



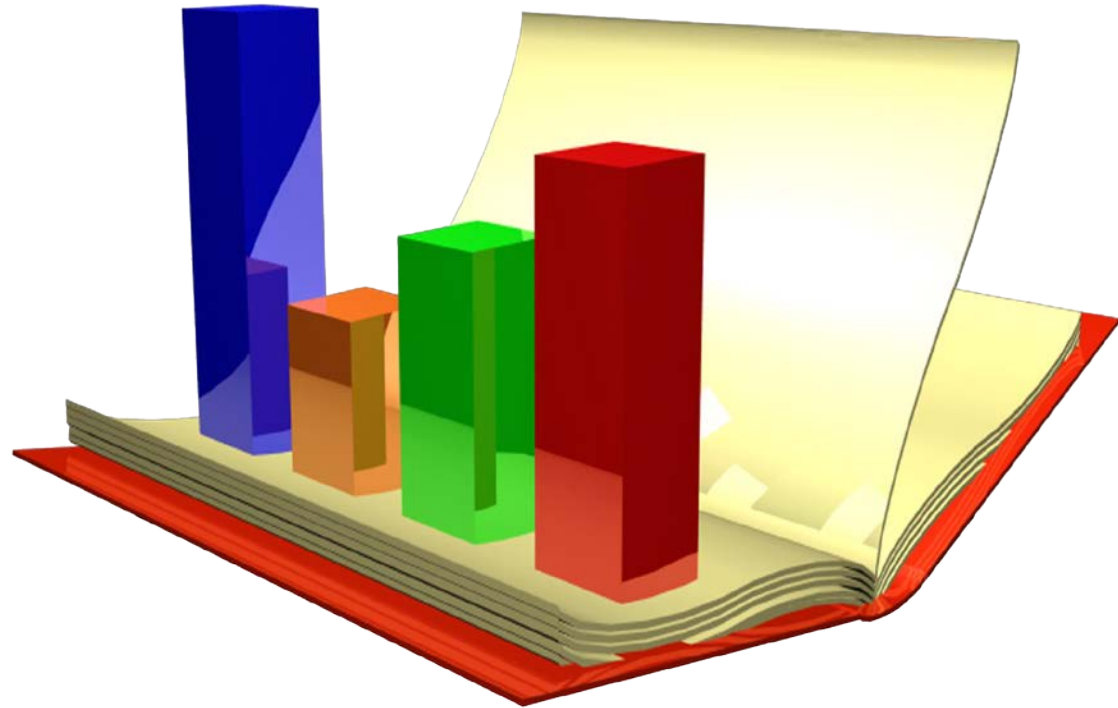
Εθνικό Μετσόβιο Πολυτεχνείο

Τομέας Βιομηχανικής Διοίκησης και Επιχειρησιακής Έρευνας

Έρευνα Αγοράς Μέρος 2^ο

-

**Έλεγχοι Συσχέτισης και
Πολυμεταβλητή Στατιστική**



Περιεχόμενα

1. Έλεγχοι Συσχετίσεων Δύο Μεταβλητών
2. Ανάλυση Παλινδρόμησης
3. Ανάλυση Παραγόντων
4. Ανάλυση Συστάδων
5. Ανάλυση Αντιστοιχιών

Έλεγχοι συσχετίσεων 2 μεταβλητών: Βασικές πληροφορίες

- Η κατηγορία των ελέγχων υποθέσεων καλύπτει ένα σύνολο στατιστικών αναλύσεων οι οποίες έχουν μια κοινή βάση αξιολόγησης: **τη διερεύνηση της συσχέτισης μεταξύ δύο μεταβλητών.**
- Στο σύνολο των μεθόδων και ελέγχων της κατηγορίας, ο στόχος έγκειται στην **στατιστική επιβεβαίωση ή απόρριψη υποθέσεων** οι οποίες είναι εκ των προτέρων προσδιορισμένες.
- Ως προς τη δομή τους, χρησιμοποιούνται δύο **εναλλακτικές μεταξύ τους υποθέσεις** και πιο συγκεκριμένα η μηδενική υπόθεση (H_0) και η εναλλακτική προς αυτή H_1 . Η πρακτική αποτύπωση των υποθέσεων είναι:
 H_0 : Οι μεταβλητές είναι ανεξάρτητες.
 H_1 : Οι μεταβλητές δεν είναι ανεξάρτητες.
- Η στατιστική ανεξαρτησία/ συσχέτιση μεταξύ των δύο εξεταζόμενων μεταβλητών προκύπτει από τη σύγκριση του παραγόμενου **p-value** με τις τιμές 0,05 (**βαθμός εμπιστοσύνης 95%**) ή 0,01 (**βαθμός εμπιστοσύνης 99%**) αντίστοιχα.

