

ΚΕΦΑΛΑΙΟ IV

ΜΕΘΟΔΟΛΟΓΙΑ ΤΗΣ ΔΙΕΡΕΥΝΗΣΗΣ ΔΕΔΟΜΕΝΩΝ ΜΕ ΒΑΣΗ ΤΗΝ ΠΑΡΑΓΟΝΤΙΚΗ ΑΝΑΛΥΣΗ ΠΟΛΥΔΙΑΣΤΑΤΩΝ ΔΕΔΟΜΕΝΩΝ

Να σχεδιάζεις χωρίς βιασύνη αλλά να εκτελείς γρήγορα
Ισοκράτης

4.0 Γενικά

Στόχος μιας παραγοντικής ανάλυσης χρησιμοποιώντας μία από τις μεθόδους της, είναι η **οπτικοποίηση** των αλληλεξαρτήσεων και των αλληλεπιδράσεων μεταξύ των «αντικειμένων» (γραμμών) και των μεταβλητών (στηλών) του πίνακα δεδομένων.

Πρέπει να τονιστεί ιδιαίτερα ότι η δημιουργία ενός παραγοντικού άξονα **ποσοτικοποιεί** την συμπεριφορά των «αντικειμένων» και των μεταβλητών, με αποτέλεσμα να οδηγεί στην διάταξή τους πάνω στον άξονα, οπότε ανάλογα με την τιμή των συντεταγμένων τους να προσδιορίζουν μία «κατεύθυνση», η οποία να υποδηλώνει τις περισσότερες φορές την **υποβόσκουσα τάση** των δεδομένων του πίνακα.

4.1 Ανάλυση ενός πίνακα δεδομένων

Όπως είναι γνωστό σε κάθε έρευνα αγοράς οι περισσότερες ερωτήσεις αντιστοιχούν σε ποιοτικές μεταβλητές και λιγότερο σε ποσοτικές μεταβλητές (ασυνεχείς ή συνεχείς) οι οποίες τελικώς μετασχηματίζονται σε ποιοτικές δημιουργώντας ένα συγκεκριμένο πλήθος διαβαθμίσεων.

Ο ερευνητής για να αναλύσει τις απαντήσεις των ερωτώμενων, έχει στη διάθεσή του δύο διαδικασίες στατιστικής ανάλυσης.

A) Η πρώτη αναφέρεται σε μεθόδους της κλασικής στατιστικής ανάλυσης.

Χρησιμοποιώντας, λοιπόν, τις μεθόδους της κλασικής στατιστικής ανάλυσης έχει την δυνατότητα δημιουργίας της κατανομής συχνοτήτων των μεταβλητών, τον υπολογισμό των παραμέτρων των κατανομών, όπως μέσο όρο, διακύμανση και τυπική απόκλιση, εύρεση ποσοστών, τη συσχέτιση μεταξύ δύο μεταβλητών, τη διαπίστωση αν υφίσταται ανεξαρτησία ή εξάρτηση μεταξύ δύο μεταβλητών από το πλήθος των k μεταβλητών, είτε να εφαρμόσει κάποιους ελέγχους παραμετρικούς ή μη παραμετρικούς χρησιμοποιώντας μία ή δύο το πολύ μεταβλητές.

Π.χ Έστω ότι έχουμε ένα ερωτηματολόγιο με 10 ερωτήματα το οποίο απάντησαν 557 ερωτώμενοι και θέλουμε να αναλύσουμε δύο από αυτά τα ερωτήματα. Το ένα ερώτημα αντιστοιχεί στην Ηλικία του ερωτώμενου και το άλλο στο ερώτημα για το «Πώς βλέπει ο ερωτώμενος την καθαριότητα της πόλης της Θεσσαλονίκης;»

