ΣΥΝΕΠΑΓΩΓΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

Η έννοια της συνεπαγωγικής στατιστικής προήλθε κατ' αρχήν ως αναγκαία απάντηση σε προβληματισμούς που έθετε η διδακτική των μαθηματικών. Ο Γάλλος καθηγητής Regis Gras του Πανεπιστημίου της Nantes, πρότεινε το 1979 [Gras R.,1979] μια νέα μεθοδολογία που απαντούσε με διαφορετικό σκεπτικό στο εξής ερώτημα:

«Αν μια ερώτηση είναι πιο σύνθετη από μια άλλη, τότε κάθε μαθητής που απαντά σωστά στη συνθετότερη ερώτηση, απαντά ορθά και στην απλούστερη; »

Οι προτάσεις του είδους αυτού είναι γενικά της μορφής:

$$a \Rightarrow b$$

έχοντας χαρακτηριστικό γνώρισμα τη σχέση συνεπαγωγής (=>).

Η κατάσταση αυτή όπως είναι γνωστό κατά κανόνα ισχύει, αλλά δεν παύει να έχει και εξαιρέσεις. Έτσι όσο ισχυρότερη είναι η πιθανότητα να μην εμφανιστεί η εξαίρεση, τόσο ισχυρότερη καθίσταται η συγκεκριμένη πρόταση.

Συνεπαγωγή μεταξύ δυαδικών μεταβλητών

Έστω δυο υποσύνολα Α και Β ενός συνόλου Ε, όπου οι αντίστοιχες μεταβλητές a και b παίρνουν αποκλειστικά τιμές 0 και 1 (0=ψευδές, 1=αληθές).
Στην περίπτωση αυτή τέσσερις είναι οι πιθανές καταστάσεις:

$$a\text{Ç}b \quad a\text{Ç}\bar{b} \quad \bar{a}\text{Ç}b \quad \bar{a}\text{Ç}\bar{b}$$

οι δε πληθάριθμοί τους συμβολίζονται αντίστοιχα με

$$n_{a\text{Ç}b} \quad n_{\bar{a}\text{Ç}b} \quad n_{a\text{Ç}\bar{b}} \quad n_{\bar{a}\text{Ç}\bar{b}}$$

Ορίζουμε ως δείκτη συνεπαγωγής την ποσότητα

$$q(a,\bar{b}) = \frac{n_{a\bar{b}} - \dfrac{n_a \cdot n_{\bar{b}}}{n}}{\sqrt{\dfrac{n_a . n_{\bar{b}}}{n}}} \qquad (10.4)$$

Η ποσότητα $q(a,\bar{b})$ είναι ένας κατάλληλος δείκτης προσδιορισμού της ανυπαρξίας συνεπαγωγής μεταξύ των μεταβλητών a και b.

Η ένταση συνεπαγωγής μεταξύ δύο μεταβλητών

Η ένταση συνεπαγωγής η οποία αποτελεί δείκτη αποδοχής της ύπαρξης συνεπαγωγής μεταξύ των μεταβλητών a και b, υπό την προϋπόθεση ότι ισχύει $n_a \leq n_b$ και $n_b < n$ ορίζεται από τη σχέση

$$\varphi(a,\bar{b}) = 1 - P[Q(a,\bar{b}) \leq q(a,\bar{b})] = \frac{1}{\sqrt{2\pi}} \int_{q(a,\bar{b})}^{\infty} e^{-\frac{t^2}{2}} \cdot dt \qquad (10.6)$$

Παράδειγμα

Έστω ο πίνακας συμπτώσεων απολύτων συχνοτήτων των δυαδικών μεταβλητών a και b.