Έλεγχοι συσχετίσεων 2 μεταβλητών : παράμετροι επιλογής μεθόδου

Θέλω να...	Χρησιμοποιώ
Ελέγξω τη συσχέτιση δύο κατηγορικών μεταβλητών	Έλεγχος χ^2
Ελέγξω τη συσχέτιση μιας δίτιμης ονομαστικής μεταβλητής (άνδρας/ γυναίκα) με μια ποσοτική μεταβλητή (π.χ. βαθμός διπλώματος)	Έλεγχος T, δύο ανεξάρτητων δειγμάτων*
Ελέγξω τη συσχέτιση μιας πολύτιμης ονομαστικής μεταβλητής (ειδικότητα ΕΜΠ) με μια ποσοτική μεταβλητή (π.χ. βαθμός διπλώματος)	Ανάλυση διακύμανσης κατά ένα παράγοντα (ANOVA) * συμπληρωμένος με ελέγχους post hoc
Ελέγξω τη γραμμική συσχέτιση δύο ποσοτικών (ή διατακτικών) μεταβλητών	Έλεγχος γραμμικής συσχέτισης κατά Pearson*

Έλεγχοι ισότητας μέσω τιμών

**Οι έλεγχοι που παρουσιάζονται αφορούν στις παραμετρικές εκδοχές των μεθόδων οι οποίες εφαρμόζονται όταν τηρούνται οι προϋποθέσεις του γραμμικού μοντέλου (κανονική κατανομή). Σε περίπτωση μη τήρησης των προϋποθέσεων, ή όταν οι μεταβλητές δεν είναι ποσοτικές αλλά διατακτικές χρησιμοποιούνται οι εναλλακτικοί, μη παραμετρικοί έλεγχοι*

Ο έλεγχος χ^2

Στον έλεγχο χ^2 ελέγχω τη συσχέτιση μεταξύ δύο κατηγορικών μεταβλητών. Σε αυτήν την περίπτωση ελέγχου, οι μεταβλητές είναι συνήθως ονομαστικές. Αν και συνήθως η μία μεταβλητή λογίζεται ως ανεξάρτητη και η άλλη ως εξαρτημένη, η συγκεκριμένη διάκριση δεν είναι απαραίτητη.

Παράδειγμα: Τόπος διαμονής διπλωματούχων ΕΜΠ – Σχέση εργασίας διπλωματούχων ΕΜΠ

Επιλέγω επιθυμητό επίπεδο σημαντικότητας (α).
Όπου, $\alpha = 1 - \beta.ε.$ (για $\beta.ε = 95\%$, $\alpha=0,05$)



Εκτελώ τον έλεγχο. Η παραγόμενη τιμή sig συγκρίνεται με το επιθυμητό επίπεδο σημαντικότητας (α). Εφόσον η τιμή sig είναι μικρότερη του α , απορρίπτεται η μηδενική υπόθεση και επομένως οι μεταβλητές συσχετίζονται



Με δεδομένη τη συσχέτιση, χρησιμοποιείται ο πίνακας συχνοτήτων προκειμένου να διαπιστωθεί ο λόγος της συσχέτισης



Οι κάτοικοι επαρχίας είναι πιο πιθανό να είναι ελεύθεροι επαγγελματίες ή αντίστροφα, οι κάτοικοι Αττικής είναι περισσότερο πιθανό να εργάζονται ως μισθωτοί

H_0 : Ο τόπος διαμονής δεν σχετίζεται με τη σχέση εργασίας
 H_1 : Ο τόπος διαμονής σχετίζεται με τη σχέση εργασίας

Αποτελέσματα ελέγχου χ^2

Chi-Square Tests					
	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	53,374 ^a	1	,000		

Πίνακας Συχνοτήτων

		Μισθωτοί	Ελεύθεροι επαγγελματίες	Σύνολο
Τόπος διαμονής	Αττική	1223	221	1444
	Εκτός Αττικής	229	110	339
Σύνολο		1452	331	1783

