ndientes.

en el caso de asociación perfecta
iación perfecta implícita.

se verifica que $V = T$, y en tablas

ue $V^2 = \rho^2$.

uevo estimar la varianza de la dis-
ada por

$$\left.\frac{}{(J-1)\})^{1/2}\hat{V}}\right)^2 \hat{\sigma}^2(\hat{\phi}^2),$$

onfianza para V al nivel $(1-\alpha)$:

$r(\hat{V})$.

las las medidas basadas en el es-
ión relativas en el sentido de que
de la asociación entre tablas di-
stas medidas son muy sensibles
tabla están desequilibrados, tal
i las tablas tienen distribuciones

abla de contingencia 4.8 para estu-
udios y la opinión sobre las drogas.

didas de asociación basadas en

48; $\hat{V} = 0.181$.

ciación global estimado con es-

4.5. Medidas basadas en Reducción Proporcional en el Error

Son medidas asimétricas basadas en la capacidad de una de las variables para predecir los niveles de clasificación en la otra.

4.5.1. *Definición general*

Consideremos que B es la variable respuesta y A la explicativa. El objetivo es predecir la categoría para la variable columna a partir de la categoría para la variable fila. Elegido un individuo al azar, la predicción del nivel j de clasificación en B se puede hacer de dos formas

1. Asumiendo que las variables son estadísticamente independientes, es decir, predecir al azar sin tener en cuenta el nivel de clasificación en A. Denotaremos por P_1 a la probabilidad de error mediante esta predicción.
2. Asumiendo que B es función de A, es decir, teniendo en cuenta el nivel de clasificación en A. La probabilidad de error al predecir de este modo se denotará por P_2.

Una medida de reducción proporcional en el error se define como

$$RPE = \frac{P_1 - P_2}{P_1}, \tag{4.9}$$

y mide la reducción en el error que se produce al tener en cuenta la clasificación en la variable explicativa. Por lo tanto, el grado de asociación dado por estas medidas representa la ganancia (mejora) relativa al predecir la variable B cuando la categoría de A es conocida en lugar de ser desconocida.

4.5.2. *Medida Lambda de Goodman y Kruskall asimétrica*

Goodman y Kruskal propusieron, basándose en esta definición, la medida *lambda de Goodman y Kruskal* cuyo valor poblacional es

$$\lambda_{B|A} = \frac{(1 - p_{.m}) - \left(1 - \sum_{i=1}^{I} p_{im}\right)}{1 - p_{.m}} = \frac{\sum_i p_{im} - p_{.m}}{1 - p_{.m}},$$

siendo

$$p_{im} = \text{máx}_j(p_{ij}) \quad i = 1, \ldots, I; \quad p_{.m} = \text{máx}_j(p_{.j}).$$

Para obtener esta expresión se calculan P_1 y P_2 como sigue y se sustituye en la ecuación general (4.9) para una medida de reducción proporcional en el error

- Predecir la columna al azar consiste en elegir aquella columna de mayor probabilidad marginal: $\text{máx}_j(p_{.j})$. En este caso la probabilidad de error es la probabilidad de que el individuo no pertenezca a esa columna, es decir,
$$P_1 = 1 - \text{máx}_j(p_{.j}) = 1 - p_{.m}.$$
- Predecir teniendo en cuenta la clasificación en la fila consiste en hacer una predicción de la clasificación en la variable columna condicionando a la fila. Dada la fila i se elige aquella columna de mayor probabilidad condicionada: $\text{máx}_j(p_{j|i})$, que es equivalente a $\text{máx}_j(p_{ij})$. Por lo tanto, aplicando el teorema de la probabilidad total, se tiene

$$\begin{aligned} P_2 &= \sum_{i=1}^{I} p(\text{fila i})\, p(\text{error}|\text{fila i}) \\ &= \sum_i p_{i.}\left(1 - \frac{p_{im}}{p_{i.}}\right) \\ &= 1 - \sum_i p_{im}. \end{aligned}$$

Interpretación y propiedades

1. $\lambda_{B|A}$ está indeterminada si y solo si la población se clasifica en una columna ($p_{.m} = 1$). En otro caso $0 \le \lambda_{B|A} \le 1$.