Πίνακας 10.3

Πίνακας συμπτώσεων απολύτων συχνοτήτων 100 στατιστικών μονάδων

Μεταβλητές	**b=1**	**b=0**	**Περιθώριο**
a=1	5	1	6
a=0	10	84	94
Περιθώριο	**15**	**85**	**100**

Προκύπτουν οι εξής πληθάριθμοι:

$n_a=6$ $n_b=15$ $n_{ab}=5$ $n_{a\bar{b}}=1$ $n_{\bar{b}}=85$

Επομένως

$$q(a,\bar{b}) = \frac{1 - \frac{6 \cdot 85}{100}}{\sqrt{\frac{6 \cdot 85}{100}}}$$

άρα $$q(a,\bar{b}) = \frac{-4{,}1}{2{,}25832} = -1{,}8155$$

οπότε

$$\varphi(a,\bar{b}) = 0{,}9653$$

Μπορούμε να υποστηρίξουμε λοιπόν ότι η συνεπαγωγή (a=>b) είναι απο-δεκτή σε επίπεδο εμπιστοσύνης 96,53%

Έννοια της συνοχής

Για να έχουμε συνεπαγωγή μεταξύ δύο τάξεων A και B απαραίτητη προϋπόθεση είναι να υφίσταται κάποια συνοχή μεταξύ τους, έτσι ώστε η συνεπαγωγική ροή της κλάσης A προς την κλάση B, να τρέφεται από μία εσωτερική ροή της A που να τροφοδοτεί την εσωτερική ροή της B [Gras R.,1991].

Ορισμός: Η συνοχή c(a,b) ενός ζεύγους μεταβλητών υπό την προϋπόθεση $n_a \leq n_b$ ορίζεται ως εξής:

-$c(a,b)=\sqrt{1-E^2}$ όταν $\varphi(a,\bar{b})=p\geq 0{,}5$ όπου $E=p\log_2 p-(1-p)\log_2(1-p)$ (10.21)

-$c(a,b)=1$ όταν $\varphi(a,\bar{b})=p=1$ (10.22)

-$c(a,b)=0$ όταν $\varphi(a,\bar{b})=p\leq 0{,}5$ (10.23)

Εξ ορισμού θέτουμε

$$c(a,a)=1 \qquad (10.24)$$

Συνεπώς διά μέσου του ορισμού της συνοχής ενός ζεύγους μεταβλητών, γίνεται περισσότερο αντιληπτή η έννοια της συνεπαγωγικής αταξίας η οποία προσδιορίζεται από την ένταση της συνεπαγωγής.

Παράδειγμα

Να υπολογισθεί η εντροπία E και η συνοχή του ζεύγους c(a,b) όταν έχουμε $\varphi(a,\bar{b})=0{,}97$.

Αφού δίνεται $\varphi(a,\bar{b})=0{,}97$ τότε p=0,97. 'Αρα η εντροπία E ισούται με

$$E=\frac{-0{,}97.\ \ln(0{,}97)-(1-0{,}97).\ \ln(1-0{,}97)}{\ln 2}=0{,}194$$

οπότε η συνοχή του ζεύγους είναι ίση με $c(a,b)=\sqrt{1-0.1942}=0{,}98$.

Ιεραρχική ταξινόμηση των τάξεων

Η στατιστική συνεπαγωγή μεταξύ δύο μεταβλητών X και Y λαμβάνει υπόψη τρεις τύπους πληροφοριών. Την θετική συσχέτιση, την ομοιογένεια και την θετικά προσανατολισμένη στατιστική εξάρτηση, δηλαδή αποκαλύπτει την ροή της αλληλεπίδρασης μεταξύ των δύο μεταβλητών X και Y, δηλαδή κάτι περισσότερο από το δεσμό που αποκαλύπτεται με τον έλεγχο ανεξαρτησίας του x^2 [Gras R.,1992].

Ο A.Totohasina [Totohasina A.,1994] προτείνει τον ακόλουθο αλγόριθμο που πρέπει να ακολουθεί κάποιος που χρησιμοποιεί την συνεπαγωγική στατιστική.