Ο έλεγχος T

Στον έλεγχο T συγκρίνω τη μέση τιμή μιας ποσοτικής μεταβλητής, μεταξύ δύο ομάδων παρατηρήσεων. Πρακτικά, προσπαθώ να εντοπίσω το κατά πόσο η μέση τιμή της μίας ομάδας είναι σημαντικά διαφορετική (μεγαλύτερη ή μικρότερη) από τη μέση τιμή της δεύτερης ομάδας. Η κατηγορική μεταβλητή λογίζεται ως ανεξάρτητη και η ποσοτική ως εξαρτημένη

Παράδειγμα: Σύγκριση βαθμού διπλώματος ΕΜΠ μεταξύ ανδρών και γυναικών αποφοίτων

Επιλέγω επιθυμητό επίπεδο σημαντικότητας (α). Όπου, $\alpha = 1 - \beta$.ε. (για $\beta = 95\%$, $\alpha = 0,05$)



Εκτελώ τον έλεγχο. Η παραγόμενη τιμή sig συγκρίνεται με το επιθυμητό επίπεδο σημαντικότητας (α). Εφόσον η τιμή sig είναι μικρότερη του α , απορρίπτεται η μηδενική υπόθεση και επομένως οι μεταβλητές συσχετίζονται



Με δεδομένη τη συσχέτιση, χρησιμοποιείται ο πίνακας περιγραφικών μέτρων προκειμένου να διαπιστωθεί η ανισότητα της μέσης τιμής



Οι γυναίκες αποφοιτούν με μεγαλύτερη βαθμολογία από τους άνδρες

H_0 : Άνδρες και γυναίκες αποφοιτούν με τον ίδιο βαθμό διπλώματος
 H_1 : Άνδρες και γυναίκες αποφοιτούν με διαφορετικό βαθμό διπλώματος

Αποτελέσματα ελέγχου T

		Levene's Test for Equality of Variances				
		F	Sig.	t	df	Sig. (2-tailed)
Βαθμός διπλώματος	Equal variances assumed	,793	,373	-7,180	1767	,000
	Equal variances not assumed			-7,151	1469,001	,000

Περιγραφικά μέτρα μέσης τιμής συγκρινόμενων ομάδων

	Φύλο	N	Mean	Std. Deviation	Std. Error Mean
Βαθμός διπλώματος	Άνδρας	1071	7,28	,64032	,01957
	Γυναίκα	698	7,51	,65276	,02471

Ο έλεγχος μονοπαραγοντικής διακύμανσης (one-way ANOVA)

Η ANOVA αποτελεί προέκταση του T-Test και χρησιμοποιείται στην περίπτωση που συγκρίνω τη μέση τιμή μιας ποσοτικής μεταβλητής μεταξύ τριών ή και περισσότερων ομάδων (τιμές κατηγορικής μεταβλητής)

Παράδειγμα: Σύγκριση βαθμού διπλώματος ΕΜΠ μεταξύ των διαφορετικών ειδικοτήτων

Αποτελέσματα ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	166,342	8	20,793	61,940	,000
Within Groups	590,823	1760	,336		
Total	757,165	1768			

Εκτελώ τον έλεγχο, κατά τρόπο αντίστοιχο με το T-Test και ελέγχω την ύπαρξη στατιστικής συσχέτισης μεταξύ των δύο μεταβλητών (ειδικότητα και βαθμός διπλώματος). Στην περίπτωση που εντοπίσω στατιστική συσχέτιση, δεν είμαι σε θέση να αναγνωρίσω την αιτία της



Με δεδομένη τη συσχέτιση, εκτελώ post hoc έλεγχο προκειμένου να εντοπίσω τις ομάδες με σημαντική διαφορά στη μέση τιμή της ποσοτικής μεταβλητής (βαθμολογία).



Οι ομάδες ταξινομούνται σε κατηγορίες, οι οποίες δεν εμφανίζουν σημαντικές διαφορές εσωτερικά και παράλληλα είναι σημαντικά διαφορετικές μεταξύ τους

Αποτελέσματα post hoc

Tukey HSD ^{a,b}						
Ειδικότητα	N	Subset for alpha = 0.05				
		1	2	3	4	5
Μηχανολόγοι	228	7,0807				
Πολιτικοί	277	7,0971	7,0971			
Τοπογράφοι	170	7,1035	7,1035			
Ναυπηγοί	87	7,1516	7,1516			
ΣΕΜΦΕ	159		7,2757	7,2757		
Μεταλλειολόγοι	120			7,3948	7,3948	
Χημικοί	221				7,4916	
Ηλεκτρολόγοι	273				7,5543	
Αρχιτέκτονες	234					8,0119

Έλεγχος γραμμικής συσχέτισης κατά Pearson

Στον έλεγχο γραμμικής συσχέτισης στοχεύω στη διερεύνηση της συγγραμμικής (ομόρροπης ή αντίρροπης) μεταβολής δύο μεταβλητών.

