Troisièmes Rencontres Internationales – Terzo Convegno Internazionale - Third International Conference
A.S.I. Analyse Statistique Implicative – Analisi Statistica Implicativa – Implicative Statistic Analysis
Palermo (Italy) 6-8 Octobre 2005

L'Analisi Statistica Implicativa : uno dei metodi di analisi dei dati nella ricerca in didattica delle Matematiche

Filippo Spagnolo
GRIM (Research Group for Mathematics Learning) Departmet of Mathematics, University of Palermo, via Archirafi 34, 90123 Palermo (Sicily). E-mail: spagnolo@math.unipa.it

Riassunto
Attraverso l'analisi di esperienze di ricerca in Didattica delle Matematiche sono stati comparati dati sia con l'analisi implicativa di Gras che con l'analsi fattoriale delle coriispondenze. I risultati di questa comparzione mettono in evidenza che l'analisi implicativa, quando vengono introdotte variabili supplementari, risulta essere più incisiva nella lettura della contingenza. L'analisi fattoriale invece può dare un utile contributo argomentativo di supporto all'analisi implicativa quando si debbono analizzare soltanto le variabili in gioco.

Parole Chiavi: Analisi Implicativa, Analisi Fattoriale, Variabili supplementari, ricerca in didattica.

Introduzione

La modellizzazione attraverso argomentazioni statistiche fornisce alla ricerca in didattica delle matematiche una maggiore possibilità di trasferibilità dell'esperienza.

Risulta evidente, come è stato ampiamente dibattuto in G.Brousseau & E. La Casta (1995), R. Gras (2000), F. Spagnolo (1998), che senza una riflessione teorica dal punto di vista della didattica e quindi della epistemologia dei contenuti matematici, l'argomentazione statistica non avrebbe alcun peso. Soltanto uno studio in parallelo di tutti i possibili percorsi argomentativi della ricerca può portare a risultati considerati attendibili.

1.0 La ricerca in didattica, alcuni strumenti

La Ricerca in Didattica si pone come un meta-paradigma rispetto ad altri paradigmi di ricerca in scienze dell'educazione in quanto utilizza sia il paradigma della disciplina oggetto di analisi che il paradigma delle scienze sperimentali. La Ricerca in Didattica può essere considerata come una sorta di "Epistemologia Sperimentale".

Lo strumento fondamentale per la ricerca in didattica é l'analisi a-priori di duna situazione didattica.

Cosa si intende per "analisi a-priori" di una situazione didattica?

Per analisi a-priori si intende una analisi delle "Rappresentazioni Epistemologiche", "Rappresentazioni Storico-epistemologiche", "Comportamenti ipotizzabili", corretti e non, per la risoluzione della data situazione didattica.

- *Le rappresentazioni epistemologiche sono le rappresentazioni degli eventuali percorsi conoscitivi*[1] *riguardo un particolare concetto. Tali rappresentazioni possono essere messe a punto da un soggetto apprendente o da una comunità scientifica in un determinato periodo storico.*

[1] I percorsi conoscitivi permettono di evidenziare le reti concettuali riguardanti la situazione didattica.

- *Le rappresentazioni storico-epistemologiche sono le rappresentazioni degli eventuali percorsi conoscitivi riguardanti la ricostruzione sintattica, semantica, pragmatica*[2] *di un determinato concetto.*
- *I comportamenti ipotizzabili dell'allievo nei confronti della situazione/problema sono tutte le possibili strategie*[3] *risolutive sia corrette che non. Tra le strategie non corrette verranno prese in considerazione quelle che possono devolvere in strategie corrette.*

L'analisi a-priori della situazione didattica consente di:

- individuare lo "*spazio degli eventi*"[4] riguardanti la particolare situazione didattica rispetto alle conoscenze professionali dell'insegnante ricercatore in un determinato periodo storico;
- individuare, attraverso lo spazio degli eventi possibili, il "buon problema"[5] e quindi una "situazione didattica fondamentale" per la classe di problemi alla quale la situazione didattica afferisce;
- individuare delle variabili della situazione problema e delle variabili didattiche[6];
- individuare delle ipotesi di Ricerca in Didattica di tipo più generale rispetto a quelle analizzabili da una prima analisi della situazione/problema.

L'analisi a-priori rappresenta quindi l'elemento fondante e tiene conto sia dell'epistemologia della disciplina che della sua storia. Riassumiamo brevemente quelle che possono essere le tappe significative di una ricerca in didattica, rimandando ad altre letture eventuali approfondimenti[7]:

[2] La prospettiva semiotica per l'analisi delle conoscenze disciplinari consente una gestione dei contenuti con riferimento ai problemi della "comunicazione" di tali contenuti. Questa posizione non è particolarmente nuova rispetto alle scienze umane, ma rappresenta una vera innovazione per le discipline tecniche e scientifiche.

[3] Una situazione didattica pone comunque un "problema" da risolvere all'allievo, vuoi come problema tradizionale (ad es. nell'ambito scientifico o matematico), vuoi come "strategia" per organizzare la conoscenza migliore per adattarsi ad una situazione.

[4] Per "spazio degli eventi" si intende l'insieme delle possibili strategie risolutive corrette e non ipotizzabili in un determinato periodo storico da una determinata comunità di insegnanti.

[5] Il "buon problema" è quello che, rispetto alla conoscenza presa in esame, permette la migliore formulazione in termin ergonomici.

[6] Le "variabili della situazione didattica" sono tutte le possibili variabili che intervengono, le "variabili didattiche" sono quelle che permettono un cambiamento dei comportamenti degli allievi. Le variabili didattiche sono quindi un sotto insieme delle variabili della situazione didattica.

[7] Guy Brousseau, *Théorie des situations didactiques*, La Pensée sauvage, Grenoble, 1998.
Filippo Spagnolo, *Insegnare le matematiche nella scuola secondaria*, La Nuova Italia, Firenze, 1998.

Troisièmes Rencontres Internationales – Terzo Convegno Internazionale - Third International Conference
A.S.I. Analyse Statistique Implicative – Analisi Statistica Implicativa – Implicative Statistic Analysis
Palermo (Italy) 6-8 Octobre 2005

Che cosa è la Ricerca in Didattica?

- Ricerca con un suo Paradigma;
- Ricerca con un linguaggio proprio;
- Ricerca Teorica: Analisi epistemologica e storico-epistemologica della disciplina relativa ad un determinato Sapere;
- Ricerca Sperimentale:
    1. Analisi a-priori della situazione-problema;
    2. Individuazione delle ipotesi di Ricerca;
    3. Falsificazione delle ipotesi[8];
    4. Analisi di dati sperimentali relativi a piccoli campioni attraverso strumenti statistici appropriati;
    5. Analisi a-posteriori dei dati sperimentali.