α) Αρχικά υπολογίζει τις θετικές συσχετίσεις μεταξύ k μεταβλητών (το k αρκετά μεγάλο)

β) Στη συνέχεια διατηρεί τις k' μεταβλητές που παρουσιάζουν τις υψηλότερες συσχετίσεις

γ) Τέλος υποβάλλει τις k' μεταβλητές στη συνεπαγωγική στατιστική η οποία αποκαλύπτει την προσανατολισμένη δυναμική ταξινόμηση των μεταβλητών με βάση δύο κριτήρια απόφασης για τον καθορισμό της σημαντικότητας κάθε τάξης που δημιουργείται:

- τη διαταξική συνεπαγωγή
- τη συνοχή των τάξεων

Ο ακολουθητέος αλγόριθμος συνένωσης δύο τάξεων, δε διαφέρει από εκείνους των μεθόδων ταξινόμησης κατ' αύξουσα ιεραρχία. Το κριτήριο συνένωσης στον οποίο βασίζεται ο αλγόριθμος είναι η συνοχή c(a,b) των τάξεων.

Στο α! βήμα υπολογίζουμε όλες τις εντάσεις συνεπαγωγής $\varphi(a_i,a_j)$ οι οποίες οδηγούν στον υπολογισμό των αντίστοιχων συνοχών $c(a_i,a_j)$.

Αρχικά επιλέγεται η $\max\{c(a_i,a_j)\}$ συνοχή. Σε περίπτωση που υπάρχουν περισσότερα ζεύγη που παρουσιάζουν την μεγαλύτερη τιμή συνοχής, επιλέγεται το ζεύγος του οποίου ο πληθάριθμος είναι μικρότερος και εμφανίζεται στη 1η θέση του ζεύγους.

Αν πάλι υπάρχει πολλαπλή λύση επιλέγεται το ζεύγος με τον μικρότερο πληθάριθμο που βρίσκεται στη 2η θέση του ζεύγους.

Ετσι συνενώνονται οι μεταβλητές a_i,a_j σε μία τάξη προσανατολισμένη από την a_i προς την a_j με την προϋπόθεση ότι ισχύει $n_{ai} \le n_{aj}$, δίχως το n_{ai} να είναι υποχρεωτικά το μικρότερο των πληθαρίθμων.

Εφαρμόζοντας διαδοχικά τον αλγόριθμο αυτό συνενώνονται κάθε φορά δύο νέες μεταβλητές, είτε μια μεταβλητή με μία ήδη δημιουργημένη τάξη, είτε δύο τάξεις μεταξύ τους. Η διαδικασία αυτή συνεχίζεται έως ότου η συνοχή κάθε νέας τάξης που προκύπτει να είναι ίση με το μηδέν ή όταν η τελευταία δημιουργημένη τάξη περιλαμβάνει όλες τις μεταβλητές.

Συμπερασματικά σε μία ιεραρχική ταξινόμηση με κριτήριο "απόστασης" την συνοχή των τάξεων, διαπιστώνουμε ότι σε κάθε βήμα η συνοχή της τάξης που δημιουργείται φθίνει συνεχώς, ενώ ο πληθάριθμός της αυξάνει κατά μία μονάδα.

ΠΑΡΑΔΕΙΓΜΑ

Οι καταστάσεις που αναλύονται με βάση την λογική της συνεπαγωγής ισχύουν κατά κανόνα, όπως αυτές διατυπώνονται, αλλά δεν παύει να υπάρχουν και εξαιρέσεις. Έτσι όσο πιο ισχυρή είναι η πιθανότητα να μην εμφανιστεί η εξαίρεση, τόσο ισχυρότερη καθίσταται η συγκεκριμένη πρόταση.

Στην εφαρμογή που παρουσιάζεται, εξετάζεται η ισχύς της παρακάτω πρότασης.

