

Métodos Estadísticos de la Ingeniería
Tema 12: Contrastes de Hipótesis no Paramétricos
Grupo B

Área de Estadística e Investigación Operativa
Licesio J. Rodríguez-Aragón

Mayo 2010

Contenidos	2
Introducción	3
Contraste χ^2 de Pearson	4
Análisis de Tablas de Contingencia	5

Contenidos

- Pruebas de Bondad de Ajuste.
- Pruebas de Independencia.
- Pruebas de Homogeneidad.

Para estimar los parámetros de una población obtenemos una muestra aleatoria simple de una distribución que suponemos, salvo por sus parámetros, conocida.

Licesio J. Rodríguez-Aragón

Tema 12, M.E.I. – 2 / 5

Introducción

En esta etapa, estudiaremos si la población cumple las hipótesis consideraras a la hora de estimar los parámetros.

- Si la distribución es consistente con los datos de la muestra: Las técnicas empleadas hasta aquí suponen en muchos casos la normalidad de la población, especialmente las inferencias respecto a varianzas.
- Si las observaciones son independientes: Cuando observaciones próximas (en el espacio, en el tiempo, etc.) tienden a ser parecidas, en estos casos las expresiones para la estimación de las varianzas propuestas son erróneas.
- Si la muestra es homogénea: En caso contrario, las observaciones no proveendrán todas de la misma población, por existencia de clases de elementos o por errores en la toma de muestras.

Licesio J. Rodríguez-Aragón

Tema 12, M.E.I. – 3 / 5

Contraste χ^2 de Pearson

H_0 : Los datos de la variable X provienen de una distribución dada, eg. $\mathcal{N}(3, 4)$.

O bien,

H_0 : Los datos de la variable X provienen de una distribución dada, eg. $\mathcal{N}(\mu, \sigma^2)$.

Sea X_1, X_2, \dots, X_n , una m.a.s. de una variable continua o discreta, con $n \geq 25$. El contraste se realiza de la siguiente manera:

- Agrupar las n observaciones en k clases, $k \geq 5$. Las clases han de cubrir todo el rango de valores, conviene que cada clase tenga el mismo número aproximado de observaciones y al menos tres observaciones en cada clase.

n_i , **frecuencia observada** en la clase i .

- Calcular la probabilidad real que el modelo asigna a cada clase, p_i . Como estas probabilidades cubren todo el rango de la variable $\sum p_i = 1$.

$E_i = n \cdot p_i$, **frecuencia esperada** en la clase i .

- Calcular la discrepancia:

$$X^2 = \sum_{i=1}^k \frac{(n_i - E_i)^2}{E_i} \equiv \chi_g^2$$

Rechazaremos H_0 cuando la probabilidad de obtener una discrepancia mayor o igual que la observada, sea suficientemente baja, $\alpha = 1 - \gamma$. Es decir,

$$X^2 \geq \chi_{g, 1-\alpha}^2$$

Los grados de libertad g de la χ^2 son:

- Si el modelo especifica completamente las probabilidades p_i , $g = k - 1$.
- Si las probabilidades p_i , se han calculado estimando r parámetros por máxima verosimilitud, $g = k - r - 1$.

Análisis de Tablas de Contingencia

Supongamos que estudiamos el número de veces que aparecen k posibles atributos mutuamente excluyentes, A_1, A_2, \dots, A_k en una muestra de n elementos, y que los elementos pueden clasificarse en c grupos disjuntos, C_1, C_2, \dots, C_c .

	G_1	...	G_j	...	G_c	Totales
A_1	n_{11}	...	n_{1j}	...	n_{1c}	$n_{1.}$
\vdots	\vdots		\vdots		\vdots	\vdots
A_i	n_{i1}	...	n_{ij}	...	n_{ic}	$n_{i.}$
\vdots	\vdots		\vdots		\vdots	\vdots
A_k	n_{k1}	...	n_{kj}	...	n_{kc}	$n_{k.}$
Totales	$n_{.1}$...	$n_{.j}$...	$n_{.c}$	n

H_0 : Los grupos no influyen y la muestra es homogénea

Bajo la hipótesis de independencia se tendría,

$$P(A_i \cap G_j) = P(A_i) \cdot P(G_j)$$

Construyéndose la tabla de **frecuencias esperadas**,

$$E_{ij} = \frac{n_{i.} \cdot n_{.j}}{n}$$

Calculándose la discrepancia:

$$X^2 = \sum_{i=1}^k \sum_{j=1}^c \frac{(n_{ij} - E_{ij})^2}{E_{ij}} \equiv \chi_g^2$$

Rechazaremos H_0 cuando la probabilidad de obtener una discrepancia mayor o igual que la observada, sea suficientemente baja, $\alpha = 1 - \gamma$. Es decir,

$$X^2 \geq \chi_{g,1-\alpha}^2$$

Los grados de libertad g de la χ^2 son:

- Si el modelo se especifica completamente, $g = (k - 1) \times (c - 1)$.
- Si se han estimado r parámetros por máxima verosimilitud, $g = (k - 1) \times (c - 1) \times (r - 1)$.