A cosa serve la Ricerca in Didattica?

- Previsione di "fenomeni didattici" attraverso "Modelli attendibili" rispetto alla Ricerca Teorico-Sperimentale. Per "Modelli attendibili" si intendono quei Modelli che consentono la possibilità di far previsioni sui fenomeni didattici;
- Comunicazione dei risultati della Ricerca alla comunità degli Insegnanti attraverso argomentazioni forti come l'analisi a-priori e gli strumenti statistici.

Di che cosa si occupa la Ricerca in Didattica?

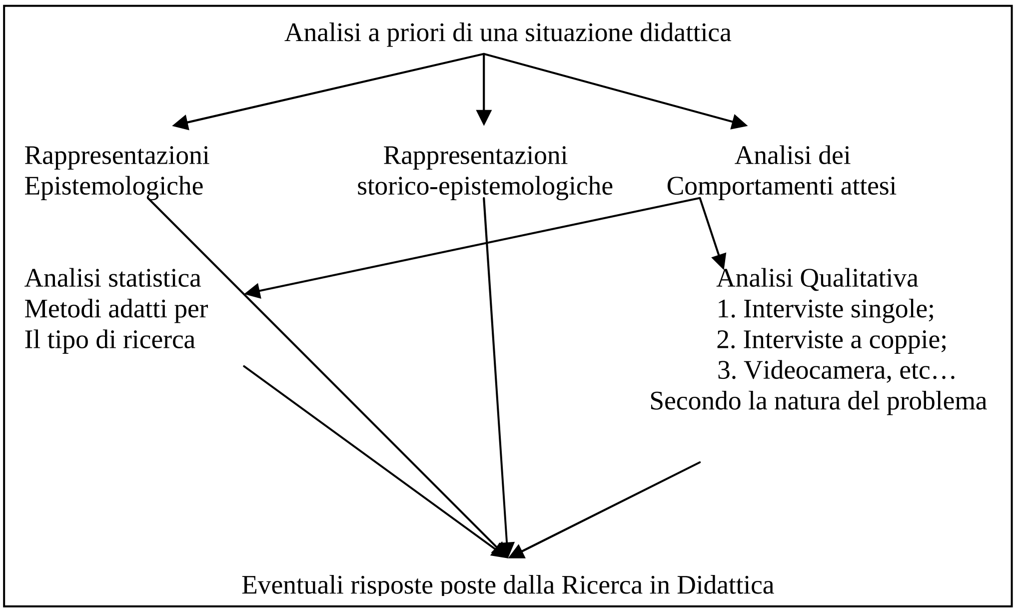
- Problemi riguardanti la "comunicazione di una determinata disciplina" attraverso:
    1. Messa a punto di situazioni a-didattiche appropriate;
    2. Analisi degli errrori ed ostacoli derivanti dai processi comunicativi;
    3. Studio degli ostacoli didattici ed epistemologici[9] come:
        - strumenti per la riflessione sulla costruzione di curricula didattici;
        - strumenti per una migliore e più profonda comprensione dei processi comunicativi;
        - strumenti per la messa a punto di situazioni a-didattiche.

[8] Una ipotesi si dice falsificabile se, sottoposta a verifica sperimentale, può essere messa a dura prova da tentativi sistematici per coglierla in fallo.

[9] Ostacoli didattici: Sono gli ostacoli che si determinano da una non pertinente trasposizione didattica.
Ostacoli epistemologici: Sono gli ostacoli che hanno un ruolo costitutivo nella conoscenza. Essi sono difficili da evidenziare e difficili da superare. Lo studio di questi ostacoli è però molto importante per l'insegnante perché questi è costretto a mettere in discussione continua le sue conoscenze sia epistemologiche che comunicative.
Gli errori possono mettere in evidenza sia gli ostacoli didattici che epistemologici, essi rappresentano lo strumento indispensabile per capire le situazioni patologiche. Altra cosa sarà il superamento degli ostacoli.

Troisièmes Rencontres Internationales – Terzo Convegno Internazionale - Third International Conference
A.S.I. Analyse Statistique Implicative – Analisi Statistica Implicativa – Implicative Statistic Analysis
Palermo (Italy) 6-8 Octobre 2005

1.1 Il seguente schema riassume la relazione tra ricerca in didattica e raccolta di dati sperimentali.



2.0 I dati.

Ogni ricerca didattica ci porta inevitabilmente a raccogliere dei dati che possiamo considerare formati da una collezione di informazioni elementari. Ogni informazione elementare riporta in generale un comportamento di un allievo in una situazione. Una statistica sarà quindi un insieme composto da: allievo, situazione, comportamento.

L'allievo appartiene ad un campione E osservato, supposto estratto da una popolazione più vasta, o a caso, o seguendo un sistema di situazioni di controllo (ad esempio: livello scolare, sesso, conoscenza personale anteriore...).

La situazione è scelta in un insieme S (di questioni, esercizi...) generata e strutturata da condizioni e parametri di varia natura (sapere in gioco, condizioni materiali, condizioni didattiche...).

I comportamenti (tipici delle conoscenze o di saperi mirati) sono presi in un insieme C di risposte possibili dell'allievo nelle condizioni nelle quali è posto.

Una classe può essere definita come un insieme di allievi E, un corso di Matematica come un insieme di esercizi S, i risultati degli allievi come una certa applicazione di E nell'insieme SxC dove C è l'insieme dei comportamenti di riuscita o di errore, una nota come una applicazione di SxC in R.

La conoscenza di un certo comportamento potrà essere rappresentato da una certa applicazione di un insieme di questioni in un insieme di comportamenti.

2.1 Utilizzazione della statistica da parte degli insegnanti e da parte dei ricercatori.

L'insegnante deve prendere rapidamente numerose decisioni e può correggerle molto velocemente se si dovessero rilevare inadatte. Non può aspettare il risultato del trattamento statistico di tutte le sue questioni. L'insegnante deve cercare di utilizzare quei trattamenti statistici che gli consentono rapidamente di trarre certe conclusioni.

Il ricercatore deve seguire un processo opposto:

- Quali ipotesi corrispondono alle questioni che ci interessano?
- Quali dati raccogliere?
- Quali trattamenti utilizzare?
- Quali conclusioni?

Più che la rapidità e l'utilità immediata, è la consistenza, la stabilità, la pertinenza e la sicurezza delle risposte che interessano il ricercatore.

La ricerca con opportuni metodi statistici consentirà:

- di comunicare tra insegnanti le informazioni di cui hanno bisogno e che raccolgono sui risultati degli allievi, il valore dei metodi impiegati...;
- di utilizzare anche con discernimento i risultati delle ricerche in didattica;
- di conoscere le possibilità e i limiti dei metodi statistici e quindi la legittimità delle conoscenze che essi utilizzano nella loro professione;
- di discutere questa legittimità;
- di formulare delle congetture suscettibili di essere sottomesse alla prova della contingenza sperimentale;
- di immaginare la plausibilità di queste congetture;
- di sapere come convertire la loro esperienza in conoscenza;
- di partecipare a delle ricerche.

