$X \equiv$ característica
$Y \equiv$ muestras

MUESTRAS

	$M_1 \cdots$	M_j	$\cdots M_K$	
A_1		$\vdots$		
$\vdots$				
A_i	$\cdots$	n_{ij}	$\cdots$	$\boxed{n_{i\cdot} = p_i}$
$\vdots$				
A_h		$\vdots$		
		$n_{\cdot j}$		n

$n_{ij} \equiv$ nº observaciones $\begin{cases} \text{modalidad } i \\ \text{muestra } j \end{cases}$

$n_{i\cdot} \equiv$ nº total ~~observaciones~~ muestras con la caract. i.

$n_{\cdot j} \equiv$ nº observaciones de la muestra j.

Contrastar la hipótesis nula de que todas las muestras proceden de la misma población equivale a contrastar la hipótesis que, para todas las muestras, se verifica que la característica está presente en la misma proporción que en la población, para cada una de sus h modalidades:

H_0: muestras homogéneas (procedentes de la misma población)

$$H_0 : P(A_i) = p_i \quad , i = 1 \text{—} h$$
$$\text{s.a. } \sum_{i=1}^{h} p_i = 1$$

Para realizar el contraste se utiliza el estadístico $\chi^2 = \sum\sum \dfrac{(O_{ij} - E_{ij})^2}{E_{ij}}$

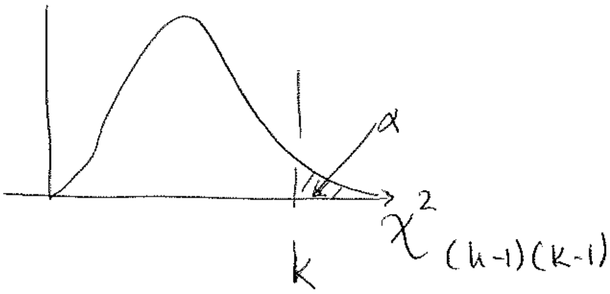
donde E_{ij} son las frecuencias esperadas de acuerdo con la hipótesis nula,

$$E_{ij} = n_{\cdot j} \cdot p_i$$

$$\chi^2_{indep} = \sum_{i=1}^{h}\sum_{j=1}^{k} \frac{\left(n_{ij} - \frac{n_{i\cdot}n_{\cdot j}}{n}\right)^2}{\frac{n_{i\cdot}n_{\cdot j}}{n}} \longrightarrow \chi^2_{(h-1)(k-1)}$$

La región crítica viene determinada por el nivel de significación α:

$$P\left[\chi^2_{(h-1)(k-1)} \geq k / H_0\right] = \alpha$$



Al ser una aplicación del contraste de bondad de ajuste, aquí también mantenemos la condición de que $E_{ij} = \frac{n_{i\cdot}n_{\cdot j}}{n} > 5$, y en caso de que no se verifique hay que reagrupar por filas o por columnas y reducir el nº de grados de libertad.

2_ CONTRASTE de HOMOGENEIDAD

El contraste de homogeneidad es otra aplicación del contraste de bondad de ajuste con el estadístico χ^2. Formalmente, es igual que un contraste de independencia, y se opera de la misma manera; pero cambian los significados de las categorías de la tabla de contingencia y el nº de parámetros a estimar.

Partimos de una tabla de contingencia de $h \times k$, donde las h filas se interpretan como las A_1 —— A_h modalidades distintas de la característica A; y las k columnas corresponden a las M_1 —— M_k muestras distintas que se presuponen procedentes de la misma población.

En el contraste de homogeneidad se quiere contrastar si las muestras son homogéneas, es decir, si todas las muestras proceden de la misma población.

Lagrangiano:

$$\phi(p_{i\cdot}, p_{\cdot j}) = \sum_{i=1}^{h}\sum_{j=1}^{K}(n_{ij} \ln p_{i\cdot} + n_{ij} \ln p_{\cdot j}) - \gamma_1\left(\sum_{i=1}^{h} p_{i\cdot} - 1\right) - \gamma_2\left(\sum_{j=1}^{K} p_{\cdot j} - 1\right)$$

La condición necesaria de óptimo lleva a calcular las derivadas parciales del lagrangiano e igualarlas a 0:

$$\frac{\partial \phi}{\partial p_{i\cdot}} = \sum_{j=1}^{K} n_{ij}\frac{1}{p_{i\cdot}} - \gamma_1 = n_{i\cdot}/p_{i\cdot} - \gamma_1 = 0 \Rightarrow \gamma_1 = \frac{n_{i\cdot}}{p_{i\cdot}} \Rightarrow p_{i\cdot} = \frac{n_{i\cdot}}{\gamma_1}$$

$$\frac{\partial \phi}{\partial p_{\cdot j}} = \sum_{i=1}^{h} n_{ij}\frac{1}{p_{\cdot j}} - \gamma_2 = n_{\cdot j}/p_{\cdot j} - \gamma_2 = 0 \Rightarrow \gamma_2 = \frac{n_{\cdot j}}{p_{\cdot j}} \Rightarrow p_{\cdot j} = \frac{n_{\cdot j}}{\gamma_2}$$

$$\frac{\partial \phi}{\partial \gamma_1} = \sum_{i=1}^{h} p_{i\cdot} - 1 = 0 \longrightarrow \frac{1}{\gamma_1}\sum_{i=1}^{h} n_{i\cdot} = 1 \Rightarrow \gamma_1 = n$$

$$\frac{\partial \phi}{\partial \gamma_2} = \sum_{j=1}^{K} p_{\cdot j} - 1 = 0 \longrightarrow \frac{1}{\gamma_2}\sum_{j=1}^{K} n_{\cdot j} = 1 \Rightarrow \gamma_2 = n$$

