	44
Total	60	70	60	190

Hipótesis:

H_0 : las tres poblaciones son homogéneas respecto del canal de compra

H_1 : al menos una distribución difiere

Idea del cálculo:

Si H_0 fuera cierta, las frecuencias esperadas de cada celda vendrían dadas por los marginales de fila y de columna.

5.5 Ejemplo resuelto de homogeneidad (cálculo)

Canal de compra	Jovenes	Adultos	Mayores	Total
Tienda física	18	30	42	90
Web	20	24	12	56
App	22	16	6	44
Total	60	70	60	190

Las frecuencias esperadas son:

$$E_{11} = \frac{90 \cdot 60}{190} = 28,42 \quad E_{12} = \frac{90 \cdot 70}{190} = 33,16 \quad E_{13} = \frac{90 \cdot 60}{190} = 28,42$$

$$E_{21} = \frac{56 \cdot 60}{190} = 17,68 \quad E_{22} = \frac{56 \cdot 70}{190} = 20,63 \quad E_{23} = \frac{56 \cdot 60}{190} = 17,68$$

$$E_{31} = \frac{44 \cdot 60}{190} = 13,89 \quad E_{32} = \frac{44 \cdot 70}{190} = 16,21 \quad E_{33} = \frac{44 \cdot 60}{190} = 13,89$$

Las **contribuciones** de las nueve celdas son, aproximadamente:

$$c_i = \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

3,82; 0,30; 6,49 ; 0,30 ; 0,55; 1,82 ; 4,74 ; 0,00 ; 4,48

Por tanto:

$$\chi^2 \approx 22,50$$

Como la tabla es 3×3 ,

$$gl = (3 - 1)(3 - 1) = 4$$

Con nivel de significación $\alpha = 0,05$, el valor crítico es

$$\chi^2_{1-\alpha;4} = \chi^2_{0,95;4} = 9,49$$

Como

$$\chi^2 > \chi^2_{0,95;4} ; 22,50 > 9,49 \quad \text{se rechaza } H_0.$$

Conclusión: La muestra aporta evidencia de que la distribución del canal de compra no es la misma en los tres grupos.

5.6 Test rápido de homogeneidad

1. **En un contraste χ^2 de homogeneidad, la hipótesis nula afirma que:**
 - a) las variables son independientes
 - b) las poblaciones tienen la misma media
 - c) las poblaciones presentan la misma distribución respecto de una variable cualitativa
 - d) todas las frecuencias observadas son iguales.
2. **En un contraste χ^2 de homogeneidad, las frecuencias esperadas se calculan:**
 - a) solo con los totales de fila
 - b) solo con los totales de columna
 - c) a partir de los marginales de fila y de columna
 - d) a partir de la media de las celdas .
3. **Rechazar H_0 en un contraste χ^2 de homogeneidad significa que:**
 - a) todas las poblaciones son idénticas
 - b) al menos una distribución difiere
 - c) las variables son causales
 - d) la muestra no es aleatoria.

5.6 Test rápido de homogeneidad

1. En un contraste χ^2 de homogeneidad, la hipótesis nula afirma que:
 - a) las variables son independientes
 - b) las poblaciones tienen la misma media
 - c) las poblaciones presentan la misma distribución respecto de una variable cualitativa
 - d) todas las frecuencias observadas son iguales.
2. En un contraste χ^2 de homogeneidad, las frecuencias esperadas se calculan:
 - a) solo con los totales de fila
 - b) solo con los totales de columna
 - c) a partir de los marginales de fila y de columna
 - d) a partir de la media de las celdas .
3. Rechazar H_0 en un contraste χ^2 de homogeneidad significa que:
 - a) todas las poblaciones son idénticas
 - b) al menos una distribución difiere
 - c) las variables son causales
 - d) la muestra no es aleatoria.

6. Contraste de Kolmogorov–Smirnov en una muestra

6.1 Idea y utilidad



El contraste de Kolmogorov–Smirnov en una muestra permite comparar la **distribución empírica** observada en la muestra con una **distribución teórica** dada $F(x)$.

