Guttmans

Asociación de variables en escala nominal

Supongamos que trabajaos con la siguientes variables nominales: el sexo y el hecho de fumar o no. Tenemos una muestra de 98 individuos con la siguiente tabla de contingencia:

$Sexo \diagdown Fumar$	Si	No	Total
Masculino	30	25	55
Femenino	15	28	43
Total	45	53	98

Si debieramos adivinar el sexo de un individuo de la muestra diriamos "Masculino" (pues es el sexo mayoritario en la muestra, la clase modal de la variable sexo) teniendo la posibilidad de equivocarnos en 43 de los 98 casos. Observemos si la otra variable nos entrega información respecto a la variable sexo. Supongamos que sabemos si el individio fuma o no y adivinamos suy sexo. Si fuma diremos que es "Masculino" equivocándonos en 15 casos (de los 45 fumadoires) si no fuma diremos que su sexo es "Femenino" con un error de 25 casos. Conociendo si fuma o no , el error de adivinar el sexo es ahora (15 + 25) casos , con lo cual hemos podido reducir el error total de 43 a 40 casos.

Podemos definir un coeficiente de asociacion entre ambas variables consistentes en el cuociente de asociacion entre ambas variables consistentes en el cuociente entre la reduccion del error y el error total primitivo.

La variable fumar en este ejemplo hace el rol de **variable predictora** y la variable "sexo" el rol de **variable dependiente**.

Llamaremos **coeficiente de preficcion de Guttmans** a:

$$\lambda \;=\; \frac{reduccion\ en\ el\ error}{error\ original\ total}$$

ejemplo:

$$\lambda \;=\; \frac{43-40}{43} \;=\; \frac{3}{43}$$

OBS.:

1. $\mathbf{0} < \lambda < \mathbf{1}$ Si λ se acerca a cero, la variable predictora reduce muy poco el error en la variable dependiente y su asociacion entre ambas es pequeña.
2. Si las variables intercambian su rol es probable que la nueva variable predictora provoque una fuerte reduccion en el error de la nueva variable dependiente

En el ejemplo si intercambiamos el rol de las variables de modo que el "sexo" sea variable predictora y "fumar o no" sea la variable dependiente se tiene:

- Error original: 45 casos
- Error conociendo la variable "sexo"
- Si el sexo es masculino: 25 casos
- Si el sexo es Femenino: 15 casos
- Total: 40 casos

Entonces:

$$\lambda \;=\; \frac{45-40}{45} \;=\; \frac{5}{45} \;=\; \frac{1}{9}$$

Algo mejor que cuando las variables desarrollaban suy rol primitivo
El coeficiente de predicción de **Guttmans** se puede generalizar a tablas de contingencia mayores de 2 x 2 a través de la siguiente fórmula:

$$\lambda \;=\; \frac{\sum n_{i*} \;-\; n_{*d}}{n \;-\; n_{*d}}$$

- n_{i*} frecuencia absoluta maxima dentro de cada subclase de la variable predictora
- n_{*d} frecuencia absoluta maxima dentro de cada subclase de la variable dependiente
- n numero de individuos

En el ejemplo (con el rol inicial de las variables)

$$\lambda \;=\; \frac{(30+28)-55}{98-55} \;=\; \frac{3}{43}$$

OBS.: Para superar el problema de cual es la variable predictoria y cual es la dependiente, usualmente se utiliza el coeficiente de **Guttmans**haciendo a ambas variables jugar los dos roles.

$$\lambda \;=\; \frac{Red\ en\ el\ error\ sist.\ 1 \;+\; Red\ en\ el\ sist.\ 2}{errororiginal\ sist,1 \;+\; error\ original\ sist.\ 2}$$

En nuestro ejemplo resultaría:

$$\lambda \;=\; \frac{3+5}{43+45} \;=\; \frac{8}{88} \;=\; \frac{1}{11}$$