Ο ερευνητής καθόρισε για τις ηλικίες τις ακόλουθες τέσσερις διαβαθμίσεις με τους αντίστοιχους κωδικούς **Ηλικία κάτω των 18 ετών** (κωδ:Η1=1), **Ηλικία 19-35 ετών** (κωδ:Η2=2), **Ηλικία 36-45 ετών** (κωδ:Η3 =3) και **Ηλικία πάνω από 46 ετών** (κωδ :Η4=4), ενώ για την στάση των ερωτηθέντων για την καθαριότητα πρότεινε τρεις διαβαθμίσεις με τους αντίστοιχους κωδικούς. Την διαβάθμιση **Αρνητική στάση** (κωδ: Α=1), την διαβάθμιση **Ουδέτερη στάση** (κωδ:Ο=2) και την διαβάθμιση **Θετική στάση** (κωδ: Θ=3). Τα δεδομένα όπως προέκυψαν παρουσιάζονται στον παρακάτω πίνακα ως εξής:

Πίνακας 4.1:Τμήμα του κωδικοποιημένου πίνακα δεδομένων

Ind	Καθαριότητα	Ηλικία
I1	3	2
I2	2	4
I3	1	1
I4	2	2
I5	2	3
I6	1	2
I7	1	1
I8	2	1
I9	2	2
I10	3	1

Επειδή τα δεδομένα είναι κωδικοποιημένα, δεν έχει κανένα νόημα να εξαχθούν οι γνωστές παράμετροι της κλασικής στατιστικής, όπως λ.χ μέσος όρος, διακύμανση κ.λ.π.

Βέβαια ο ερευνητής μπορεί αρχικά να υπολογίσει, αν υφίσταται συσχέτιση μεταξύ των μεταβλητών **Ηλικία** ερωτώμενων και **Στάση** ως προς την καθαριότητα υπολογίζοντας τον συντελεστή συσχέτισης ρ , η τιμή του οποίου βρέθηκε ίση με $\rho = -0,02625$.

Με δεδομένο τον σχετικά μικρό αριθμό δείγματος ($n=557$) ο έλεγχος συσχέτισης $H_0: \rho=0$ γίνεται δεκτός σε επίπεδο σημαντικότητας $\alpha=5\%$.

Επομένως από την συγκεκριμένη τιμή του δείγματος συμπεραίνεται ότι οι δύο μεταβλητές είναι **ασυσχέτιστες**.

Ακολουθως ο ερευνητής δημιουργεί ένα απλό πίνακα διπλής εισόδου, όπου διασταυρώνονται οι τρεις διαβαθμίσεις της Στάσης (Α,Ο,Θ) ως προς την καθαριότητα με τις τέσσερις διαβαθμίσεις της Ηλικίας (Η1,Η2,Η3,Η4), ώστε να διαπιστωθεί αν οι δύο μεταβλητές είναι ανεξάρτητες. Αυτό γίνεται επειδή δύο μεταβλητές εφόσον είναι ασυσχέτιστες, δεν σημαίνει ότι είναι και ανεξάρτητες (Καραπιστόλης Δ, 2011 Στατιστική Επιχειρήσεων). Δημιουργεί λοιπόν τον παρακάτω πίνακα στον οποίο εφαρμόζεται ο έλεγχος ανεξαρτησίας.

Πίνακας 4.2: Πίνακας διπλής εισόδου

Διαβαθμίσεις	H1	H2	H3	H4	Άθροισμα γραμμής
A	9	55	27	17	108
O	3	83	29	58	173
Θ	11	136	69	60	276

Ο έλεγχος ανεξαρτησίας σε επίπεδο σημαντικότητας 5% δίνει τα ακόλουθα αποτελέσματα.

Για $n=6$ βαθμούς ελευθερίας $\Rightarrow X^2=20,85$

Σε επίπεδο σημαντικότητας 5% έχουμε $X_{6,0.05}^2 = 12,59$

Επειδή $X_{6,0.05}^2 < X^2$ συμπεραίνεται ότι η **Στάση ως προς την καθαριότητα** με την **Ηλικία** είναι **εξαρτημένες**.

