

La distribución de χ^2

- Introducción.
- Ejemplo.
- Definición general de χ^2 .
- Grados de libertad. χ^2 reducido.
- La distribución de χ^2 .
- Probabilidades de χ^2 .
- Ejemplos:
 1. Distribución de Poisson.
 2. Bondad de un ajuste.

Introducción

Dado un conjunto de medidas cabe plantearse:

- ¿Son compatibles con la distribución límite esperada?
- ¿Son compatibles los datos con la función teórica a la que se ajustan?

¿Cómo decidir si nuestras medidas son consistentes con lo que esperamos?

 Test de χ^2

Ejemplo

Sean las siguientes medidas del alcance de un proyectil:

731	772	771	681	722	688	653	757	733	742
739	780	709	676	760	748	672	687	766	645
678	748	689	810	805	778	764	753	709	675
698	770	754	830	725	710	738	638	787	712

Con valor medio y desviación típica dadas por:

$$\bar{x} = \frac{\sum x_i}{N} = 730.1 \quad \sigma_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N-1}} = 46.8$$

¿Vienen gobernadas por una distribución gaussiana con media

$X = \bar{x}$ y desviación típica $\sigma = \sigma_x$?

$$G_{X,\sigma}(x) = G_{\bar{x},\sigma_x}(x)$$

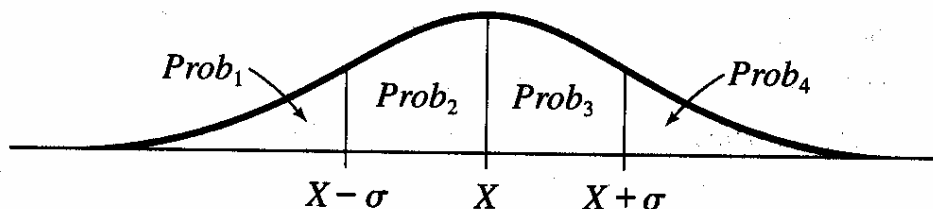
Comparación entre:

El resultado observado realmente:

Número de bin k	Valores de x en el bin		Observaciones O_k
1	$x < X - \sigma$	$(x < 683.3)$	8
2	$X - \sigma < x < X$	$(683.3 < x < 730.1)$	10
3	$X < x < X + \sigma$	$(730.1 < x < 776.9)$	16
4	$X + \sigma < x$	$(776.9 < x)$	6

con el resultado esperado si la hipótesis es correcta:

Número de bin k	Probabilidad $Prob_k$	Sucesos esperados E_k	Sucesos observados O_k
1	15.87 %	6.4	8
2	34.13 %	13.6	10
3	34.13 %	13.6	16
4	15.87 %	6.4	6



Conclusión:

- Si las diferencias $E_k - O_k$ son pequeñas, la hipótesis debe ser correcta.
- Si las diferencias $E_k - O_k$ son grandes, la hipótesis debe ser falsa.

Sin embargo:

¿Hasta qué punto pueden ser grandes las diferencias $E_k - O_k$?

Las diferencias $E_k - O_k$ han de ser del tamaño de las fluctuaciones de E_k , en este caso:

$$|E_k - O_k| \approx \sqrt{E_k}$$

Para evitar cancelaciones elevamos al cuadrado y sumamos para todos los bins $k = 1, \dots, n$ ($n = 4$):

$$\chi^2 = \sum_{k=1}^n \frac{(E_k - O_k)^2}{E_k}$$

En general:

Si $\chi^2 \leq n$ (χ^2 menor o similar a n)	Los valores esperados y observados concuerdan. Hipótesis compatible.
Si $\chi^2 \gg n$ (χ^2 mucho mayor que n)	Los valores observados no concuerdan con los esperados. Hipótesis falsa.

Ejemplo:

$$\begin{aligned} \chi^2 &= \sum_{k=1}^4 \frac{(E_k - O_k)^2}{E_k} = \\ &= \frac{(1.6)^2}{6.4} + \frac{(-3.6)^2}{13.6} + \frac{(2.4)^2}{13.6} + \frac{(1.6)^2}{6.4} = 1.8 \end{aligned}$$

Resultado compatible con la hipótesis

Definición general de χ^2

Estimación del acuerdo entre lo observado y lo esperado:

$$\chi^2 = \sum_1^n \left(\frac{\text{Valor observado} - \text{Valor esperado}}{\text{Desviación estándar}} \right)^2$$

- Si el acuerdo es bueno χ^2 es del orden de n .
- Si el acuerdo es malo χ^2 es mucho mayor que n .

Distribuciones límites.- Comparamos k valores observados O_k con k valores esperados E_k cuyas desviaciones típicas son $\sqrt{E_k}$:

$$\chi^2 = \sum_{k=1}^n \frac{(E_k - O_k)^2}{E_k}$$

Ajuste de funciones .- Comparamos i valores medidos y_i con i valores esperados $f(x_i)$ cuyas desviaciones típicas son σ_i :

$$\chi^2 = \sum_{i=1}^n \left(\frac{y_i - f(x_i)}{\sigma_i} \right)^2$$

Grados de libertad. χ^2 reducido.