2.2 Le osservazioni.

Una osservazione consiste in una attribuzione di un valore a una variabile a proposito di un individuo: l'oggetto osservato.

La statistica permette principalmente di trattare il caso dove:

- parecchie osservazioni sono raccolte.
- e dove queste osservazioni nel loro insieme:

i) riguardino individui differenti per una stessa proprietà;
ii) riferiscano proprietà differenti per uno stesso individuo;
iii) riguardino sia la i) che la ii).

Se "24" (valore osservato) è attribuito all'allievo X (oggetto d'osservazione), come "risultato dell'esame di matematica" (variabile osservata), l'insieme dei valori o dei casi possibili nel nostro caso è un intero compreso tra 0 e 30.

Le variabili possono essere numeriche, d'intervallo, ordinali, nominali.

- *Variabile Numerica*: quando i valori sono espressi in numeri (appartenenti agli insiemi N, Z, Q, R) e le operazioni che si possono fare con essi hanno un senso per la variabile.
- *Variabile d'Intervallo*: quando solo le differenze tra valori hanno senso mentre la somma non ne ha. Per esempio il punteggio ottenuto in una disciplina sportiva può costituire una variabile d'intervallo.
- *Variabile Ordinale*: quando i valori esprimono soltanto un ordine tra le osservazioni. In una variabile ordinale la somma di due valori non è un valore.
- *Variabile Nominale*: quando i valori sono dei caratteri o attributi. Questa variabile può essere a due valori: un suo attributo e la sua negazione. Anche se essa è espressa da numeri come 0 e 1, una variabile nominale non è numerica: la somma tra due caratteri non è definita, né il loro ordine in generale. Le sole operazioni sono quelle logiche (insiemistiche).

Una variabile numerica è sempre possibile trasformarla in variabile d'intervallo, ordinale o nominale (perdendo delle informazioni); una Variabile d'intervallo può essere trasformata in variabile ordinale o nominale; una variabile ordinale può essere trasformata in variabile nominale. L'inversa non è vera.

3.0 Analisi fattoriale delle corrispondenze e l'analisi implicativa tra variabili nella ricerca in didattica della matematica: confronto sperimentale.

La ricerca in didattica si serve di strumenti quantitativi e qualitativi. In questa esposizione si cercherà di affrontare il tema degli strumenti quantitativi soffermandoci soprattutto su relazioni teorico-sperimentali tra l'analisi fattoriale e l'analisi implicativa. Verranno analizzati alcune situazioni sperimentali significative. Naturalmente le riflessioni vengono fuori da numerosi lavori sperimentali.

3.1 L'analisi implicativa

Il problema che ha cercato di affrontare R.Gras[10] è stato quello di poter rispondere alla seguente questione: "Date delle variabili binarie **a** e **b**, in quale misura posso assicurare che in una popolazione, da ogni osservazione di **a** segue necessariamente quella di **b**?". O anche in maniera più lapidaria :"E' vero che se a allora b?".

In generale la risposta non è possibile ed il ricercatore si deve accontentare di una implicazione "quasi" vera. Con l'analisi implicativa di R. Gras si cerca di misurare il grado di validità di una proposizione implicativa tra variabili binarie e non. Questo strumento statistico viene messo a punto su ricerche riguardanti la Didattica delle Matematiche.

Viene presentata la modellizzazione del caso binario.

Siano date una popolazione E e un insieme di variabili V, si vuole dare significato statistico alla implicazione larga $\mathbf{a} \Rightarrow \mathbf{b}$.

Siano A e B gli insiemi delle sotto popolazioni rispettive dove la variabile a e b prendono il valore 1 (vero). L'intensità della implicazione viene espressa formalmente:

$$\varphi(a,\overline{b}) = 1 - \Pr ob\left[card\left(X \cap \overline{Y}\right) \le card\left(A \cap \overline{B}\right)\right]$$

[10]Régis Gras, *L'implication statistique..., op. cit.*.

X e Y sono due sotto insiemi di E, parti aleatorie di E e che hanno la stessa cardinalità rispettivamente di A e B. $\overline{Y}$ é il complementare di Y rispetto ad E. $\overline{B}$ é il complementare di B rispetto ad E. $\overline{b}$ rappresenta il fatto di non possedere il carattere b.

E si dirà:

$$\left[a \Rightarrow b \text{ accettabile alla soglia } \varphi(\mathrm{a},\overline{\mathrm{b}}) = 1 - \alpha\right] \Leftrightarrow \Pr ob\left[\left[card(X \cap \overline{Y}\right] \le card(A \cap \overline{B})\right] \le \alpha$$

Non ci dilungheremo sull'analisi implicativi. Molto è già stato pubblicato e molti riferimenti si trovano negli atti di questo convegno

3.2 Alcune osservazioni sull'analisi Fattoriale*.*

Gli approcci all'analisi fattoriale sono di due tipi, il primo attraverso lo studio degli autovalori di equazioni, il secondo attraverso una interpretazione geometrica (vettori) e con contributi della meccanica razionale.

L'approccio che qui viene presentato è il secondo.

Consideriamo il prodotto cartesiano E (in generale allievi, n, n∈N) e V (in generale variabile m∈N). Questa è una tipica situazione di rilevazione di dati in didattica. Si pone il problema di poter rappresentare geometricamente in uno spazio ad nxm dimensioni la distribuzione dei due insiemi. L'analisi fattoriale interpreta le rappresentazioni geometriche. Sorta nell'ambito delle Scienze Umane ha avuto parecchie applicazioni nel campo della Psicologia, ma consentendo una analisi su piccoli campioni nel campo della Statistica non parametrica contribuisce ad interpretare significativamente i fenomeni didattici. In questo paragrafo introdurremo il metodo senza entrare molto nel particolare in quanto richiederebbe una trattazione a parte.

Il punto di partenza è una tabella a doppia entrata I x J del tipo:

I \ **J**	**j**	Colonna di margine
i	**k(i,j)**	**k(i)** [totale della linea i]
Linea di margine	**k(j)** [totale della colonna j]	**k** [totale]

- $\mathbf{f_i = k(i)/k}$ **massa della linea i;**
- $\mathbf{f_j = k(j)/k}$ **massa della colonna j.**

  La massa di un elemento i o j misura l'importanza relativa di questo elemento.
- $\mathbf{f_j^i} = \{\mathbf{f_j^i} \ / \ \mathbf{j \in J}\}$, insieme degli f_j^i per j che percorre J. **Profilo di una linea.** $\mathbf{f_j^i = k(i,j)/k(i)}$ è la parte relativa, la porzione di j nella i^{esima} linea, $\mathbf{f_j^i}$ insieme di tutti i termini corrispondenti ai diversi elementi di J, è la composizione di questa linea.