Αυτό σημαίνει ότι οι διαφορετικές ηλικίες έχουν συγκεκριμένη στάση ως προς την καθαριότητα. Βέβαια αυτό **το συμπέρασμα δεν μας πληροφορεί ποια ηλικία συνδέεται με ποια στάση, βασικό ερώτημα προς διερεύνηση.**

B) Η δεύτερη διαδικασία στατιστικής ανάλυσης είναι αυτή που αναφέρεται σε μεθόδους της πολυπαραγοντικής στατιστικής ανάλυσης και ιδιαίτερα στις μεθόδους της Παραγοντικής Ανάλυσης των Αντιστοιχιών και της Ανιούσας Ιεραρχικής Ταξινόμησης, οι οποίες όχι μόνο μας προσδιορίζουν αν συσχετίζονται ή όχι οι δύο ερωτήσεις, αλλά και ποια ηλικία έχει συγκεκριμένη στάση,

Οι μέθοδοι αυτές οπωσδήποτε προσφέρουν πληρέστερη πληροφόρηση από τις κλασικές μεθόδους και είναι ιδιαίτερα χρήσιμες όταν μελετά κάποιος ποιοτικές μεταβλητές.

4.2 Εφαρμογή της Παραγοντικής Ανάλυσης των Αντιστοιχιών

Εφόσον αναλύσουμε ένα πίνακα διπλής εισόδου με την μέθοδο της Παραγοντικής Ανάλυσης των Αντιστοιχιών, έχουμε στη διάθεσή μας, μετά την δημιουργία του παραγοντικού επιπέδου, τις ακόλουθες πληροφορίες.

1. το ποσοστό της αντλούμενης πληροφορίας που παρέχει το παραγοντικό επίπεδο
2. τις αλληλεπιδράσεις μεταξύ μεταβλητών που παρίστανται από τις στήλες
3. τις αλληλεπιδράσεις μεταξύ ιδιοτήτων που παρίστανται από τις γραμμές
4. ποια μεταβλητή στήλη συνδέεται με ποια ιδιότητα γραμμή

Η ανάλυση του πίνακα 4.2 δίνει τις παρακάτω πληροφορίες:



εικόνα 4.1

Από την εικόνα 4.1 προκύπτει ότι οι δύο μεταβλητές είναι **εξαρτημένες** σε επίπεδο σημαντικότητας 0,05.

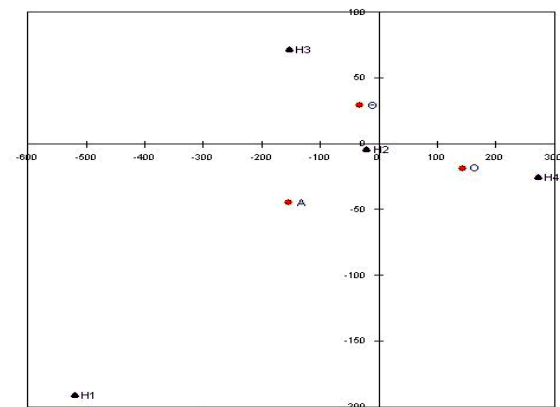
Στη συνέχεια ακολουθεί η εικόνα 4.2. Με δεδομένο ότι η τάξη του πίνακα 4.2 είναι ίση με 3, η ανάλυση με την AFC παρέχει δύο παραγοντικούς άξονες, καθόσον η τάξη ενός πίνακα διπλής εισόδου όταν είναι k, η παραγοντική ανάλυση προσδιορίζει k-1 παραγοντικούς άξονες.

ΣΥΝΟΛΙΚΗ ΑΔΡΑΝΕΙΑ 0,03743				
ΑΞΩΝ	ΑΔΡΑΝΕΙΑ	%ΕΡΜΗΝΕΙΑΣ	ΑΘΡΟΙΣΜΑ	ΙΣΤΟΓΡΑΜΜΑ ΧΑΡΑΚΤΗΡ.ΡΙΖΩΝ
1	0,0345987	92,43	92,43	*****
2	0,0028343	7,57	100,00	**

εικόνα 4.2:Ιστόγραμμα χαρακτηριστικών ριζών

Η εικόνα 4.2 πληροφορεί ότι ο 1^{ος} παραγοντικός άξονας ερμηνεύει το 92,43% της συνολικής πληροφορίας που αντλείται από τα δεδομένα του πίνακα 5, ενώ ο 2^{ος} παραγοντικός άξονας το υπόλοιπο 7,57%.