«Αν οι επιδόσεις ενός μαθητή στις Πανελλήνιες εξετάσεις στα Μαθηματικά, είναι αντίστοιχες με τις επιδόσεις του στα Μαθηματικά Κατεύθυνσης και στα Μαθηματικά Γενικής παιδείας;

Σε ένα δείγμα 146 σπουδαστών του Α.Τ.Ε.Ι.Θ η διαδικασία εφαρμογής της μεθόδου είχε ως εξής:

Κατ' αρχή η ποσοτική κλίματα βαθμολογίας 1-20 μετασχηματίστηκε σε ποιοτική με τρεις διαβαθμίσεις : την υψηλή βαθμολογία (από 20 έως 17), την μεσαία βαθμολογία (από 16,9-10) και την χαμηλή βαθμολογία (κάτω από 10)

Για τις ανάγκες επεξεργασίας των δεδομένων με το λογισμικό MAD έγιναν οι παρακάτω κωδικοποιήσεις:

Η βαθμολογία στις

στο μάθημα της Γενικής παιδείας κωδικοποιήθηκε με MG

στο μάθημα της Κατεύθυνσης κωδικοποιήθηκε με MK

Πανελλήνιες εξετάσεις κωδικοποιήθηκε με MP

ενώ οι τιμές για τις τρεις διαβαθμίσεις της κλίμακας βαθμολόγησης ήταν αντίστοιχα

Για την υψηλή βαθμολογία το 1

Για την μεσαία » το 0,5

Για την χαμηλή βαθμολογία το 0

Τμήμα του πίνακα δεδομένων παρουσιάζεται παρακάτω:

Πίνακας 1: Τμήμα του πίνακα δεδομένων. Βαθμολογίες-Κωδικοποιήσεις

	ΒΑΘΜΟΛΟΓΙΑ			ΚΩΔΙΚΟΙ		
A/A	**MG**	**MK**	**MP**	**MG**	**MK**	**MP**
I1	14,5	6,5	6,5	0,5	0	0
I2	19	18	17	1	1	1
I3	13	9	5	0,5	0	0
I4	18	17	12	1	1	0,5
I5	19	13	19	1	0,5	1
I6	18	13	15	1	0,5	0,5
I7	15,9	8,5	20	0,5	0	1
I8	18	13	7,5	1	0,5	0,5
I9	19	13	18	1	0,5	1
I10	19	17	18	1	1	1
I11	10	9	2	0,5	0	0
I12	11	9	1	0,5	0	0

Η ανάλυση των κωδικοποιημένων βαθμολογιών έδωσε τα παρακάτω αποτελέσματα:

Εντάσεις Συνεπαγωγής μεταξύ των μεταβλητών

	MG	MK	MP
MG	1	0	0
MK	0,8430	1	0
MP	0,8415	0,6428	1

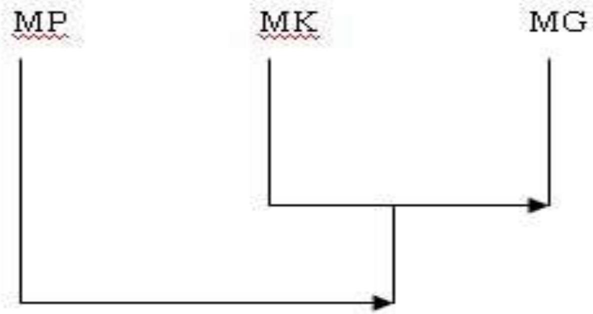
Συνοχές μεταξύ των μεταβλητών

	MG	MK	MP
MG	1	0	0
MK	0,77895	1	0
MP	0,77601	0,34025	1

Περιγραφή της ταξινόμησης

Επίπεδο	Τάξη	Προσανατολισμός	Συνοχή τάξης
1^0	A1	MK->MG	0,7790
2^0	A2	MP->A1	0,5903

Διάγραμμα της ταξινόμησης



Διαπιστώνεται ότι με πιθανότητα 84,3% η βαθμολογία που έλαβε στις Πανελλήνιες εξετάσεις ένας μαθητής στα Μαθηματικά ήταν αντίστοιχες με εκείνες που έλαβε στο Λύκειο στα Μαθηματικά Κατεύθυνσης και στα Μαθηματικά Γενικής Παιδείας Ως επιπλέον συμπέρασμα θα μπορούσε να υποστηριχθεί ότι οι καθηγητές της μέσης εκπαίδευσης αξιολόγησαν σε πολύ μεγάλο βαθμό σωστά τις ικανότητες των μαθητών τους στο μάθημα των μαθηματικών.