- **$f_i^j = \{ f_i^j \ / \ i \in I \}$, Profilo di una colonna.**

Spazio dei profili su J: Un punto di questo spazio è un profilo su J, cioè un insieme di numeri positivi o nulli indicizzati da J e di totale 1.

$\pi_j = \{\pi_j \ / \ j \in J\}, \pi_j \in R^+_0, \ \sum\{\pi_j \ / \ j \in J\} = 1.$

Ad ogni elemento j di J corrisponde un termine π_j e uno solo del profilo π_j ; ad ogni termine π_j di π_j corrisponde un elemento e uno solo di J: il suo indice j; un profilo π_j è composto da parametri che variano su J con cardinalità **Card J**, i parametri sono legati dalla relazione d'avere come somma 1. Lo spazio dei profili su J è uno spazio a (**Card J** -1) dimensioni.

Spazio dei profili su I:

$\pi_i = \{\pi_i \ / \ i \in I\}, \pi_i \in R^+_0, \ \sum\{\pi_i \ / \ i \in I\} = 1.$

Lo spazio dei profili su I è uno spazio a (**Card I** -1).

3.2.1 Le rappresentazioni grafiche: le nuvole.

La nuvola N(I):

Nello spazio dei profili su J, ogni linea i della tabella è rappresentata dal suo profilo:

$f_j^i = \{ k(i,j) / k(i) \ / \ j \in J \}$,

al quale si associa la massa di **i : $f_i = k(i)/k$**; l'insieme dei profili delle diverse linee i, ciascuna munita della massa della linea che rappresenta, costituisce la nuvola N(I):

$N(I) = \{ f_j^i, f_i \ / \ i \in I \}$; un elemento della nuvola N(I) è una coppia formata da un profilo di linea e della massa di questa linea.

La nuvola N(J):

Nello spazio dei profili su I, ogni linea i della tabella è rappresentata dal suo profilo:

$f_i^j = \{ k(j,i) / k(j) \ / \ i \in I \}$,

al quale si associa la massa di **j : $f_j = k(j)/k$**; l'insieme dei profili delle diverse linee i, ciascuna munita della massa della linea che rappresenta, costituisce la nuvola N(J):

$N(J) = \{ f_i^j, f_j \ / \ j \in J \}$; un elemento della nuvola N(J) è una coppia formata da un profilo di linea e della massa di questa linea.

3.2.2 Media e centro di gravità; Dispersione e inerzia.

Il centro di gravità di un sistema di punti muniti di massa (numeri positivi o nulli) è una generalizzazione spaziale della nozione di media: il centro di gravità dei punti f_j^i con una massa f_i è come la media della Nuvola N(I), ma una media dove ogni punto f_j^i gioca un ruolo proporzionale alla sua massa f_i.

Per determinare il centro di gravità o baricentro utilizzeremo la nozione di media ponderata[11] e applicheremo questa nozione al nostro caso delle nuvole N(I) e N(J). In questo caso il centro di gravità è una sorta di media spaziale, ogni punto gioca un ruolo proporzionale alla sua massa prendendo la media ponderata delle coordinate asse per asse.

Per N(I), lo spazio ambiente è lo spazio dei profili su J; i punti di N(I) sono i profili su J, ciascuno munito di massa f_i. Il centro di gravità di N(I) è un profilo su J che chiameremo g_j, la sua j^{esima} coordinata g_j è la media ponderata della j^{esima} coordinata di punti di N(I):

---

[11]La media ponderata tra due nuemri x_1 e x_2 di massa rispettivamente m_1 e m_2 sarà definita da: $(m_1x_1+m_2x_2)/(m_1+m_2)$.

$$g_j = \frac{\sum \left\{ f_i f_j^i \,/\, i \in I \right\}}{\sum \left\{ f_i \,/\, i \in I \right\}}$$

Ma siccome la somma degli f_i vale 1, la formula si può anche scrivere:

$$g_j = \sum \left\{ f_i f_j^i \,/\, i \in I \right\}$$

e rimpiazzando in questa espressione f_i e f_j^i con le espressioni in funzione di k(i,J), k(i), K(j) e k avremo:

$$g_j = f_j = \left\{ \frac{k(j)}{k} \,/\, j \in J \right\}$$

Analogamente avremo nello spazio dei profili di I la seguente formula:

$$g_I = f_I = \left\{ \frac{k(i)}{k} \,/\, i \in I \right\}$$

Dispersione della Nuvola e Varianza:

Una nuvola N(I) sarà più o meno dispersa attorno al suo centro di gravità f_j e la sua dispersione può essere calcolata prendendo le medie degli scarti dei diversi punti rispetto al centro di gravità. Verrà definito lo scarto di una nuvola N in rapporto ad un punto P qualunque con la somma $I_P(N)$:

Si consideri una nuvola N di n punti M^i con massa rispettivamente m_i : $N\{(M^i, m_i) / i=1,...,n\}$e P un punto qualunque situato nello stesso spazio di N.

$\mathbf{I_P(M^i,m_i)=m_i(d(P,M^i))^2=m_i d^2(P,M^i)}$ Inerzia del punto (M^i,m_i) rispetto a P, dove $d^2(P,M^i)$ è la distanza tra P e M^i elevata al quadrato.

$$\mathbf{I_P(N)=\sum\{m_i d^2(P,M^i) \,/\, i= 1,...,n\}.}$$

e rispetto al suo centro di gravità:

$$\mathbf{I_G(N)=\sum\{m_i d^2(G,M^i) \,/\, i= 1,...,n\}.}$$

Le inerzie di N rispetto a G e rispetto ad un punto P qualunque sono legate dalla seguente formula dovuta a Huyghens:

$\mathbf{I_P(N)= I_G(N) + m_{tot}d^2(G,P)}$ ($\mathbf{m_{tot}=m_1+m_2+...+m_n}$ massa totale di N)

Cioè l'inerzia della nuvola N rispetto ad un punto qualunque P è eguale all'inerzia della nuvola N rispetto al suo centro di gravità, aumentato della quantità positiva $m_{tot}d^2(G,P)$ *prodotto della massa totale per il quadrato della distanza tra G e P.*

3.2.3 Distanza distribuzionale.

Si definisce una distanza distribuzionale tra profili introducendo un principio di equivalenza distribuzionale.

Consideriamo il seguente esempio. Supponiamo che vi siano una lista di nomi I ed una lista di verbi J. Per ogni coppia (i,j) di un nome i e di un verbo j si calcola quante volte il nome i è soggetto del verbo j. Nella tabella dei dati, ad ogni nome i di I corrisponde una linea che chiameremo anche i, e ad ogni verbo j di J corrisponde una colonna j. All'incrocio della linea i e della colonna j è scritto il numero di volte, k(i,j), che il nome i è stato trovato soggetto del verbo j.

