Bondad de ajuste

Todos los de χ^2 son de cola derecha, en si la formula del χ^2_c de cola derecha es $\chi^2_c = \chi^2_{1-\alpha, v}$ con $v = k-1$ pero aqui la significancia va directo en el $1-\alpha$

"use 0,05 de significancia" $\rightarrow \chi^2_{0.05, v}$ ✓ ~~$\chi_{1-0.05}$~~

OJO!!! Se pueden usar otras distribuciones, no solo χ^2

(ver ejemplo 2)

k = cantidad de eventos n = cantidad de muestra

O = Valores observados P_i = Probabilidad

e = Valores esperados $n \cdot P_i$ $v = k-1$

$\chi^2_c = \chi^2_{\text{significancia dada}, v}$

$$\chi^2_{obs} = \sum_{i=p}^{n} \frac{(O_i - e_i)^2}{e_i}$$, hacer una tabla

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

1. Se toma un muestra de tamaño n
2. El experimento se subdivide en $i = 1,2,3,\dots,k$ eventos
3. Cada evento i tiene una probabilidad de ocurrencia p_i
4. El resultado esperado para cada evento i, se calcula $e_i = np_i$ (Frecuencias esperadas nivel teórico)
5. De la muestra de tamaño n, se identifica $o_i = n_i$ observaciones para cada evento (frecuencias observadas)
6. o_i: valor o dato observado experimentalmente
7. e_i: valor esperado (teórico)

Aclaraciones sobre bondad de ajuste

	K-Eventos				
i	**1**	**2**	**3**	**...**	**k**
Frecuencias observadas (o_i)	o_1	o_2	o_3	$\dots$	o_k
Frecuencias esperadas (e_i)	e_1	e_2	e_3	$\dots$	e_k
$\frac{(o_i-e_i)^2}{e_i}$	$\frac{(o_1-e_1)^2}{e_1}$	$\frac{(o_2-e_2)^2}{e_2}$	$\frac{(o_3-e_3)^2}{e_3}$	$\dots$	$\frac{(o_k-e_k)^2}{e_k}$

$$\mathcal{X}^2_{obs} = \sum_{i=1}^{k} \frac{(o_i-e_i)^2}{e_i} = \frac{(o_1-e_1)^2}{e_1} + \frac{(o_2-e_2)^2}{e_2} + \frac{(o_3-e_3)^2}{e_3} + \dots + \frac{(o_k-e_k)^2}{e_k}, \text{con } gl: v = k-1$$

Contraste que se intenta estudiar es del tipo siguiente:

$$\left.\begin{array}{l} H_0\text{: La distribución de } X \text{ es de tipo } X_0 \\ H_1\text{: La distribución de } X \textbf{ no} \text{ es de tipo } X_0 \end{array}\right\}$$

Dónde X_0 es un tipo de distribución conocida, a saber normal, binomial, Poisson, etc.

- Identifique el tamaño de la muestra: n
- Identifique las frecuencias observadas: $O_1, O_2, O_3, \dots, O_k$
- Calcule las estimaciones teóricas esperadas: $e_1, e_2, e_3, \dots, e_k$, donde $\boxed{e_i \geq 5, i = 1,2,3,\dots,k}$
- Calcule el estadístico de prueba u observado: $\mathcal{X}^2 = \sum_{i=1}^{k} \frac{(o_i-e_i)^2}{e_i}$, con $gl: v = k-1$

$e_3 = 6$ $\Rightarrow e_3 = 6+3 = 9$
$e_4 = 3$ si no cumple

Si χ^2_{obs} es pequeño, entonces se puede asumir:

1. O_i es similar a e_i
2. Habrá un buen ajuste
3. No existe suficiente evidencia en contra para rechazar H_0

Reglas de decisión: Las pruebas de bondad de ajuste son de cola derecha

Dónde $\chi^2_c = \chi^2_{1-\alpha,k-1}$ y $ValorP = P(\chi^2 > \chi^2_{obs})$

Bondad de ajuste – Pruebas de Cola derecha

Ejemplo 1:

Un grupo de ratas baja 90 veces por una rampa que conecta con 3 puertas. Se observa que en 23 ocasiones las ratas pasaron por la puerta 1, 36 veces por la puerta 2 y 31 veces por la puerta 3.

