ALLEGARE A PARTE 9

I LIVELLI TARIFFARI PER OGNI FATTORE DI RISCHIO

Obiettivo del raggruppamento delle determinazioni in livelli è la riduzione del numero delle classi tariffarie che può essere giustificata da motivi commerciali o da motivi di applicabilità della metodologia statistica

I livelli con la Cluster Analysis. Metodo Ward

Si aggregano determinazioni di un fattore di rischio in modo da ottenere classi omogenee rispetto ad una variabile risposta

L'omogeneità è ottenuta con la minima varianza totale interna

Variabile risposta: frequenza sinistri

Fattore di rischio con N determinazioni da ripartire in K livelli

Abbiamo un file di dati in cui per ogni individuo abbiamo il tempo di esposizione f_i, la determinazione h del fattore di rischio ed il numero di sinistri n_i

Possiamo ottenere le esposizioni complessive $g_1,\ldots,g_N$:

$g_h = \sum_i g_{ih}$ dove la somma è estesa a tutti gli individui con determinazione h del fattore di rischio

Per l'i-esimo individuo della classe con determinazione h la frequenza sinistri è: $f_{ih} = \dfrac{n_{ih}}{g_{ih}}$

con n_{ih} e g_{ih} numero dei sinistri ed esposizione

La frequenza sinistri del gruppo con determinazione h è:

$$f_h = \sum_{i=1}^{m_h} \frac{g_{ih}}{g_h} f_{ih},$$

e la varianza interna:

$$\sigma_h^2 = \sum_{i=1}^{m_h} \frac{g_{ih}}{g_h} (f_{ih} - f_h)^2 .$$

La varianza totale interna è

$$_N\sigma^2 = \sum_{h=1}^{N} \sigma_h^2 \frac{g_h}{g} .$$

S.I.F.A. s.r.l.

Il metodo di Ward è di tipo aggregativo ed il generico passo del procedimento è il seguente:

supposto di avere una ripartizione in K classi, con varianza totale interna ${}_K\sigma^2 = \sum_{h=1}^{K} {}_K\sigma_h^2 \frac{{}_K g_h}{g}$, ${}_K g_h$ essendo l'esposizione nella *h*-esima classe della ripartizione, si passa a K-1 classi scegliendo l'aggregazione di due classi che rende minima la varianza totale interna ${}_{K-1}\sigma^2$, con ${}_{K-1}\sigma^2 = \sum_{h=1}^{K-1} \frac{{}_{K-1} g_h}{g} {}_{K-1}\sigma_h^2$.

Il numero finale K di clusters può essere prefissato o individuato tramite i valori di una funzione g(K) che misura la perdita di informazione connessa al raggruppamento delle determinazioni della variabile tariffaria in K cluster.

Di solito si prende

$$g(K) = \frac{{}_{(K)}\sigma^2}{\sigma^2}$$

con ${}_{(K)}\sigma^2 = \sigma^2$

e si sceglie il massimo K tale che $g(K) \geq 0.95$.

Con due variabili risposta A e B indipendenti, dopo averle standardizzate, si considera la varianza totale interna ${}_K\sigma_A^2 + {}_K\sigma_B^2$ e si passa da K a K-1 classi raggruppando le due classi che danno il minimo di ${}_{K-1}\sigma_A^2 + {}_{K-1}\sigma_B^2$.

Consideriamo ad esempio la frequenza sinistri ed il costo medio.

E' necessario procedere preventivamente ad una standardizzazione delle variabili risposta.

La frequenza sinistri standardizzata è data

$$f'_i = \frac{f_i - \bar{f}}{\sigma_f} \qquad \text{dove } \sigma_f^2 = \sum_{i=1}^{N} (f_i - \bar{f})^2 \frac{g_i}{g},$$

mentre i costi dei sinistri standardizzati sono

$$c'_i = \frac{c_i - \bar{c}}{\sigma_c} \qquad \text{dove } \sigma_c^2 = \sum_{i=1}^{N} (c_i - \bar{c})^2 \frac{n_i}{n}.$$

La varianza totale interna si definisce come somma delle varianze totali interne misurate per le singole variabili risposta.

Usando come variabile risposta il costo medio, il procedimento di aggregazione è fortemente influenzato da eventuali sinistri eccezionali. Occorre ridurre il loro effetto.

S.I.F.A. s.r.l.

L'esigenza di ottenere classi omogenee per sinistrosità può essere mediata da esigenze di tipo commerciale. Ad esempio quella di ottenere livelli con determinazioni vicine (es. età 18-25)

Per forzare il procedimento a raggruppare determinazioni contigue si può decidere di inserire tra le variabili risposta la variabile esplicativa stessa.