Che significato dobbiamo attribuire a due verbi che hanno lo stesso profilo ?

Essi hanno gli stessi soggetti con le stesse frequenze relative. I due verbi j e j' li chiameremo sinonimi distribuzionali.

In N(J) sono rappresentati dai punti f_I^j e $f_I^{j'}$ che coincidono rispettivamente con le masse f_j e $f_{j'}$. Nell'analisi è come se si considerasse un solo verbo j_0 che ingloba gli usi di j e j' con massa $f_{j0}=f+f_{j'}$.
Cosa avviene alla nuvola N(I) quando la tabella è modificata cumulando le due colonne in una sola?

La formula della distanza sarà allora del tipo: $\alpha_j=1/f_j$.

La formula della distanza distribuzionale tra due linee nello spazio dei profili su J (N(I)):

$$d^2(f_J^i, f_J^{i'}) = \sum\left\{\left(\frac{1}{f_j}\right)\left(f_j^i - f_j^{i'}\right)^2 / j \in J\right\}$$

La formula della distanza distribuzionale tra due colonne nello spazio dei profili su I (N(J)):

$$d^2\left(f_I^j, f_I^{j'}\right) = \sum\left\{\left(\frac{1}{f_i}\right)\left(f_i^j\, f_i^{j'}\right)^2 / i \in I\right\}$$

3.2.4 Gli assi principali d'inerzia.

La ricerca degli assi principali d'inerzia di una nuvola di punti muniti di massa in uno spazio Euclideo è pregiudiziale per qualsiasi tipo di analisi Fattoriale. Oggi con il termine "Analisi Fattoriale" si intende l'Analisi delle Componenti Principali (ACP): ricerca degli assi principali d'inerzia. In questa trattazione seguiremo questa impostazione.

L'Analisi Fattoriale delle Corrispondenze (AFC) seguirà l'analisi delle componenti principali.

Dal punto di vista dell'analisi dei dati multidimensionale è importante ridurre la nuvola N ad una rappresentazione accessibile alla nostra visione (cioè di dimensioni accettabili ottenuta per proiezione su una retta o su un piano) e fedele alla complessità del reale (se la dispersione della nuvola proiettata è quasi uguale a quella di N stesso). E' per questo che bisogna considerare non solamente l'inerzia della nuvola N rispetto ad un punto ma anche l'inerzia trasversalmente ad un sotto spazio L (retta o piano), misura dello scarto della nuvola a L, e, correlativamente, l'inerzia lungo un sotto spazio L, misura della fedeltà della rappresentazione della nuvola dalla sua proiezione ortogonale su L. Il teorema di Huyghens gioca un ruolo fondamentale in quanto permette di restringere lo studio dell'inerzia ai sotto spazi passanti per il centro di gravità G di una nuvola N. In particolare alle rette passanti per G o ai piani passanti per G.

Le coordinate f_j^i sugli assi sono i fattori che si studieranno: l'Analisi Fattoriale si occupa del cambiamento delle coordinate e del cambiamento degli assi.

Generalmente per ottenere la massima percentuale d'inerzia totale è sufficiente considerare i primi 4 assi d'inerzia[12]:

- Δ_1 : Asse principale d'inerzia della nuvola N(I);
- Δ_2 : E' tra le rette perpendicolari a Δ_1, quella sulla quale la nuvola si proietta con la più grande dispersione (lungo la quale l'inerzia della nuvola è più grande);
- Δ_3 : E' tra le rette perpendicolari a Δ_1 e Δ_2, quella sulla quale la nuvola si proietta con la più grande dispersione;
- ecc..

Δ_α : Insieme di rette a due a due perpendicolari è utilizzato come sistema d'assi di coordinate che chiameremo fattori.

L'indipendenza dei fattori è assicurata dall'indipendenza degli assi ortogonali tra loro.

L'inerzia della nuvola lungo l'asse α è il valore λ_α relativo a questo asse ed è definita:

$\lambda_\alpha = \sum\{f_i \ F_\alpha^2 (i) \ / \ i \in I\}$, dove f_i = massa del punto f_j^i , e $F_\alpha^2(i)$ = quadrato della distanza da f_j alla proiezione ortogonale di f_j^i sull'asse α.

La media ponderata del quadrato di una funzione centrata a zero è chiamata varianza.

Per covarianza si intende la media del prodotto di due funzioni centrate (di media nulla) $F_\alpha F_\beta$.

Due fattori distinti hanno covarianza nulla e quindi non sono correlati od anche le funzioni sono indipendenti (condizione necessaria).

Per fare l'analisi su N(I) o N(J) si seguono due processi analoghi e questi due processi conducono agli stessi risultati per due vie simmetriche.

3.2.5 Analisi delle Componenti Principali (ACP).

Consiste nel ricercare i piani principali determinati dagli assi d'inerzia ed il centro del centro di gravità della nuvola per rappresentare le migliori proiezioni.

Permette anche di precisare in quale misura le variabili sono correlate o legate tra loro.

Nell'interpretazione statistica gli assi principali saranno chiamati "fattori principali".

3.2.6 Analisi Fattoriale delle corrispondenze (AFC).

Segue lo stesso principio della ACP ma la distanza utilizzata è quella di X^2 (leggermente diversa da quella utilizzata nella ACP) e che permette di meglio trattare il caso di una matrice d'incidenza (variabili booleane).

La distanza utilizzata è la seguente:

$$d^2(V_i, V_j) = \sum_k \frac{S}{S_k}\left(\left|\frac{S_{ki}}{S_i} - \frac{S_{kj}}{S_j}\right|\right)^2$$

[12]Teoricamente il numero degli assi fattoriali della nuvola N(I) non può superare il più piccolo dei due numeri (Card J - 1) e (Card I - 1).

Dove S_k è la somma della linea k (valori corrispondenti al soggetto k), S_i e S_j le somme delle colonne corrispondenti alle variabili V_i e V_j, S_{ki} e S_{kj} i valori delle variabili V_i e V_j osservate per il soggetto k e S la somma dei valori della tabella di riferimento.

Inoltre i soggetti e le variabili possono essere messi nello stesso spazio. E' possibile allora interpretare le prossimità dei soggetti tra loro, quelle delle variabili tra loro, e quelle dei soggetti con le variabili.

Il "significato" da attribuire ai fattori è tutto a carico del ricercatore, il quale deve interpretare le informazioni che sono più nascoste e che discendono da questo significato. Ci si interesserà ai contributi di certi punti a questi fattori e alle posizioni relative dei sottogruppi di popolazione studiata. Risulta molto interessante l'introdurre delle variabili supplementari e/o dei soggetti supplementari.