Ακολουθεί η εικόνα 4.3



εικόνα 4.3

Η εικόνα 4.3 παρουσιάζει το παραγοντικό επίπεδο 1x2, από το οποίο μπορεί να διαπιστωθεί η αλληλεπίδραση μεταξύ μεταβλητών και ιδιοτήτων.

Για την ερμηνεία των πληροφοριών που μπορεί να εξαχθούν από το παραγοντικό επίπεδο, γίνεται με τη βοήθεια των ενδείξεων του πίνακα 4.3, ο οποίος παρουσιάζει τις τιμές d του δείκτη έλξης- άπωσης μεταξύ «γραμμών» και «στηλών» του πίνακα 4.2, προσδιορίζοντας την μεταξύ τους εξάρτηση.

Πίνακας 4.3: Δείκτης έλξης- άπωσης

ΕΝΑ	H1	H2	H3	H4
A	2,0181	1,0352	1,114	0,6494
O	0,4199	0,9752	0,7469	1,3832
Θ	0,9651	1,0016	1,114	0,8969

Όταν η τιμή του δείκτη d έλξης-άπωσης είναι μεγαλύτερη του 1, σημαίνει ότι υφίσταται έλξη μεταξύ των διαβαθμίσεων των γραμμών και των στηλών, ενώ όταν η τιμή είναι μικρότερη του 1, υποδεικνύει άπωση. Στην περίπτωση που η τιμή είναι ίση με 1 δεν υφίσταται καμία εξάρτηση.

Συνεπώς από την απεικόνιση των σημείων πάνω στο παραγοντικό επίπεδο και την επικύρωση με βάση τις τιμές του δείκτη d προκύπτουν οι εξής παρατηρήσεις: οι ηλικίες **κάτω των 18** ετών (H1) και οι ηλικίες **19 έως 35** ετών (H2) συνδέονται με την **Αρνητική στάση** (A), οι ηλικίες **35 έως 45** ετών (H3) συνδέονται με την **Θετική στάση** (Θ), ενώ τέλος **οι ηλικίες άνω των 45** ετών συνδέονται με την **Ουδέτερη στάση** (O).

Η ανάγλυφη αυτή διαπίστωση προκύπτει από την εξής λογική που διέπει την Παραγοντική Ανάλυση των Αντιστοιχιών.

Ως γνωστόν όταν από μια σειρά αριθμών υπολογιστεί ο μέσος όρος, οι αποκλίσεις κάθε τιμής από αυτή την παράμετρο θέσης, προσδιορίζουν κάτι συγκεκριμένο. Έτσι όσο η θετική απόκλιση είναι μεγαλύτερη τόσο πιο χαρακτηριστική είναι η ιδιότητα που μετρά η μεταβλητή που αντιστοιχεί στον μέσο όρο. Ανάλογη ερμηνεία δίνεται για την αρνητική απόκλιση. Δηλαδή όσο μεγαλύτερη είναι η απόκλιση τόσο λιγότερο χαρακτηρίζει η μεταβλητή την ιδιότητα που μετρά.

Συνεπώς αν από ένα δείγμα 100 ατόμων προέκυψε ότι το μέσο ετήσιο εισόδημα είναι 12.000 ευρώ, όταν κάποιος από αυτούς έχει εισόδημα 35.000 ευρώ θεωρείται εύπορος, ενώ αν παρουσιάζει εισόδημα 55.000 ευρώ θεωρείται πλούσιος. Στην αντίθετη περίπτωση ένας άλλος που δηλώνει 5.000 ευρώ θεωρείται φτωχός.

Ανάλογη είναι η διαδικασία ερμηνείας στην Παραγοντική Ανάλυση των Αντιστοιχιών.

Αντί του μέσου όρου, επειδή διασταυρώνονται δύο τουλάχιστον μεταβλητές, υπολογίζεται ο πίνακας ανεξαρτησίας, δηλαδή το πως θα παρουσιαζόταν ο πίνακας διπλής εισόδου, αν η αλληλεπίδραση των δύο μεταβλητών ήταν ιδανική.

Η ιδανική αυτή σχέση υπολογίζεται βάσει της σχέσης (4.1)