BIBLIOGRAFIA

Anderberg M.R. (1973), Cluster Analysis for Applications, Academic Press, New York.

Dickmann H. (1978), Einsatz der clusteranalyse bei klassifikationsproblemen in der versicherungswirtschaft, Blatter der Deutschen Gesellschaft fur Versicherungsmathemitik, Vol. XIII, n. 4.

Van Eeghen J., Greup E.K., Nijssen J.A. (1983), Rate Making, Surveys of Actuarial Studies, 2, Nationale Nederlande N.V.

Ward J.H. (1963), Hierachical grouping to optimise an objective function, J. Am. Statist. Ass., 58.

1,3

ANALISI STATISTICHE PRELIMINARI

Sui dati di portafoglio

Mettere in relazione la sinistrosità con i fattori di rischio

Dapprima su base univariata:

Ad esempio: esprimere la frequenza sinistri per le varie età

Successivamente su base multivariata

Ad esempio: considerati due fattori di rischio correlati (in base all'esperienza) come età e potenza del veicolo si esprime la frequenza sinistri per le varie età e per ogni livello di potenza

VARIABILI QUALITATIVE, NUMERICHE E VARIABILI CON DETERMINAZIONI SUDDIVISE IN CLASSI (LIVELLI)

Variabili (o fattori di rischio) quali sesso, professione, stato civile, zona di residenza,.., sono dette nominali o qualitative.

Non hanno una determinazione numerica "naturale".

Consideriamo una codifica numerica

Esempio. Possiamo codificare

$$\text{sesso} = \begin{cases} 1 & \text{se femmina} \\ 0 & \text{se maschio} \end{cases}$$

Esempio. Stato civile con le quattro determinazioni di single, coniugato/a, vedovo/a, divorziato/a. Definiamo

$$X_1 = \begin{cases} 1 \text{ se single} \\ 0 \text{ altrimenti} \end{cases}, \quad X_2 = \begin{cases} 1 \text{ se coniugato/a} \\ 0 \text{ altrimenti} \end{cases},$$

$$X_3 = \begin{cases} 1 \text{ se vedovo/a} \\ 0 \text{ altrimenti} \end{cases}, \quad X_4 = \begin{cases} 1 \text{ se divorziato/a} \\ 0 \text{ altrimenti} \end{cases}.$$

Dato che $X_1 + X_2 + X_3 + X_4 = 1$, possono bastare le determinazioni di X_1, X_2, X_3 per determinare X_4. Le terne (1,0,0), (0,1,0), (0,0,1), (0,0,0) identificano rispettivamente la qualifica di single, coniugato/a, vedovo/a, divorziato/a.

In generale, data una variabile nominale con k determinazioni, possiamo descriverla con k-1 variabili 0-1, nel modo indicato nei due esempi precedenti.

Le variabili numeriche possono essere considerate con le loro determinazioni naturali (si parla di variabili continue)

Oppure, soprattutto per ragioni commerciali, si raggruppano le determinazioni in classi dette livelli.

Una volta che le determinazioni sono ripartite in livelli la variabile può essere descritta con variabili 0-1 come per le variabili qualitative. La descrizione con variabili 0-1 può essere fatta in diversi modi come mostra l'esempio seguente.

Esempio . Consideriamo la variabile età con la seguente suddivisione in livelli.

Età
18-29
30-39
40-54
55-68
69-88

Per descriverla possiamo usare le quattro variabili linearmente indipendenti

$$X_1 = \begin{cases} 1 \text{ se } 18 \le x \le 29 \\ 0 \text{ altrimenti} \end{cases}, X_2 = \begin{cases} 1 \text{ se } 30 \le x \le 39 \\ 0 \text{ altrimenti} \end{cases},$$

$$X_3 = \begin{cases} 1 \text{ se } 40 \le x \le 54 \\ 0 \text{ altrimenti} \end{cases}, X_4 = \begin{cases} 1 \text{ se } 55 \le x \le 68 \\ 0 \text{ altrimenti} \end{cases}$$

ma anche le

$$X_1 = \begin{cases} 1 \text{ se } x \le 29 \\ 0 \text{ altrimenti} \end{cases}, X_2 = \begin{cases} 1 \text{ se } x \le 39 \\ 0 \text{ altrimenti} \end{cases},$$

$$X_3 = \begin{cases} 1 \text{ se } x \le 54 \\ 0 \text{ altrimenti} \end{cases}, X_4 = \begin{cases} 1 \text{ se } x \le 68 \\ 0 \text{ altrimenti} \end{cases}$$