4.0 Comparazione sperimentale tra analisi implicativa ed analisi fattoriale

La comparazione tra i due strumenti statistici ha una pregiudiziale epistemologica:

1. L'analisi fattoriale è nel dominio della statistica descrittiva: medie, distanze anche con metodi geometrici. La misura tra variabili utilizzata è simmetrica.
2. L'analisi implicativa è nel dominio della statistica inferenziale. Confronto con un campione ideale (vedi paragrafo 3.1). La misura tra variabili utilizzata è asimmetrica.

Nonostante queste due grandi differenze epistemologiche i due strumenti statistici vengono utilizzati per la ricerca in didattica delle Matematiche in quanto strumenti adatti per l'analisi multivariata di piccoli campioni.

E' chiaro che l'Analisi Fattoriale delle Corrispondenze (AFC) può portare a studio di variabili di variabili (fattori) che possono essere utili per analizzare sperimentalmente concezioni, relazioni tra concezioni riferite all'individuazione dei fattori. Altra cosa è invece quando vengono introdotte variabili supplementari. Nel caso della ricerca in didattica le variabili supplementari sono degli individui costruiti in funzione delle ipotesi e dell'analisi a priori. Sulla matrice trasposta le variabili diventano gli allievi con l'aggiunta delle variabili supplementari (allievi teorici che corrispondo a caratteristiche ben definite dalla natura del problema di ricerca). La conseguente AFC da delle informazioni riguardo gli eventuali fattori individuati danno informazioni su quanti individui si raggruppano su di una variabile supplementare e questo ci porta ad inferire ragionamenti di tipo implicativi: *Se n individui si raggruppano sul quella variabile supplementare allora la variabile supplementare individua una concezione significativa*.

L'Analisi Statistica Implicativa (ASI) si occupa proprio di implicazioni e questa attività viene fornita senza l'introduzione di variabili supplementari. Ma quando si introducono le variabili supplementari l'analisi implicativi risulta essere molto più chiara ed evidente.

La comparazione dei due metodi può avere un senso solo nel caso dell'introduzione delle variabili supplementari per quanto detto finora.

Vengono comparati due lavori sperimentali analizzati separatamente su di uno stesso campione e con le stesse variabili supplementari.

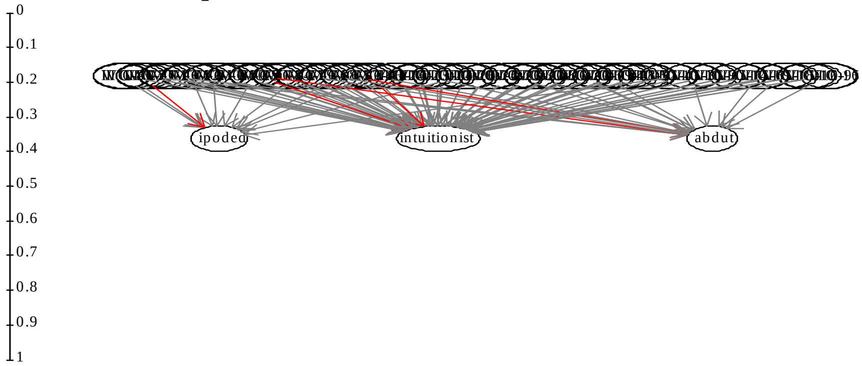
4.1 Le variabili supplementari e la congettura di Goldbach.

In questo lavoro di tesi di dottorato di Aldo Scimone, attraverso l'introduzione di variabili supplementari riguardanti alcuni schemi di ragionamento sulla risoluzione della congettura di Goldbach, sono stati comparati gli stessi dati con CHIC e con SPSS per l'analisi fattoriale.

I grafi che vengono fuori dall'esperienza sono i seguenti.

Troisièmes Rencontres Internationales – Terzo Convegno Internazionale - Third International Conference
A.S.I. Analyse Statistique Implicative – Analisi Statistica Implicativa – Implicative Statistic Analysis
Palermo (Italy) 6-8 Octobre 2005

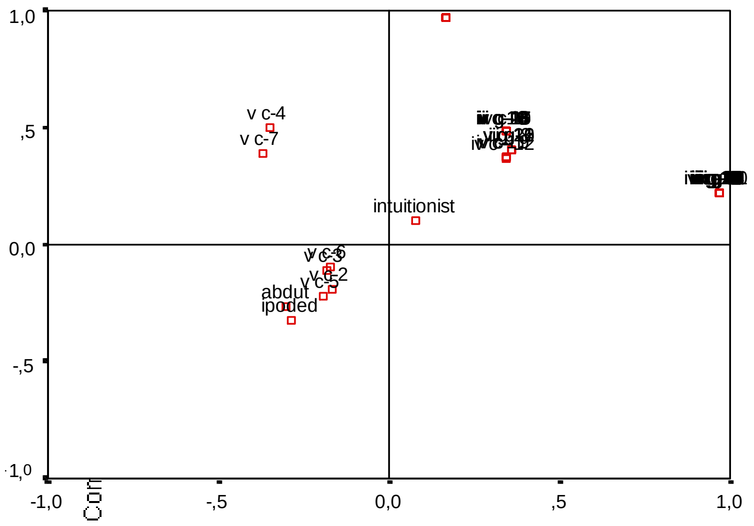
Grafo implicativo.



Graphe implicatif : C:\WINDOWS\Desktop\tesiFP-Dott\aldo\Goldtrasp2.csv 99 95 90 85

Ipotetic, Intuizionist e abdut, rappresentano gli schemi di ragionamento utilizzati dagli studenti. In questa rappresentazione tutto è molto chiaro. Gli allievi si suddividono nelle tre variabili supplementari.

Grafico componenti ruotato



Componente 1

In questa rappresentazione sembrerebbe che Intuizionist discrimini le altre due variabili…ipotetic e abdut che risultano assieme e che invece nella precedente rappresentazione non lo erano!!

4.2 Il caso del passaggio dal linguaggio aritmetico a quello algebrico.

Il lavoro di tesi di Elsa Malisani affronta l'aspetto relazionale-funzionale della variabile nel problem-solving, considerando i contesti semiotici dell'algebra e della geometria analitica. L’obiettivo è indagare se la nozione di incognita interferisce con l’interpretazione dell’aspetto funzionale, e se i procedimenti in lingua naturale e/o il linguaggio aritmetico prevalgono come strategie risolutive in mancanza di un’adeguata conoscenza del linguaggio algebrico.

Anche per questo lavoro si cercherà di analizzare le differenze tra l’analisi fattoriale e l’analisi implicativa.