HIPÓTESIS HABITUALES



H_0 :

La muestra procede de una población con distribución $F(x)$.

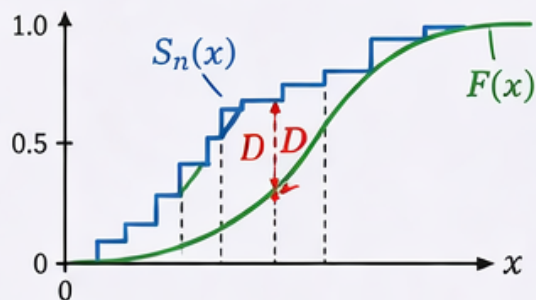


H_1 :

La muestra **no** procede de esa población.

¿QUÉ SE COMPARA?

Se compara la función de distribución empírica de la muestra $S_n(x)$ con la distribución teórica $F(x)$.



$$D = \sup_x |S_n(x) - F(x)|$$

D es la mayor discrepancia vertical entre ambas distribuciones.



IDEA CLAVE

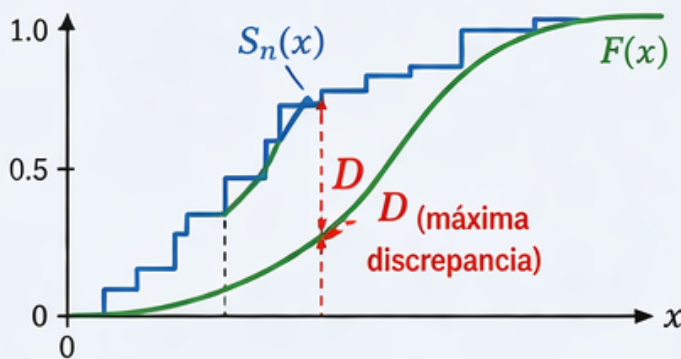
El contraste se basa en la mayor discrepancia entre la **distribución empírica muestral** y la distribución teórica.

Cuanto mayor sea esa discrepancia, más evidencia habrá contra H_0 .

¿CÓMO FUNCIONA EL CONTRASTE?

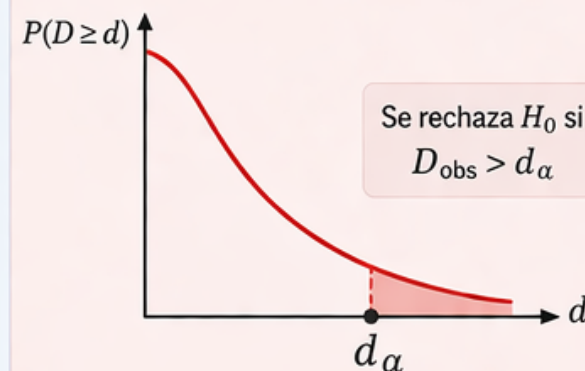
- 1 Se plantea H_0 : la muestra $\sim F(x)$.
- 2 Se calcula la función de distribución empírica $S_n(x)$.
- 3 Se calcula el estadístico:
$$D = \sup_x |S_n(x) - F(x)|$$
- 4 Se obtiene el valor p (o se compara con el valor crítico).
- 5 Si D es grande (p pequeño), se rechaza H_0 .

DISTRIBUCIÓN EMPÍRICA vs. TEÓRICA



$S_n(x)$: empírica
 $F(x)$: teórica

REGIÓN CRÍTICA (COLA DERECHA)



6.2 Estadístico D y regla de decisión



INTERPRETACIÓN

Si $S_n(x)$ está muy próxima a $F(x)$ en todo el rango, D será pequeño \rightarrow no se rechaza H_0 .

Si hay grandes diferencias en algún punto, D será grande \rightarrow se rechaza H_0 .



IDEA PRÁCTICA

El contraste de Kolmogorov–Smirnov no depende de cómo se agrupen los datos en clases: usa toda la información de la muestra a través de $S_n(x)$. Es especialmente útil con muestras pequeñas o cuando no se quieren suponer posibles puntos de corte.