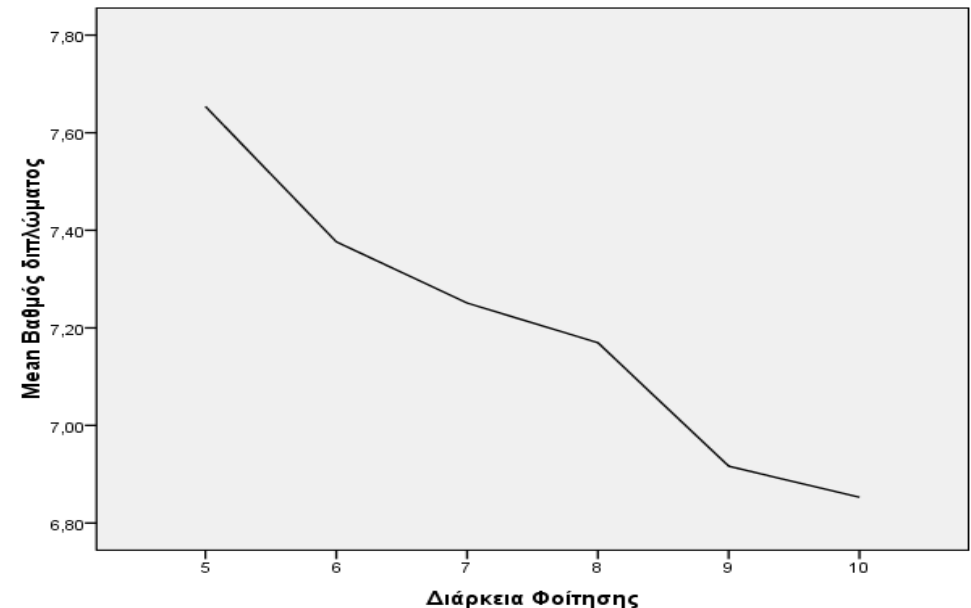
Στις παραγόμενες τιμές της μεθόδου, εκτός από τη στατιστική συσχέτιση (sig) με ενδιαφέρει και η τιμή Pearson, η οποία λαμβάνει τιμές από -1 έως 1. Η τιμή -1 αντιστοιχεί σε απόλυτη αντίρροπη συμμεταβολή και η τιμή 1 σε απόλυτη ομόρροπη συμμεταβολή

Παράδειγμα: Διάρκεια Φοίτησης στο ΕΜΠ – Βαθμός διπλώματος

		Βαθμός διπλώματος	Διάρκεια Φοίτησης
Βαθμός διπλώματος	Pearson Correlation	1	-,308**
	Sig. (2-tailed)		,000
	N	1622	1622
Διάρκεια Φοίτησης	Pearson Correlation	-,308**	1
	Sig. (2-tailed)	,000	
	N	1622	1622

****.** Correlation is significant at the 0.01 level (2-tailed).

→ Αντίρροπη συσχέτιση (μακρά φοίτηση => μικρός βαθμός)
→ Σημαντική συσχέτιση σε β.ε. 99%



Μέθοδοι πολυμεταβλητής στατιστικής

Εισαγωγή στη πολυμεταβλητή στατιστική (1/2)

- ✓ Η ανάλυση δεδομένων μέσω της ταυτόχρονης αξιοποίησης πολλών μεταβλητών, αναφέρεται συχνά ως πολυμεταβλητή (multivariate) στατιστική
- ✓ Αντίθετα με τις προηγούμενες κατηγορίες αναλύσεων, στις οποίες ο στόχος ήταν δεδομένος, οι πολυπαραμετρικές μέθοδοι ανάλυσης εξυπηρετούν εναλλακτικά αντικείμενα μιας έρευνας
- ✓ Χωρίς να αποκλείεται η αξιοποίηση των μεθόδων της κατηγορίας για λόγους ελέγχου υποθέσεων, οι αναλύσεις συνήθως εξυπηρετούν την ανακάλυψη μοτίβων με επαγωγικό τρόπο.