2. $\lambda_{B|A} = 0$ si y solo si el conocimiento de la fila no es de ayuda para predecir la columna ($P_1 = P_2$).
3. Si A y B son independientes entonces $\lambda_{B|A} = 0$. Sin embargo, el recíproco no es cierto, es decir, $\lambda_{B|A}$ puede ser cero sin que las variables sean independientes.

   Esta propiedad no es una desventaja porque la medida lambda proporciona una interpretación predictiva de la asociación. Mientras que medidas simétricas de asociación reflejan ausencia de independencia, puede ocurrir que $\lambda_{B|A}$ sea cero para la misma tabla, lo que indica ausencia de asociación predictiva cuando se quiere predecir la columna a partir de la fila.
4. $\lambda_{B|A} = 1$ si y solo si el conocimiento de la fila de clasificación de un individuo determina totalmente la columna en la que se clasifica, es decir, si y solo si cada fila de la tabla contiene como mucho una probabilidad no nula (asociación perfecta implícita de tipo I).
5. $\lambda_{B|A}$ es invariante frente a permutaciones de filas o columnas.

La medida lambda asimétrica tiene la desventaja de ser muy sensible frente a totales marginales desequilibrados, tomando valores excesivamente bajos que hace su utilización inapropiada.

La estimación MV de $\lambda_{B|A}$ se obtiene sustituyendo en su expresión por las proporciones muestrales y tiene la siguiente forma:

$$\hat{\lambda}_{B|A} = \frac{\sum_i n_{im} - n_{.m}}{n - n_{.m}}, \tag{4.10}$$

siendo

$$n_{im} = \text{máx}_j(n_{ij}), \qquad n_{.m} = \text{máx}_j(n_{.j}).$$

Goodman y Kruskal (1963) demostraron que cuando $\hat{\lambda}_{B|A}$ no es ni 0 ni 1, tiene distribución normal asintótica de media $\lambda_{B|A}$ y varianza estimada

$$\hat{\sigma}^2\left(\hat{\lambda}_{B|A}\right) = \frac{\left(n - \sum_i n_{im}\right)\left(\sum_i n_{im} + n_{.m} - 2\sum_i^{*} n_{im}\right)}{(n - n_{.m})^3}$$

donde $\sum_i^* n_{im}$ es la suma de las máximas frecuencias en aquellas filas en las que n_{im} cae en la misma columna que $n_{.m}$.

4.5.3. *Medida Lambda de Goodman y Kruskall simétrica*

Para el caso en que no es posible distinguir de forma objetiva entre factor explicativo y respuesta se define la medida de asociación *lambda de Goodman y Kruskal simétrica muestral* en la forma:

$$\hat{\lambda} = \frac{\sum_i n_{im} + \sum_j n_{mj} - n_{.m} - n_{m.}}{2n - n_{.m} - n_{m.}} \qquad (4.11)$$

donde

$$n_{mj} = \text{máx}_i(n_{ij}) \quad j = 1, \ldots, J; \qquad n_{m.} = \text{máx}_i(n_{i.})$$

Ejemplo 4.5. *Consideremos de nuevo la Tabla de contingencia 4.8 en la que se puede considerar el nivel de estudios como variable explicativa de la opinión respecto a la legalización de las drogas.*

La medida lambda que mide el grado de asociación predictiva entre ambas variables se estima en la forma

$$\hat{\lambda}_{B|A} = \frac{237 + 426 + 138 - 801}{1425 - 801} = 0,$$

y se interpreta diciendo que conocer el nivel de estudios no mejora la predicción de la opinión con respecto a la liberalización de las drogas.

Finalmente, el valor estimado de la lambda simétrica es también muy pequeño: $\hat{\lambda} = 0.043$.

4.6. Medidas de Proporción de Varianza Explicada

Este tipo de medidas son para las variables cualitativas el concepto análogo al coeficiente de determinación para variables continuas. Considerando una variable como respuesta y la otra como