4.2.1 Profilo degli alunni

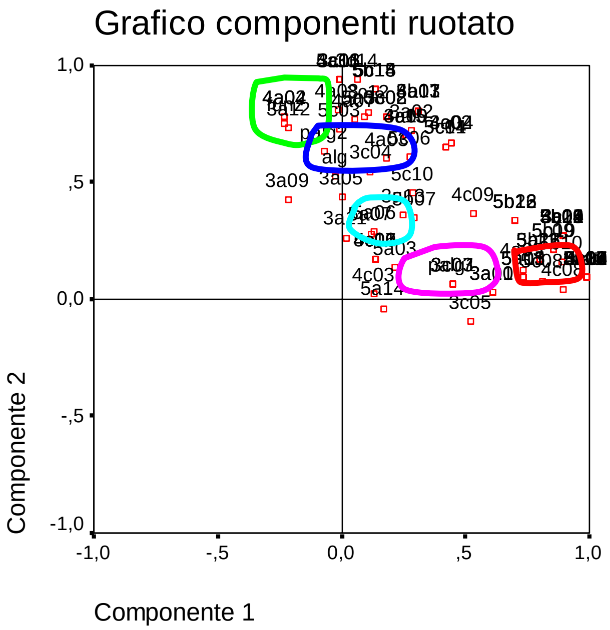
A partire dall’analisi effettuato in precedenza emergono chiaramente i possibili profili degli alunni che affrontano il primo problema. Essi sono:

- **NAT**: questo profilo corrisponde all’alunno che esibisce un procedimento in lingua naturale. Quindi egli aggiunge un dato considerando che le vincite sono uguali (generalmente dividendola a metà) o che le scommesse sono uguali (5) e risolve il problema trovando soltanto una soluzione particolare che verifica l’equazione. Questo profilo è caratterizzato dalla presenza delle seguenti variabili sperimentali: AL1, AL2, AL4, AL15, ALb1 e ALb2.
- **FUNZ**: corrisponde all’alunno che applica una strategia per tentativi ed errori in lingua naturale e/o in linguaggio semi-formalizzato. Egli generalmente assegna più valori ad una delle variabili e trova i valori corrispondenti dell’altra variabile mostrando alcune soluzioni che verificano le equazioni e/o considerando che essa ha una pluralità di soluzioni. Le variabili sperimentali che descrivono questo profilo sono: AL1, AL3, AL11, ALb1, ALb3, ALb4, ALb5 e ALb6).
- **PALG1**: corrisponde all’alunno che utilizza il procedimento pseudo-algebrico. Egli traduce il testo del problema ad un’equazione di primo grado in due incognite ed applica il metodo di “sostituzione in se stessa” (3). Quando arriva all’identità non riesce a darle un’interpretazione adeguata e allora riprende la risoluzione dell’equazione effettuando alcuni errori di tipo sintattico per tentare di trovare la soluzione unica Questo profilo è caratterizzato dalla presenza delle variabili sperimentali: AL1, AL5, AL7, AL13, AL15, ALb1 e ALb2.
- **PALG2**: è una variazione del profilo PALG1, in quanto l’alunno che arriva all’identità cambia procedimento risolutivo abbandonando quello pseudo-algebrico. Le variabili sperimentali che descrivono questo profilo sono le seguenti: AL1, AL3, AL5, AL7, AL9, AL11, AL15, ALb1e ALb3).
- **ALG**: corrisponde all’alunno che applica un procedimento algebrico. Egli traduce il problema ad un’equazione di primo grado in due incognite, considera in modo implicito o esplicito che essa rappresenta una relazione funzionale e quindi che è verificata da una pluralità di soluzioni. Le variabili sperimentali implicate sono: AL1, AL5, AL11, AL15, ALb1, ALb4 e ALb6).

4.2.2 Analisi fattoriale

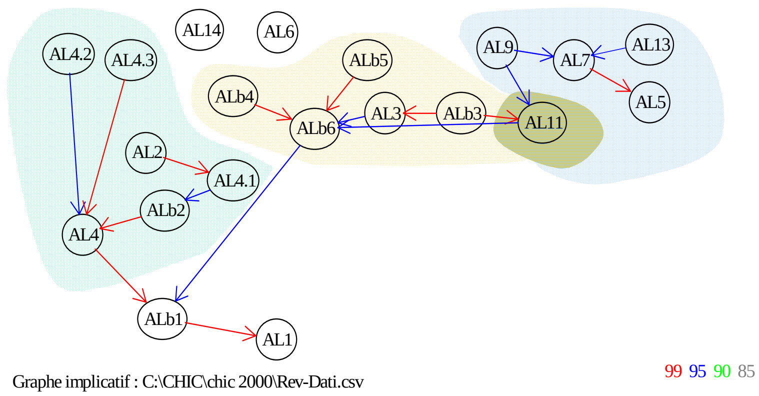
Dall’analisi fattoriale si osserva che, rispetto alla componente orizzontale, le variabili supplementari NAT (cerchiata in rosso) e PALG1 (evidenziata in fucsia) assumono un ruolo determinante e lo caratterizzano fortemente.

I profili ALG (cerchiata in turchese), PALG2 (in blu elettrico) e FUNZ (in verde) formano una nuvola che caratterizza fortemente la componente verticale. La variabile supplementare PALG2 è molto vicina a FUNZ perché l’alunno che abbandona il procedimento pseudoalgebrico, generalmente adotta quello descritto in FUNZ.



Le strategie vincenti sono precisamente quelle descritte nei profili ALG, PALG2 e FUNZ che portano alla pluralità di soluzioni, mentre NAT e PALG1 conducono alla soluzione unica. Questo trova una forte corrispondenza con le diverse concezioni del concetto di "variabile". Quindi l'asse orizzontale rappresenta la concezione di "variabile" come incognita, l'asse verticale, invece, riproduce il suo aspetto relazionale-funzionale. Questi risultati confermano nuovamente la validità della prima ipotesi.

4.2.3 ANALISI IMPLICATIVA



Il grafico implicativo mostra, con percentuali del 95 % e 99 %, tre raggruppamenti ben definiti delle variabili sperimentali che si collegano direttamente o indirettamente con le variabili ALb1 *"l'alunno risponde sulla quantità di soluzioni"* e AL1 *"l'alunno risponde al quesito"*. Ad ogni raggruppamento corrisponde un tipo differente di strategia utilizzata dagli alunni:

- ***Procedimento in lingua naturale*** (nuvola verde): l'alunno aggiunge un dato, considerando che le vincite o le scommesse sono uguali, e trova una soluzione particolare che verifica l'equazione. Questo risultato viene confermato dai legami implicativi tra le variabili sperimentali AL2, ALb2 ed AL4 (con le sue varianti AL4.1, AL4.2 ed AL4.3).

Il procedimento in lingua naturale è il più utilizzato dagli alunni. La concezione di variabile predominante è quella di incognita.