Se desea probar que las ratas no tienen preferencias por alguna de las puertas en particular. Use un nivel de significancia de 0.05.

p_i = Proba de puerta i, $i = 1,2,3$

$H_0: p_1 = p_2 = p_3 = \frac{1}{3}$ Hay 3 puertas, la proba de escoger cualquiera es de $\frac{1}{3}$

H_1: Las ratas tienen preferencia por alguna de las puertas

$k = 3 \rightarrow 3$ datos

$n = 90$ $\alpha = 0{,}0$

Valores observados

Puerta 1: 23 Puerta 2: 36 Puerta 3: 31

Intentos esperados $n \cdot p_i$

Puerta 1: $90 \cdot \frac{1}{3} = 90$ Puerta 2: $90 \cdot \frac{1}{3} = 30$ Puerta 3: $90 \cdot \frac{1}{3} = 30$

Eventos	1	2	3
Observados O	23	36	31
Esperados e	30	30	30
$\frac{(O_i - E_i)^2}{E_i}$	$\frac{(23-30)^2}{30}$	$\frac{(36-30)^2}{30}$	$\frac{(31-30)^2}{30}$
	1,63	1,2	0,03

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

La suma es el valor Chi-Cuadrado esperado: $1{,}6 + 1{,}2 + 0{,}03 = 2{,}86$

Por ser cola derecha pero se usa directo α

$\chi^2_c = \chi^2_{1-\alpha, v}$ con $v = k - 1$

RA | RR

χ^2_c χ^2_{obs}

$\alpha = 0{,}05$ $n = 90$

$1 - \alpha = 0{,}95$ $v = 89$

5,99146

$\chi^2_c = \chi^2_{0.05, 2} = 5{,}99146$ χ_c^2

Como $2{,}86 < 5{,}99146$, No se rechaza H_0, las ratas no muestran preferencia por alguna puerta

Con valor P:

Hacer la tabla para encontrar en el χ^2_{obs} y meter en la app

$P(\chi^2 > \chi^2_{obs})$ $\quad \chi^2_{obs} = 2{,}86 \quad v = 2 \quad \alpha = 0{,}05$

$= P(\chi^2 > 2{,}86)$ Valor p

$= 0{,}23931$ ✓

$\boxed{\text{Como } P = 0{,}23931 > 0{,}05 \text{ NO se rechaza } H_0}$

El curso Estructura de Datos II, es impartido en el segundo año de la carrera de Ingeniería en Sistemas de una Universidad. Un profesor considera que las edades de los estudiantes que llevan dicho curso, sigue una distribución normal. En una muestra de 200 estudiantes se obtuvieron los resultados de la tabla adjunta. ¿Considera correcta la afirmación del profesor?

Edad: X	Frecuencia Observada
$X < 18$	30
$18 \leq X < 19$	100
$19 \leq X < 20$	50
$20 \leq X < 21$	15
$X \geq 21$	5

Considera para los datos:

μ $\quad \bar{x} = 18.825$

σ^2 $\quad s = 0.90746486^2$

Si no dice α, entonces $\alpha = 0{,}05$

X = edad de estudiantes del curso $\quad k = 5 \quad n = 200 \quad v = 4$

$\alpha = 0{,}05$

$H_0: X \sim N(18.825, 0.90746486^2)$

$H_1: X \nsim N(18.825, 0.90746486^2)$

Valores observados: 30, 100, 50, 15, 5

Valores esperados (Recordar que estamos usando normal)