EN RESUMEN

Compara $S_n(x)$ con $F(x)$ y se fija en la mayor diferencia vertical.

D pequeño \rightarrow compatible con H_0 .

D grande \rightarrow evidencia contra H_0 .

El estadístico del contraste es:

$$D_n = \sup_x |S_n(x) - F(x)|$$

Donde:

- $S_n(x)$: función de distribución empírica de la muestra;
- $F(x)$: función de distribución teórica bajo H_0 ;
- \sup_x : de todas las discrepancias posibles, tomamos la más grande.

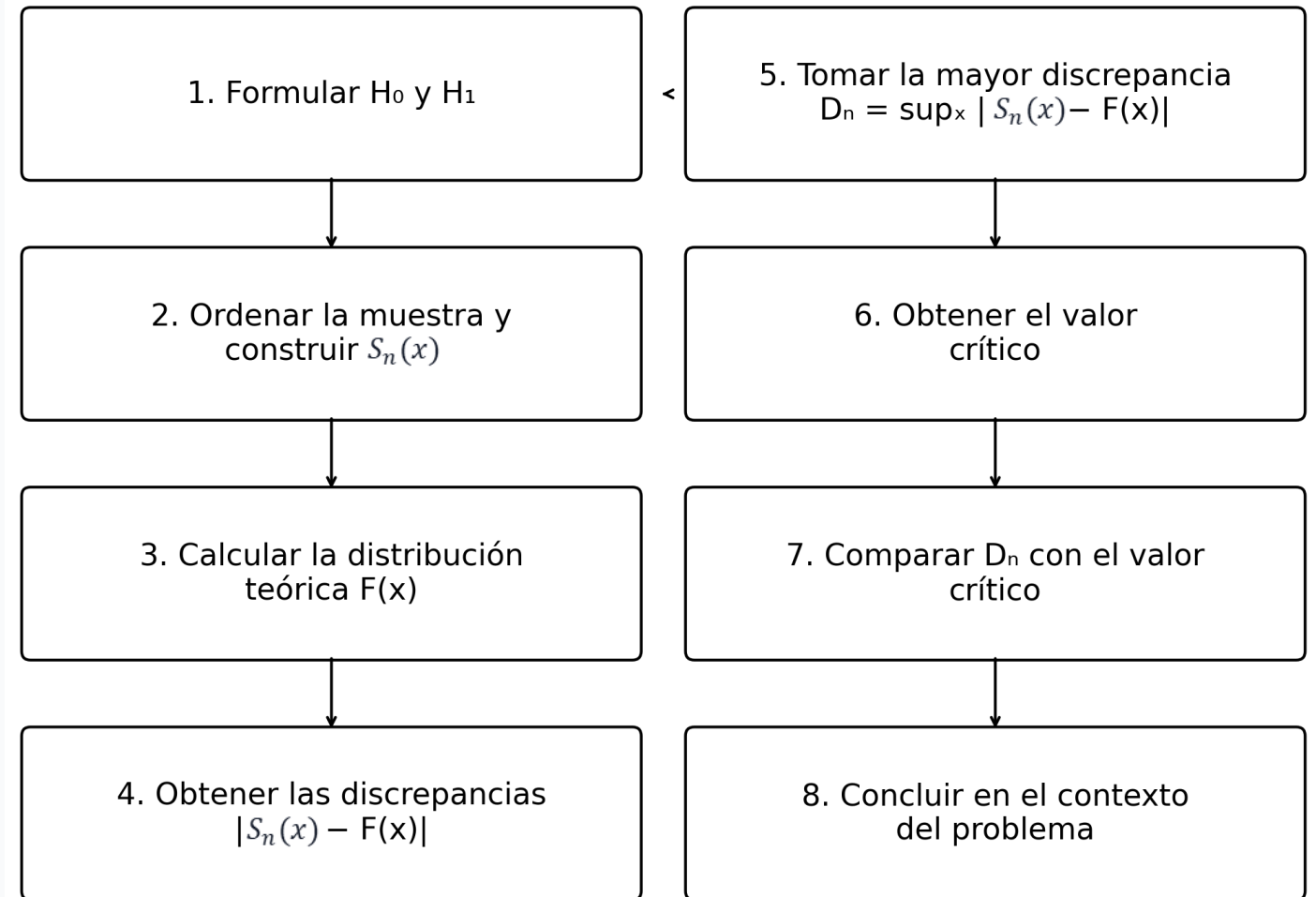
Interpretación básica:

