21/03/2018

= Two-way contingency tables =

Consideramos un conjunto de observaciones que toman valores en dos dimensiones categóricas, $i = 1, \dots, I$ y $j = 1, \dots, J$. Así, cada observación y_i está definida en $\mathcal{I} \times \mathcal{J}$, donde

$$\mathcal{I} = \{1, 2, \dots, I\}$$

$$\mathcal{J} = \{1, 2, \dots, J\},$$

definiendo así el soporte para $(y_\ell)_{\ell=1}^{n}$, n - datos.

Un conjunto de 'n' variables de este tipo puede reacomodarse como una matriz de contingencia de dos dimensiones:

Dim 1 \ Dim 2	1	2	$\cdots$	J	
1	n_{11}	n_{12}	$\cdots$	n_{1J}	$n_{1\circ}$
2	n_{21}	n_{22}	$\cdots$	n_{2J}	$n_{2\circ}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	
I	n_{I1}	n_{I2}	$\cdots$	n_{IJ}	$n_{I\circ}$
	$n_{\circ 1}$	$n_{\circ 2}$		$n_{\circ J}$	n

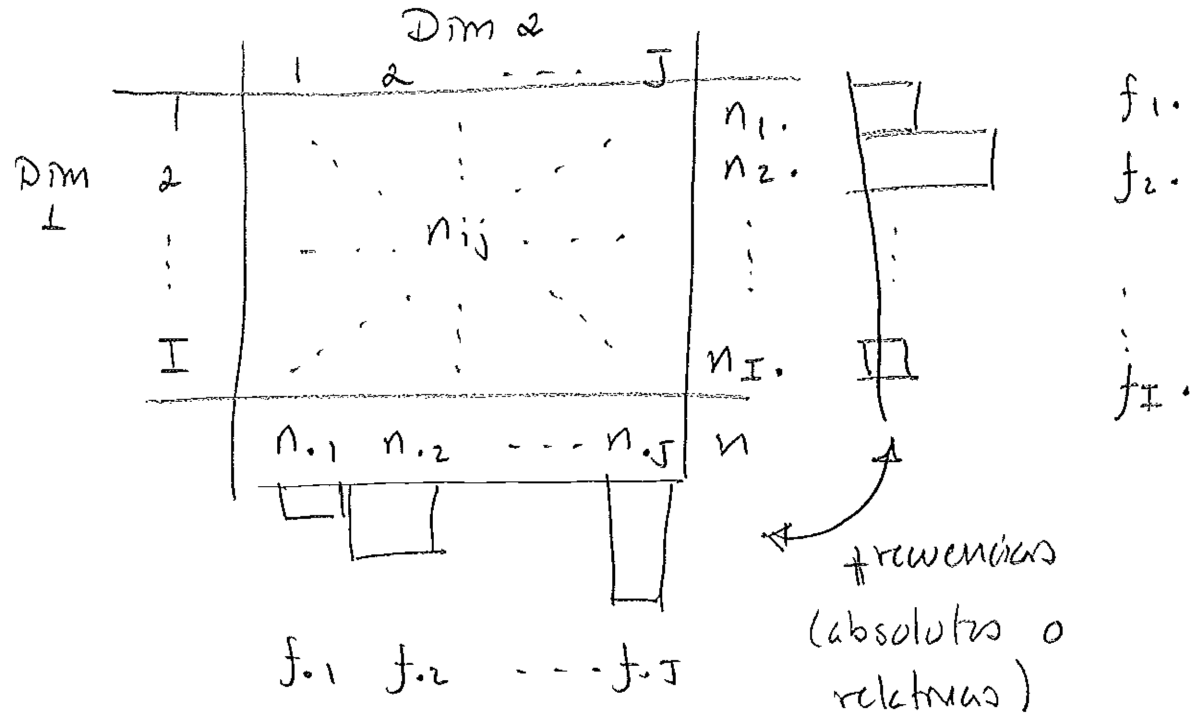
n - # de observaciones / datos / variables.

donde
$$n = \sum_{i=1}^{I} \sum_{j=1}^{J} n_{ij} = \sum_{i=1}^{I} n_{i\circ} = \sum_{j=1}^{J} n_{\circ j}$$

donde $n_{i\circ} = \sum_{j=1}^{J} n_{ij}$ y $n_{\circ j} = \sum_{i=1}^{I} n_{ij}$.

con $n_{ij} = \#\{\cancel{\#} y_\ell : y_{\ell 1} \in \{i\} \text{ y } y_{\ell 2} \in \{j\}\}$.

A partir de la tabla de contingencia, recuperamos las distribuciones marginales de las <u>dos</u> dimensiones:



Para la <u>dimensión 1</u>, las frecuencias $\{f_{i\cdot}\}_{i=1}^{I}$ describen la variabilidad intrínseca a esta dimensión. De igual forma lo hacen $\{f_{\cdot j}\}_{j=1}^{J}$ para la <u>dimensión 2</u>.