$P(a \leq x < b)$

$= P(b < x) - P(a < x)$

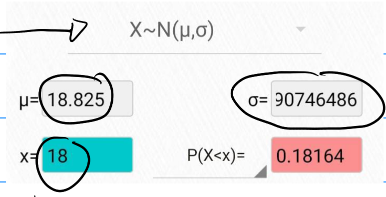
Edad: X	Frecuencia Observada
$X < 18$	30
$18 \leq X < 19$	100
$19 \leq X < 20$	50
$20 \leq X < 21$	15
$X \geq 21$	5

Valores esperados (Recordar que estamos usando normal)

$e_1 = 200 \cdot P(X < 18) =$

$= 200 \cdot 0{,}1816$

$= 36{,}32$



Los demas se meten igual en la app

$e_2 = 200 \cdot P(18 \leq x < 19)$

$= 200 \cdot [P(x < 19) - P(x < 18)]$

$= 200 [0{,}5765 - 0{,}1816]$

$= 200 \cdot 0{,}3949$

$= 78{,}98$

$e_3 = 200 \cdot P(19 \leq x < 20)$

$= 200 \cdot [P(x < 20) - P(x < 19)]$

$= 200 [0{,}9023 - 0{,}5765]$

$= 200 \cdot 0{,}3258$

$= 65{,}16$

$e_4 = 200 \cdot P(20 \leq x < 21)$

$= 200 \cdot [P(x < 21) - P(x < 20)]$

$= 200 [0{,}9917 - 0{,}9023]$

$= 200 \cdot 0{,}0894$

$= 17{,}88$

$e_5 = 200 \cdot P(x \geq 21)$

$= 200 \cdot 0{,}0827$

$= 1{,}654 < 5$, no cumple el requisito, entonces

$e_4 = 17{,}88 + 1{,}654 = 19{,}53$

Y el evento se vuelve

Edad: X	Frecuencia Observada
$X < 18$	30
$18 \leq X < 19$	100
$19 \leq X < 20$	50
$20 \leq X < 21$	15
$X \geq 21$	5

$15 + 5 = 20$

NO Cumple

Eventos	1	2	3	4	~~5~~
Observados O	30	100	50	20	~~5~~
Esperados e	36,32	78,98	65,16	19,53	~~1,6594~~
$\frac{(O_i - e_i)^2}{e_i}$	$\frac{(30-36,32)^2}{36,32}$	$\frac{(100-78,98)^2}{78,98}$	$\frac{(50-65,16)^2}{65,16}$	$\frac{(20-19,53)^2}{19,53}$	~~$\frac{(5-1,6594)^2}{1,6594}$~~
Resultado	1,09974	5,59433	3,52710	0,01131	

Entonces $x^2_{obs} = 1,09974 + 5,59433 + 3,52710 + 0,01131$

$x^2_{obs} = 10,23248$

$x^2_c = x^2_{0,05,\,3} = 7,81473$

K era 5 pero como uno no cumple ahora es 4 entonces $v = 3$

Cola derecha

Si se cumple
$x^2_{obs} < x^2_c$, NO se rechaza H_0

Si se cumple
$x^2_{obs} > x^2_c$, Se rechaza H_0

Como $x^2_{obs} = 10,23248 > x^2_c = 7,81473$

Se rechaza H_0, por lo que no es posible asumir normalidad

Valor Pi: $P(x^2 > x^2_{obs})$

$= P(x^2 > 10,23248) = 0,01669$

$\boxed{\text{Como } 0,01669 \leq 0,05 \text{, se rechaza } H_0}$

Con el objetivo de redactar un informe anual del curso de Estadística se registraron las ausencias por semana durante todo el año lectivo (38 semanas).