Sustituyendo los valores de $\hat{\gamma}_i$, los estimadores máximo-verosímiles de las probabilidades marginales son:

$$\hat{p}_i = \frac{n_{i\cdot}}{n}, \quad i = 1 \dots h$$

$$\hat{p}_j = \frac{n_{\cdot j}}{n}, \quad j = 1 \dots K$$

Para efectuar el contraste de independencia utilizamos el estadístico χ^2 del contraste de bondad de ajuste,

$$\chi^2 = \sum_{i=1}^{h}\sum_{j=1}^{K} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

siendo E_{ij} las frecuencias estimadas bajo la hipótesis nula cierta de independencia, es decir:

$$E_{ij} = n \cdot \hat{p}_{ij} = n \cdot \hat{p}_{i\cdot} \cdot \hat{p}_{\cdot j} = n \cdot \frac{n_{i\cdot}}{n} \cdot \frac{n_{\cdot j}}{n} = \frac{n_{i\cdot}\, n_{\cdot j}}{n}$$

donde el nº de grados de libertad es igual al nº total de observaciones − nº parámetros estimados − 1 $= h \cdot K - (h+K-2) - 1$

$= hK - h - K + 1 = h(K-1) - K + 1 = h(K-1) - (K-1) =$

$= (h-1)(K-1)$

Procedimiento: De una m.a.s. de tamaño n, se observan las características X e Y y se presentan en una tabla de contingencia.

$$\begin{cases} H_0 : X \text{ e } Y \text{ son independientes} \\ H_1 : X \text{ e } Y \text{ no son independientes} \end{cases} \longrightarrow \left\{ H_0 : p_{ij} = p_{i\cdot} \cdot p_{\cdot j} \;, \forall i,j \right.$$

Designamos por

$p_{ij} \equiv P(A_i \cap B_j) \rightarrow$ probab. de que un elemento presente las características A_i y B_j conjuntamente.

$p_{i\cdot} \equiv P(A_i) \rightarrow$ probab. de que un elemento presente A_i

$p_{\cdot j} \equiv P(B_j) \rightarrow$ probab. de que un elemento presente B_j

Se verifica:

$$\sum_{i=1}^{h} p_{i\cdot} = \sum_{j=1}^{k} p_{\cdot j} = \sum_{i=1}^{h} \sum_{j=1}^{k} p_{ij} = 1$$

Bajo H_0 cierta, la distrib. de probabilidad conjunta de las caract. X e Y vendrá dada por la expresión $p_{ij} = p_{i\cdot} \cdot p_{\cdot j}$, y contendrá $h * k - 2$* parámetros desconocidos $p_{i\cdot}, p_{\cdot j}$ que habrá que estimar, a partir de la función de verosimilitud conjunta.

$$\mathcal{L}(p_{i\cdot}, p_{\cdot j}) = \sum_{i=1}^{h} \sum_{j=1}^{k} (p_{i\cdot} \cdot p_{\cdot j})^{n_{ij}}$$ ¿¿PRODUCTO??

Tomando logaritmos:

$$\ln \mathcal{L}(p_{i\cdot}, p_{\cdot j}) = \sum_{i=1}^{h} \sum_{j=1}^{k} n_{ij} \ln (p_{i\cdot} \cdot p_{\cdot j})$$

Añadimos las restricciones lineales $\begin{cases} \sum_i p_{i\cdot} = 1 \\ \sum_j p_{\cdot j} = 1 \end{cases}$ para obtener el lagrangiano a optimizar.

---

* Ya que $\sum p_{i\cdot} = \sum p_{\cdot j} = 1$, uno de cada $p_{i\cdot}$ y $p_{\cdot j}$ se obtendrá por diferencia lineal

1- CONTR. de INDEPENDENCIA.

Tanto el contraste de independencia como el contraste de homogeneidad parten de las tablas de contingencia, que expondremos a continuación.

Una tabla de contingencia $h \times k$ es una tabla de doble entrada, con h filas y k columnas, donde:

Factor Y

Factor X	B_1	...	B_j	...	B_k	
A_1						$n_{1.}$
⋮			⋮			⋮
A_i	...	...	n_{ij}			$n_{i.}$
⋮						⋮
A_h						$n_{h.}$
$n_{.j}$	$n_{.1}$	...	$n_{.j}$	...	$n_{.k}$	$n_{..} = n$ (tamaño muestral)

Características X e Y

$A_1 \dots A_h$, $B_1 \dots B_k$ modalidades de las características

$n_{ij} \equiv$ frec. absoluta conjunta → nº veces que se presentan conjuntamente la modalidad A_i de X y la modalidad B_j de Y.

$n_{i.} \equiv$ frec. marginal de A_i → nº de veces que se presenta la modalidad A_i de X, sin tener en cuenta los valores de Y.

$n_{.j} \equiv$ frec. marginal de B_j → nº de veces que se presenta la modalidad B_j de Y, sin considerar los valores de X.

$$\sum_{j=1}^{k} n_{ij} = n_{i.} \quad , \; i = 1 \dots h \longrightarrow \sum_{i=1}^{h} n_{i.} = n_{..}$$

$$\sum_{i=1}^{h} n_{ij} = n_{.j} \quad , \; j = 1 \dots k \longrightarrow \sum_{j=1}^{k} n_{.j} = n_{..}$$

$$\sum_{i=1}^{h} \sum_{j=1}^{k} n_{ij} = n$$

ESTAD_ T.28. CONTRASTES de INDEPENDENCIA.
CONTRASTE de HOMOGENEIDAD.
CONTRASTES de ALEATORIEDAD.
CONTRASTES de LOCALIZACIÓN.
CONTRASTES de COMPARACIÓN de POBLACIONES

0_ <u>INTRODUCCIÓN</u>.