Lo correcto no es comparar χ^2 con n , sino con el número de grados de libertad ν :

Grados de libertad (ν).- Se define como el número de datos observados, menos el número de parámetros calculados a partir de los datos (ligaduras) y utilizados en los cálculos.

$$\nu = n - l$$

Ejemplos:

a) Test de χ^2 para distribuciones gaussianas:

Ligaduras:

$$N = \sum O_k \quad ; \quad \bar{x} = \frac{\sum x_i}{N} \quad ; \quad \sigma_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N-1}} \quad ; \quad (l=3)$$

$$\text{Grados de libertad} \Rightarrow \nu = n - 3$$

b) Test de χ^2 para el ajuste a una recta ($y=ax+b$):

Ligaduras: a, b ($l=2$)

$$\text{Grados de libertad} \Rightarrow \nu = n - 2$$

Grados de libertad. χ^2 reducido.

El valor que esperamos al realizar el test de χ^2 será:

$$(\text{Valor esperado en promedio de } \chi^2) = \nu$$

Se define entonces el chi-cuadrado reducido como, el valor del chi-cuadrado dividido por los grados de libertad:

$$\bar{\chi}^2 = \chi^2 / \nu$$

cuyo valor esperado será entonces:

$$(\text{Valor esperado en promedio de } \bar{\chi}^2) = 1$$

Ejemplo (test de distribución gaussiana):

$$\text{Chi-cuadrado} \rightarrow \chi^2 = 1.80$$

$$\text{Ligaduras} \rightarrow l = 3$$

$$\text{Grados de libertad} \rightarrow \nu = n - l = 4 - 3 = 1$$

$$\text{Chi-cuadrado reducido} \rightarrow \bar{\chi}^2 = \chi^2 / \nu = 1.80$$

¿Cuán diferente de 1 ha de ser $\bar{\chi}^2$ para que podamos desechar la hipótesis de que es una gaussiana?

La distribución de χ^2

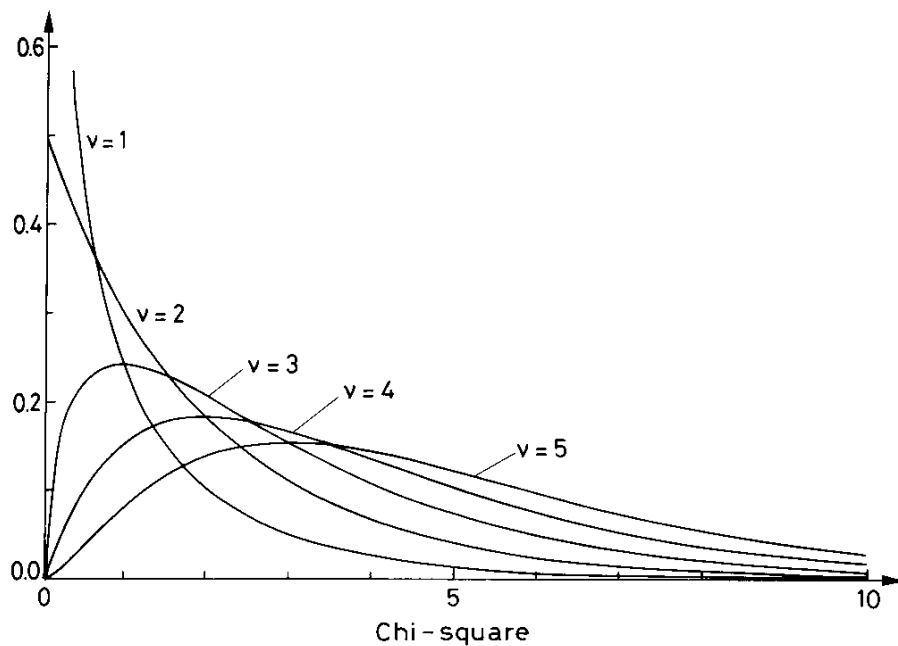
Supongamos que tenemos n variables aleatorias $y_1, y_2, \dots, y_n$ cuyas distribuciones son gaussianas con medias $\mu_1, \mu_2, \dots, \mu_n$ y con desviaciones típicas $\sigma_1, \sigma_2, \dots, \sigma_n$. Si calculamos entonces:

$$\chi^2 = \sum_{i=1}^n \left(\frac{y_i - \mu_i}{\sigma_i} \right)^2$$

¿Cuál es la probabilidad de obtener un valor entre χ^2 y $\chi^2 + d\chi^2$?