	1	2	3	4	5
Ausencias	0	1	2	3	4 o más
observados **Frecuencia**	6	9	10	6	7

$\Sigma = 38$

$u = \lambda$

$\lambda = 1{,}94$

Con una significancia de 5 %, ¿puede concluirse que el número de ausencias por semana sigue una distribución Poisson $\left(P(X=k) = \dfrac{\lambda^k e^{-\lambda}}{k!}\right)$ con media 1.94? **(5 puntos)**

$H_0: X \sim P(1{,}94)$ $\quad n = 38 \quad k = 5 \ (0,1,2,3,4) \quad \lambda = 1{,}94$

$H_1: X \nsim P(1{,}94)$ $\quad v = 4 \quad \alpha = 0{,}05$

Observados: 6, 9, 10, 6, 7

$$e_1 = 38 \cdot \frac{1{,}94^0 \cdot e^{-1{,}94}}{0!} = 5{,}461$$

$$e_2 = 38 \cdot \frac{1{,}94^1 \cdot e^{-1{,}94}}{1!} = 10{,}594$$

$$e_3 = 38 \cdot \frac{1{,}94^2 \cdot e^{-1{,}94}}{2!} = 10{,}276$$

$$e_4 = 38 \cdot \frac{1{,}94^3 \cdot e^{-1{,}94}}{3!} = 6{,}645$$

$$e_5 = 38 \cdot \frac{1{,}94^4 \cdot e^{-1{,}94}}{4!} = 5{,}024$$

Eventos	0	1	2	3	4
Observados O	6	9	10	6	7
Esperados e	5,461	10,594	10,276	6,645	5,024
$\dfrac{(O_i - e_i)^2}{e_i}$	$\dfrac{(6-5{,}461)^2}{5{,}461}$	$\dfrac{(9-10{,}594)^2}{10{,}594}$	$\dfrac{(10-10{,}276)^2}{10{,}276}$	$\dfrac{(6-6{,}645)^2}{6{,}645}$	$\dfrac{(7-5{,}024)^2}{5{,}024}$

$X^2_{obs} = 1{,}1784$ $\quad X^2_c = X^2_{0{,}05,\,4} = 9{,}48773$

$\boxed{\text{Como } 1{,}1784 < 9{,}48773, \text{ no se rechaza } H_0}$

Hay seis tipos diferentes de colores en un paquete de chocolates M&Ms: azul, rojo, naranja, amarillo, verde y café. Se quiere saber si la distribución de estos 6 colores sucede en igual proporción. En la fábrica, antes de empacarlos, se toma una muestra simple de 600 chocolates M&Ms y se observan las distribuciones que se muestran en el Cuadro 1. Constraste la hipótesis de que los colores de los M&Ms se adaptan a una distribución uniforme. **(5 puntos)**

	1	2	3	4	5	6
Color	Azul	Naranja	Verde	Rojo	Amarillo	Café
Cantidad	212	147	103	50	46	42

$H_0: P_1 = P_2 = P_3 = P_4 = P_5 = P_6 = \frac{1}{6}$ $n = 600$ $k = 6$

$\alpha = 0{,}05$ $v = 5$

$H_1: X \not\sim U$

Observados: 212, 147, 103, 50, 46, 42

$e_{1 \dots 6} = 600 \cdot \frac{1}{6} = 100$

Eventos	1	2	3	4	5	6
Observados O	212	147	103	50	46	42
Esperados e	100	100	100	100	100	100

$$\sum_{i=1}^{k} \frac{(O_i - e_i)^2}{e_i}$$

$$= \frac{(212-100)^2}{100} + \frac{(147-100)^2}{100} + \frac{(103-100)^2}{100}$$

$$\frac{(50-100)^2}{100} + \frac{(46-100)^2}{100} + \frac{(42-100)^2}{100}$$

$X^2_{obs} = 235{,}42$ $X^2_c = X^2_{0{,}05,\,5} = 11{,}07050$

$\boxed{\text{Como } 235{,}42 > 11{,}07050 \text{ se rechaza } H_0}$

$P = P(\chi^2 > \chi^2_{obs}) = P(\chi^2 > 235.42) \approx 0$