- para cada x , se compara la diferencia $|F_n(x) - F(x)|$;
- el estadístico D_n se queda con la mayor de cada una de esas diferencias;
- **cuanto mayor sea D_n , mayor será la discordancia** entre la muestra y la distribución teórica.
- Se rechaza H_0 cuando D_n supera el valor crítico correspondiente

6.3 Procedimiento paso a paso

Para aplicar el **contraste de Kolmogorov–Smirnov en una muestra** seguimos estos pasos:

1. **Formular** H_0 y H_1 .
2. **Ordenar** la muestra y **construir** la función de **distribución empírica** $S_n(x)$.
3. Calcular la función de **distribución teórica** $F(x)$ bajo H_0 .
4. Obtener las **discrepancias** $|S_n(x) - F(x)|$.
5. Tomar la **mayor discrepancia** para obtener D_n .
6. Comparar D_n con el **valor crítico** correspondiente.
7. Tomar la **decisión**.
8. **Concluir** en el contexto del problema.



6.4 Ejemplo resuelto de Kolmogorov–Smirnov

Se quiere contrastar, con un nivel de significación $\alpha = 0,02$, si una muestra de tamaño $n = 50$ puede proceder de una población continua con función de distribución:

$$F(x) = x^2 \quad ; \quad 0 \leq x \leq 1$$

La muestra, agrupada en intervalos, ha dado las siguientes frecuencias:

Intervalo	[0; 0,2)	[0,2; 0,4)	[0,4; 0,6)	[0,6; 0,8)	[0,8; 1]
n_i	1	7	10	18	14

Hipótesis:

H_0 : la muestra procede de una población con función de distribución $F(x) = x^2$

H_1 : la muestra no procede de esa población

Idea del cálculo:

Se compara la distribución empírica acumulada de la muestra con la distribución teórica y se toma la mayor discrepancia.

6.5 Ejemplo resuelto de Kolmogorov–Smirnov (cálculo)

Intervalo	[0; 0,2)	[0,2; 0,4)	[0,4; 0,6)	[0,6; 0,8)	[0,8; 1]
n_i	1	7	10	18	14

A partir de las frecuencias observadas, las frecuencias acumuladas son

$$N_i = 1; 8; 18; 36; 50$$

y, por tanto, la distribución empírica acumulada es:

$$F_n(x) = \frac{N_i}{50} = \{0,02; 0,16; 0,36; 0,72; 1,00\}$$

En los extremos derechos de los intervalos, la distribución teórica vale

$$F(0,2) = 0,04; F(0,4) = 0,16; F(0,6) = 0,36; F(0,8) = 0,64; F(1) = 1$$

Extremo derecho	0,2	0,4	0,6	0,8	1
$S_n(x)$	0,02	0,16	0,36	0,72	1
$F(x)$	0,04	0,16	0,36	0,64	1
$ F_n(x) - F(x) $	0,02	0	0	0,08	0

Por tanto,

$$D_n = \sup_x |S_n(x) - F(x)| = 0,08$$

Con nivel de significación $\alpha = 0,02$, en la tabla KS aparece $c_\alpha = c_{0,02} = 1,517$

Luego, el valor crítico es

$$K_{\alpha;n} \approx \frac{c_\alpha}{\sqrt{n}} = \frac{1,517}{\sqrt{50}} \approx \frac{1,517}{7,07} \approx 0,214$$

Como

$$D_n < K_{\alpha;n}; 0,08 < 0,214 \quad \text{no se rechaza } H_0.$$

Conclusión: La muestra es compatible con una población cuya función de distribución es $F(x) = x^2$.