La función densidad de probabilidad de χ^2 o la distribución de χ^2 viene dada por:

$$p_\nu(\chi^2) = \frac{e^{-\chi^2/2} (\chi^2)^{\nu/2-1}}{2^{\nu/2} \Gamma(\nu/2)}$$



Probabilidades de χ^2

¿Cuál es la probabilidad de obtener un valor de χ_o^2 o mayor?

$$P(\chi^2 \geq \chi_o^2) = \int_{\chi_o^2}^{\infty} p_v(\chi^2) d\chi^2$$

Método General

- Realizar una serie de medidas.
- Calcular el valor de χ_o^2 .
- Calcular el valor de chi-cuadrado reducido $\bar{\chi}_o^2$.
- Calcular la probabilidad de obtener un valor de $\bar{\chi}^2$ igual o mayor:

$$Prob(\bar{\chi}^2 \geq \bar{\chi}_o^2)$$

- Si el valor obtenido es alto no hay razón para desechar la hipótesis.
- Si el valor obtenido es muy bajo, desechar la hipótesis.

En general:

$Prob(\bar{\chi}^2 \geq \bar{\chi}_o^2) < 5\%$	Hay un desacuerdo significativo
$Prob(\bar{\chi}^2 \geq \bar{\chi}_o^2) < 1\%$	Hay un desacuerdo muy significativo

Ejemplo 1: Distribución de Poisson

Realizamos 100 medidas de 1 minuto cada una, del número de rayos cósmicos que llegan a un contador Geiger.

Cuentas x en 1 minuto	Sucesos	Número de bin k	Sucesos Observados O_k	Sucesos Esperados E_k
Ninguna	7	1	7	7.5
Una	17	2	17	19.4
Dos	29	3	29	25.2
Tres	20	4	20	21.7
Cuatro	16	5	16	14.1
Cinco	8			
Seis	1			
Siete	2	6	11	12.1
Ocho o más	0			
Total	100			

Con valor medio $\bar{x} = \frac{\sum x_i}{N} = 2.59$

¿Vienen gobernadas por una distribución de Poisson con media $\mu = \bar{x}$?

Chi-cuadrado $\rightarrow \chi^2 = 1.39$

Ligaduras $\rightarrow l = 2$

Grados de libertad $\rightarrow \nu = n - l = 6 - 2 = 4$

Chi-cuadrado reducido $\rightarrow \bar{\chi}^2 = \chi^2 / \nu = 1.39 / 4 = 0.35$

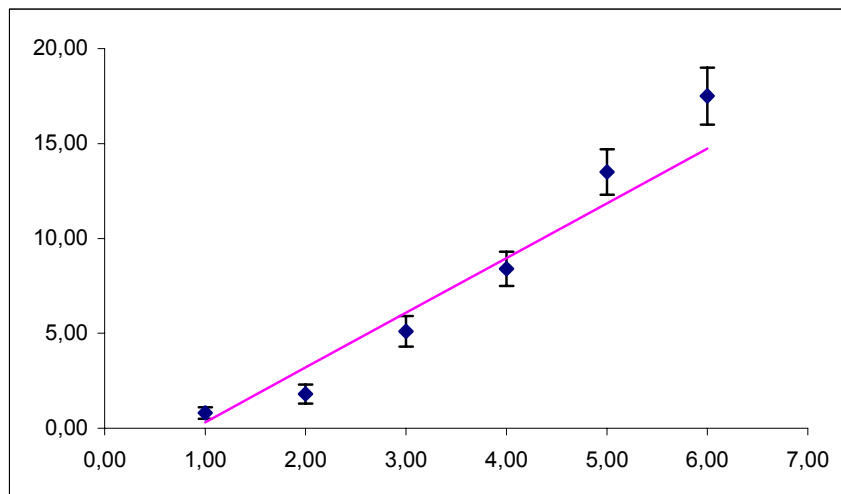
$\text{Prob}(\bar{\chi}^2 \geq 0.35) \approx 85\%$

No hay razón para pensar que la distribución no es de Poisson



Ejemplo 2: Bondad de un ajuste

x_i	1	2	3	4	5	6
y_i	0.8	1.8	5.1	8.4	13.5	17.5
σ_i	0.3	0.5	0.8	0.9	1.2	1.5



Resultado $y = (2.88 \pm 0.18)x + (-2.6 \pm 0.4)$

¿Se trata de un buen ajuste? ¿La hipótesis lineal es correcta?

$$\chi^2 = \sum_{i=1}^6 \left(\frac{y_i - f(x_i)}{\sigma_i} \right)^2 = \sum_{i=1}^6 \left(\frac{y_i - ax_i - b}{\sigma_i} \right)^2 = 17.6$$

Chi-cuadrado $\rightarrow \chi^2 = 17.6$

Grados de libertad $\rightarrow \nu = n - l = 6 - 2 = 4$

Chi-cuadrado reducido $\rightarrow \bar{\chi}^2 = \chi^2 / \nu = 17.6 / 4 = 4.4$

$$\text{Prob}(\bar{\chi}^2 \geq 4.4) \approx 0.14\%$$

La hipótesis lineal no es correcta