Ambas descripciones de variabilidad, por separado, son estándar y corresponden al caso de variabilidad descrita por una distribución <u>multinomial</u>.

▷ Lo que resulta <u>sumamente</u> <u>interesante</u> es describir la variabilidad al interior de la tabla de contingencia (renglones o columnas) y su <u>interacción</u> ~~con los~~ entre los demás casos de cada dimensión.

Para esto, necesitamos definir algunas medidas de utilidad (cinco):

1) Risk (Riesgo)

$$\text{Risk} = \begin{array}{l}\text{proporción con una característica}\\ \text{poco deseable}\end{array} = \frac{\#\{y_i : y_{i1} \in \{i^*\} \text{ y } y_{i2} \in \{j^*\}\}}{n}$$

$$= \text{Ris}\{i^*, j^*\}$$

donde $\{i^*\}$ y $\{j^*\}$ son características no deseables en $\mathcal{I}$ y $\mathcal{J}$, respectivamente.

= se usa en casos de tabla 2×2 =

2) Riesgo relativo

se define como el cociente de los riesgos entre dos características poco deseables, i.e.

$$RR(\{i^*, j^*\}, \{i', j'\}) = \frac{\text{Risk}(\{i^*, j^*\})}{\text{Risk}(\{i', j'\})} = \frac{\#\{y_i : y_{i1} \in \{i^*\}, y_{i2} \in \{j^*\}\}}{\#\{y_i : y_{i1} \in \{i'\}, y_{i2} \in \{j'\}\}}.$$

= Esta medida puede definirse en una sola dimensión, ya sea en "i" o en "j" =.

3) Riesgo incremental

Porcentaje de incremento del Riesgo $\{i^*, j^*\}$ relativo al riesgo Risk $\{j', j'\}$.

$$\Delta RR(\{i^*, j^*\}, \{i', j'\}) = \left(RR(\{i^*, j^*\}, \{i', j'\}) - 1\right) 100\%$$

4) Momios (odds)

Proporción de casos con una cierta característica $\{i^*, j^*\}$ relativo a los casos sin esa característica, i.e.

$$\text{odd}\{i^*, j^*\} = \frac{\#\{y_i : y_{i1} \in \{i^*\},\ y_{i2} \in \{j^*\}\}}{\#\{y_i : y_{i1} \notin \{i^*\},\ y_{i2} \in \{j^*\}\}}.$$

= Se aplica en ambas dimensiones =

5) Momios relativos

Cociente, ~~Proporción~~ de casos con una cierta característica $\{i^*, j^*\}$ relativo a la proporción de casos con otra característica $\{i', j'\}$, en términos de momios; i.e.

$$R\,\text{odd}\left(\{i^*, j^*\}, \{i', j'\}\right) = \frac{\#\{y_i : y_{i1} \in \{i^*\},\ y_{i2} \in \{j^*\}\} \Big/ \#\{y_i : y_{i1} \notin \{i^*\},\ y_{i2} \notin \{j^*\}\}}{\#\{y_i : y_{i1} \in \{i'\},\ y_{i2} \in \{j'\}\} \Big/ \#\{y_i : y_{i1} \notin \{i'\},\ y_{i2} \notin \{j'\}\}}$$

$$= \frac{\#\{y_i : y_{i1} \in \{i^*\},\ y_{i2} \in \{j^*\}\}}{\#\{y_i : y_{i1} \in \{i'\},\ y_{i2} \in \{j'\}\}} \times \frac{\#\{y_i : y_{i1} \notin \{i'\},\ y_{i2} \notin \{j'\}\}}{\#\{y_i : y_{i1} \notin \{i^*\},\ y_{i2} \notin \{j^*\}\}}$$

Paradoja de Simpson

La asociación entre dos variables puede cambiar o incluso invertirse cuando existe otra variable que interactúa fuertemente con ~~otra vari~~ ambas variables.

<u>Motivación</u>: The Civil Rights Act 1964 (EEUU)

House Vote	Democrat Yes	Democrat No	Republican Yes	Republican No
Northern	145	9	138	24
Southern	7	87	0	10
Total	152	96	138	34

Senate Vote	Democrat Yes	Democrat No	Republican Yes	Republican No
Northern	45	1	27	5
Southern	1	20	0	1
Total	46	21	27	6

<u>Preguntas</u>:

P1) % Demócratas en House votan a favor?

P2) % Republicanos — — — ?

P3) % Demócratas en Senado — — ?

P4) % Republicanos — — — ?

P5) En cada cámara, qué partido votó proporcionalmente más a favor?

Split the results by region?
Extend answers accordingly.