TEST DEL CHI QUADRO

PER MISURARE IL GRADO DI DIPENDENZA DI DUE CARATTERI X ED Y OCCORRE CONFRONTARE LA TABELLA DELLE FREQUENZE OSSERVATE CON QUELLA TEORICA DI INDIPENDENZA; IL GRADO DI DIPENDENZA SARÀ TANTO PIÙ ELEVATO QUANTO PIÙ LA TABELLA DELLE FREQUENZE OSSERVATE È LONTANA DALLA TABELLA DELLE FREQUENZE TEORICHE DI INDIPENDENZA. GLI INDICI STATISTICI CHE MISURANO TALE "DISTANZA" SI BASANO SULLE DIFFERENZE TRA LE FREQUENZE OSSERVATE E QUELLE TEORICHE. TALI DIFFERENZE SONO DETTE CONTINGENZE, E SONO DEFINITE COSÌ:

$$\underbrace{c(x_i;y_j)}_{\text{CONTINGENZA DELLA COPPIA } i,j} = \underbrace{f(x_i;y_j)}_{\text{FREQUENZA CONGIUNTA DELLA COPPIA } i,j} - \underbrace{f'(x_i;y_j)}_{\text{FREQUENZA TEORICA DI INDIPENDENZA DELLA COPPIA } i,j}$$

SI PUÒ DIMOSTRARE CHE LA SOMMA DI TUTTE LE CONTINGENZE È SEMPRE NULLA; PERTANTO, NON CI SI PUÒ BASARE SEMPLICEMENTE SULLA SOMMA DELLE CONTINGENZE. L'INDICE STATISTICO PIÙ NOTO È DOVUTO AL MATEMATICO KARL PEARSON (1857–1936) E, COME LA VARIANZA E LO SCARTO QUADRATICO MEDIO, SI BASA SUI QUADRATI DELLE CONTINGENZE. ESSENDO INDICATO CON LA LETTERA GRECA χ^2 (CHI), SI PARLA DI TEST DEL CHI QUADRO. DATI DUE CARATTERI X ED Y E LE DIVERSE MODALITÀ $x_1, x_2, \dots, x_i$ E $y_1, y_2, \dots, y_j$ CON CUI ESSI SI MANIFESTANO, IL CHI QUADRO È COSÌ DEFINITO:

$$\chi^2 = \sum_{i=1}^{k}\sum_{j=1}^{h} \frac{c^2(x_i;y_j)}{f'(x_i;y_j)}$$

IL χ^2 È UGUALE A ZERO SE E SOLO SE X E Y SONO INDIPENDENTI (CIOÈ SE TUTTE LE CONTINGENZE SONO NULLE), E CRESCE AL CRESCERE DELLE CONTINGENZE. PURTROPPO LA FORMULA SOPRA SCRITTA È DI DIFFICILE UTILIZZO; SI PREFERISCE PERCIÒ USARE LA SEGUENTE, AD ESSA EQUIVALENTE, CHE HA IL VANTAGGIO DI EVITARE IL CALCOLO DELLE CONTINGENZE E DELLE FREQUENZE TEORICHE DI INDIPENDENZA:

$$\chi^2 = N\left(\sum_{i=1}^{k}\sum_{j=1}^{h} \frac{f^2(x_i;y_j)}{f(x_i)\,f(y_j)} - 1\right)$$

L'INDICE VA NORMALIZZATO, CIOÈ TRASFORMATO IN UN NUMERO COMPRESO TRA 0 E 1 IN MODO DA INTERPRETARLO FACILMENTE; PER QUESTO LO SI DIVIDE PER IL SUO VALORE MASSIMO; E TALE VALORE MASSIMO È DATO DAL PRODOTTO DI N PER IL MINIMO TRA I DUE VALORI $(k-1)$ E $(h-1)$.

ESEMPIO:

LAVORO \ SESSO	M	F	TOT
OCCUPATO	12	13	25
DISOCCUPATO	18	7	25
TOTALE	30	20	50

$$\chi^2 = 50\left(\frac{12^2}{30\cdot 25} + \frac{13^2}{20\cdot 25} + \frac{18^2}{30\cdot 25} + \frac{7^2}{20\cdot 25} - 1\right) = 3$$

$$\text{MAX } \chi^2 = 50\cdot\min(2-1;2-1) = 50\cdot 1 = 50$$

$$\chi^2_{\text{NORMALIZZATO}} = \frac{3}{50} = 0{,}06 = 6\%$$

IL GRADO DI CONNESSIONE È QUINDI MOLTO BASSO.