- ***Metodo per tentativi ed errori in lingua naturale o in linguaggio semi-formalizzato*** (nuvola gialla): l'alunno mostra alcune soluzioni che verificano le equazioni e/o considera che essa ha una pluralità di soluzioni, cioè, tiene conto generalmente in maniera implicita che il problema rappresenta una relazione funzionale. Il risultato descritto si ottiene dai legami implicativi tra le variabili AL3, ALb3, AL11 e ALb6.
- ***Strategia pseudo-algebrica*** (nuvola celeste): l'alunno traduce il testo del problema ad un'equazione di primo grado in due incognite ed applica il metodo di "sostituzione in se stessa"(2). Siccome l'alunno non riesce ad interpretare l'identità, cambia procedimento risolutivo abbandonando quello pseudo-algebrico o riprende la risoluzione dell'equazione effettuando alcuni errori per tentare di trovare la soluzione unica. Il risultato descritto si evince dai legami implicativi tra le variabili sperimentali AL9, AL13, AL7 e AL5.

L'alunno che abbandona questa strategia considera, in modo implicito o esplicito, che il problema rappresenta una relazione funzionale. Questo risultato viene confermato dal legame implicativo tra le variabili AL9 ed AL11 che permette il collegamento tra i due procedimenti: per tentativi ed errori ed pseudo-algebrico (nuvola rigata celeste-gialla).

Conclusioni

Sia l'Analisi Fattoriale delle Corrispondenze (AFC) che l'Analisi Statistica Implicativa (ASI) fanno delle analisi su piccoli campioni con la differenza che l'AFC è una statistica descrittiva mentre l'ASI è di tipo inferenziale. Questa prima differenza pone l'ASI in una situazione differente per poter inferire dal campione alla popolazione. Ma come già detto nel paragrafo 4.0 l'introduzione di vraibili supplementari ci può permettere una comparazione sperimentale sui due strumenti statistici.

La misura utilizzata dall'AFC è simmetrica mentre la misura introtta dall'ASI è asimmetrica. L'asimmetria è dovuta all'introduzione della relazione di inclusione e quindi per la sua natura inferenziale (dalla causa all'effetto).

I numerosi dati raccolti in lavori sperimentali riguardanti tesi di laurea e tesi di dottorato portano ad affermare che l'analisi implicativa, quando vengono introdotte variabili supplementari, risulta essere più incisiva nella lettura della contingenza. L'analisi fattoriale invece può dare un utile contributo argomentativi di supporto all'analisi implicativa quando si debbono analizzare soltanto le variabili in gioco. E questo è in perfetto accordo con l'analisi epistemologica dei due strumenti statistici. L'ASI ci consente quindi un miglior approccio all'analisi statistica dell'evoluzione delle concezioni nella dinamica delle classi.

Bibliografia

Agrawal R. et als (1993). *Mining association rules between sets of items in large databases*, Proc. of the ACM SIGMOD'93.

Bodin, A. (1996). *Improving the Diagnostic and Didactic Meaningfulness of Mathematics Assessment in France*, Annual Meeting of the American Educational Research Association AERA - New-York

Brousseau G. , *Theory of didactical situations in mathematics*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1997, Edited and translated by N. Balacheff, M. Cooper, R. Sutherland, and V. Warfield.

Couturier R. (2001). *Traitement de l'analyse statistique implicative dans CHIC*, Actes des Journées sur la « Fouille dans les données par la méthode d'analyse implicative »

Gras R. (1979). *Contribution à l'étude expérimentale et à l'analyse de certaines acquisitions cognitives et de certains objectifs didactiques en mathématiques*, Thèse d'Etat, Université de Rennes 1.

Gras R. (2000). *Les fondements de l'analyse implicative statistique*, Quaderni di Ricerca in Didattica, Palermo, http://dipmat.math.unipa.it/~grim/quaderno9.htm

Gras R., Diday E., Kuntz P. et Couturier R. (2001). *Variables sur intervalles et variables-intervalles en analyse implicative*, Actes du 8ème Congrès de la Société Francophone de Classification de Pointe à Pitre, 17-21 décembre 2001, pp 166-173.

R. Gras, R. Couturier, F. Guillet, F. Spagnolo (2005), Extraction de règles en incertain par la méthode statistique implicative, *Comptes rendus des 12èmes Rencontres de la Société Francophone de Classification, Montréal 30 mai-1er juin 2005, UQAM*, p. 148-151.

Lagrange J.B. (1998). *Analyse implicative d'un ensemble de variables numériques ; application au traitement d'un questionnaire aux réponses modales ordonnées*, Revue de Statistiques Appliquées, XLVI (1), p. 71-93.

Lerman I.C. (1981) Classification et analyse ordinale des données, Dunod, Paris

Lerman I.C., Gras R., Rostam H. (1981). *Elaboration et évaluation d'un indice d'implication pour des données binaires*, I et II, Mathématiques et Sciences Humaines n° 74, p. 5-35 et n° 75, p. 5-47.

Scimone A., How much can the History of Mathematics help mathematics Education? An interplay via Goldbach's conjecture, Zbornik,Bratislavskehoseminara z teorie vyucovania matematiky, Bratislava, 2003, pp. 89-101.

Spagnolo F. (1997). *L'analisi a priori e l'indice di implicazione di Regis Gras*, Quaderni di Ricerca in Didattica, Palermo, http://dipmat.math.unipa.it/~grim/quaderno7.htm

Spagnolo F**.** – R. Gras (2004), *Fuzzy implication through statistic implication: a new approach in Zadeh's framework*, 23rd International Conference of the North American Fuzzy Information Processing Society, NAFIPS (IEEE), Banff, Canada, Edited by Scott Dick-Lukasz Kurgan-Petr Musilek-Witold Pedrycz-Mark Reformat (IEEE Catalog 04TH8736, ISBN 0-7803-8376-1), pagg 425-429, Vol I.

Abstract

Through the analysis of experiences of search in Didactics of the Mathematics they have been comparative data both with the implicative analysis of Gras that with the factorial analysis of the correspondences. The results of this comparison put in evidence that the implicative analysis, when they are introduced supplementay variables, it results to be more incisive in the reading of the contingency. Factorial analysis can give a profit instead argumentative contribution of support to the implicative analysis when they are had to analyze only the variable in game.

Résume

À travers l'examen de situations de recherche en Didactique des Mathématiques, nous comparons les traitements de données obtenus à la fois par l'analyse implicative de Gras et par l'analyse factorielle des correspondances. Les résultats de cette comparaison mettent en évidence que l'analyse implicative, quand on a introduit des variables supplémentaires, est plus efficace dans l'interprétation de la contingence. L'analyse factorielle peut, par contre, offrir une contribution argumentative meilleure lorsqu'on se limite à l'analyse des seules variables en jeu.