6.6 Test rápido de Kolmogorov–Smirnov

1. **En el contraste de Kolmogorov–Smirnov de una muestra, la hipótesis nula suele afirmar que:**
 - a) la muestra sigue una distribución teórica dada
 - b) la media poblacional es igual a un valor concreto
 - c) las variables son independientes
 - d) las poblaciones son homogéneas.
2. **El estadístico D_n representa:**
 - a) la suma de todas las discrepancias entre $S_n(x)$ y $F(x)$;
 - b) la mayor discrepancia entre $S_n(x)$ y $F(x)$;
 - c) la media de las discrepancias entre $S_n(x)$ y $F(x)$;
 - d) la diferencia entre dos medias muestrales.
3. **Rechazar H_0 en un contraste χ^2 de homogeneidad significa que:**
 - a) D_n es pequeño
 - b) D_n coincide con cero
 - c) D_n supera el valor crítico correspondiente
 - d) la muestra tiene muchos intervalos.

6.6 Test rápido de Kolmogorov–Smirnov

1. **En el contraste de Kolmogorov–Smirnov de una muestra, la hipótesis nula suele afirmar que:**
 - a) la muestra sigue una distribución teórica dada ✓
 - b) la media poblacional es igual a un valor concreto
 - c) las variables son independientes
 - d) las poblaciones son homogéneas.
2. **El estadístico D_n representa:**
 - a) la suma de todas las discrepancias entre $S_n(x)$ y $F(x)$;
 - b) la mayor discrepancia entre $S_n(x)$ y $F(x)$; ✓
 - c) la media de las discrepancias entre $S_n(x)$ y $F(x)$;
 - d) la diferencia entre dos medias muestrales.
3. **Rechazar H_0 en un contraste χ^2 de homogeneidad significa que:**
 - a) D_n es pequeño
 - b) D_n coincide con cero
 - c) D_n supera el valor crítico correspondiente ✓
 - d) la muestra tiene muchos intervalos.

7. Cierre: mapa de decisión y errores típicos

7.1 Mapa de decisión

Objetivo	Contraste recomendado
Comprobar si una muestra es compatible con una distribución teórica dada	Bondad del ajuste
Comprobar si una muestra es compatible con una distribución continua comparando funciones de distribución	Kolmogorov–Smirnov
Estudiar la relación entre dos variables cualitativas en una misma población	χ^2 de independencia
Comparar varias poblaciones respecto de una variable cualitativa	χ^2 de homogeneidad

Idea clave:

La elección del contraste depende del tipo de pregunta y de la estructura de los datos.

7.2 Errores típicos

Errores frecuentes en este tema:

- confundir **independencia** con **homogeneidad**;
- interpretar **asociación como causalidad**;
- calcular mal las frecuencias esperadas;
- usar mal los grados de libertad;
- olvidar que en χ^2 la región crítica está en la **cola derecha**;
- **confundir** en KS la mayor discrepancia con una **suma de discrepancias**.

Idea clave:

En muchos errores no falla el cálculo; falla la identificación correcta del contraste o la interpretación final.

Recursos

Bibliografía:

- Ruiz-Maya, L., Martín-Pliego López, F. J. (3.ª ed.). **Fundamentos de Inferencia Estadística. Thompson–Paraninfo.**
- **Material de apoyo del curso:**
 - Tablas de probabilidad y cuantiles críticos elaboradas para la asignatura. **Atribuciones y transparencia:**
- **Gráficas e imágenes** de estas diapositivas: imágenes generadas con ChatGPT y gráficas con Python (SciPy/Matplotlib) y apoyo de ChatGPT.
- **Imágenes externas:** cuando se incluyen, se usan bajo licencia CC-BY / CC-BY-SA, citadas individualmente.