PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

FACULTAD DE MATEMÁTICAS / DEPARTAMENTO DE ESTADÍSTICA

ELM2400 Métodos Estadísticos

Test Chi Cuadrado de Independencia

Profesor: Alexis Peña
Ayudante: Reinaldo González S.

Supongamos que dos variables aleatorias han sido categorizadas en k y l niveles, respectivamente. Podemos, a partir de una muestra de tamaño n, observar los datos de la siguiente manera.

	1	2	$\cdots$	l	
1	n_{11}	n_{12}	$\cdots$	n_{1l}	$n_{1\cdot}$
2	n_{21}	n_{22}	$\cdots$	n_{2l}	$n_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	
k	n_{k1}	n_{k2}	$\cdots$	n_{kl}	$n_{k\cdot}$
	$n_{\cdot 1}$	$n_{\cdot 2}$		$n_{\cdot l}$	$n_{\cdot\cdot} = n$

Donde:

$$n_{i\cdot} = \sum_{j=1}^{l} n_{ij} \quad \text{y} \quad n_{\cdot j} = \sum_{i=1}^{k} n_{ij}$$

La tabla anterior se conoce como “Tabla de Contingencia”. Construiremos una tabla de valores esperados bajo la hipotesis H_0 : Existe Independencia.

Sea e_{ij}: el valor esperado en la celda (i,j), O_{ij}: el valor observado en la celda (i,j). Y por ultimo p_{ij}: la probabilidad de pertenecer a la celda (i,j).

$$e_{ij} \;=\; n \cdot p_{ij} = n \cdot \frac{n_{ij}}{n}$$

$$\text{Bajo } H_0 \qquad p_{ij} = \frac{n_{ij}}{n} = \underbrace{\frac{n_{i\cdot}}{n}}_{p_i} \cdot \underbrace{\frac{n_{\cdot j}}{n}}_{p_j}$$

$$\text{Luego} \qquad e_{ij} = \frac{n_{i\cdot} n_{\cdot j}}{n}$$

En español, se lee:

$$e_{ij} = \frac{\text{Total de Fila * Total de Columnas}}{\text{Gran Total}}$$

Se Rechaza H_0 si:

$$\chi^2_c = \sum_{j=1}^{l} \sum_{i=1}^{k} \frac{(O_{ij} - e_{ij})^2}{e_{ij}} > \chi^2_{1-\alpha,(k-1)(l-1)}$$

Ejemplo: En una muestra aleatoria de 100 universitarios se clasifico cada uno de ellos segun si habia consumido alguna vez droga o no y el promedio de notas. A partir de los datos tabulados en la tabla ¿Proporcionan estos datos evidencia suficiente como para concluir que hay una relacion entre las dos variables? Use $\alpha = 0{,}05$.

Promedio de Notas	Si Consume	No Consume	Total
$\leq 4{,}0$	10	29	39
$> 4{,}0$	20	41	61
Total	30	70	100

Las hipotesis son:

$$\begin{aligned} H_0 &: \text{Existe independencia entre el consumo de drogas y el promedio de notas} \\ H_1 &: \text{Existe asociacion entre el consumo de drogas y el promedio de notas} \end{aligned}$$

Para testear tales hipotesis, se ocupa el estadistico:

$$\chi_c^2 = \sum_{j=1}^{l} \sum_{i=1}^{k} \frac{(O_{ij} - e_{ij})^2}{e_{ij}}$$

Donde $e_{ij} = \dfrac{n_{i\cdot} n_{\cdot j}}{n_{\cdot\cdot}}$. El cual rechaza H_0 cuando $\chi_c^2 > \chi^2_{1-\alpha,(I-1)(J-1)}$.

Luego, la tabla de Valores Esperados es:

Promedio de Notas	Si Consume	No Consume	Total
$\leq 4{,}0$	11.7	27.3	39
$> 4{,}0$	18.3	42.7	61
Total	30	70	100

Por lo tanto el estadistico calculado queda:

$$\begin{aligned} \chi_c^2 &= \frac{(10-11{,}7)^2}{11{,}7} + \frac{(29-27{,}3)^2}{27{,}3} + \frac{(20-18{,}3)^2}{18{,}3} + \frac{(41-42{,}7)^2}{42{,}7} \\ &= 0{,}578 \end{aligned}$$

Como $\chi_c^2 = 0{,}578 < 3{,}841 = \chi^2_{0{,}95,1}$, no se rechaza H_0, es decir, con un 95 % de confianza el consumo de droga no influye en el promedio de notas de los estudiantes.