Στη συνέχεια της παρουσίασης, παρουσιάζονται τρεις από τις πλέον διαδεδομένες μεθόδους πολυμεταβλητής στατιστικής

Η ανάλυση
παλινδρόμησης
(regression analysis)

Η ανάλυση
παραγόντων
(factor analysis)

Η ανάλυση
συστάδων
(cluster analysis)

Εισαγωγή στη πολυμεταβλητή στατιστική (2/2)

Η ανάλυση
παλινδρόμησης
(regression analysis)

Βασικός στόχος: Η αποτύπωση του προβλήματος με τη μορφή εξίσωσης

Χωρίς να αποκλείεται η αξιοποίηση των μεθόδων της κατηγορίας για λόγους ελέγχου υποθέσεων, οι αναλύσεις συνήθως εξυπηρετούν την ανακάλυψη μοτίβων με επαγωγικό τρόπο.

Η ανάλυση
παραγόντων
(factor analysis)

Βασικός στόχος: ομαδοποίηση μεταβλητών

Οι αρχικές μεταβλητές ομαδοποιούνται σε κοινές συνιστώσες ή παράγοντες (components ή factors), που λειτουργώντας ως νέες μεταβλητές συνοψίζουν τις μεταξύ αυτών συσχετίσεις και κατά συνέπεια επιτρέπουν την εξαγωγή πληροφοριών που αρχικά ήταν δυσδιάκριτες

Η ανάλυση
συστάδων
(cluster analysis)

Βασικός στόχος: ομαδοποίηση παρατηρήσεων

Η ανάλυση συστάδων (cluster analysis) χρησιμοποιεί την πληροφορία που περιέχεται στις μεταβλητές που εισέρχονται στην ανάλυση, με σκοπό να κατατάξει τις παρατηρήσεις ενός δείγματος σε ομάδες

Περιεχόμενα

1. Έλεγχοι Συσχετίσεων Δύο Μεταβλητών
2. Ανάλυση Παλινδρόμησης
3. Ανάλυση Παραγόντων
4. Ανάλυση Συστάδων
5. Ανάλυση Αντιστοιχιών

Ανάλυση Παλινδρόμησης (1/8)

Η ανάλυση παλινδρόμησης χρησιμοποιείται για την:

- Διατύπωση με τη μορφή εξίσωσης, της συσχέτισης μεταξύ μιας εξαρτημένης μεταβλητής και μίας ή περισσότερων ανεξάρτητων μεταβλητών, και επομένως για την δυνατότητα:
 - Αναγνώρισης ή υπολογισμού της τιμής μιας (εξαρτημένης) μεταβλητής, μέσω των τιμών μίας ή και περισσότερων (ανεξάρτητων) μεταβλητών ή παραγόντων
- Αναγνώριση των πιο δυνατών συσχετίσεων, μεταξύ μίας & πολλών μεταβλητών. Η υπό αυτό το πρίσμα, η ανάλυση παλινδρόμησης μπορεί να αντικαταστήσει πολλαπλούς δυαδικούς ελέγχους συσχετίσεων, με μια και μοναδική ανάλυση.

Σαν τελικό αποτέλεσμα, η ανάλυση παλινδρόμησης καταλήγει σε συναρτήσεις του τύπου:

$$Y = a_0 + f(x_i) + \varepsilon$$

Ανάλυση Παλινδρόμησης (3/8)

Στην πραγματικότητα, η ανάλυση παλινδρόμησης αποτελεί μια οικογένεια μεθόδων η οποία εξυπηρετεί ένα μεγάλο σύνολο διαφορετικών στόχων.

Τύπος παλινδρόμησης	Τύπος εξαρτημένης μεταβλητής (Y)	Πλήθος ανεξάρτητων μεταβλητών	Τύπος ανεξάρτητων μεταβλητών
Απλή γραμμική	Ποσοτική	1	Ποσοτική
Πολλαπλή γραμμική	Ποσοτική	n	Ποσοτικές
Μη γραμμική (εκτίμηση καμπύλης)	Ποσοτική	n	Ποσοτικές
Διχοτομική λογιστική	Ονομαστική - Δίτιμη	n	Κάθε τύπου
Πολυονομαστική λογιστική	Ονομαστική - Πολύτιμη	n	Κάθε τύπου
Διατακτική λογιστική	Διατακτική	n	Κάθε τύπου

Στην πραγματικότητα, η ανάλυση παλινδρόμησης αποτελεί μια οικογένεια μεθόδων η οποία εξυπηρετεί ένα μεγάλο σύνολο διαφορετικών στόχων.

Ανάλυση Παλινδρόμησης (4/8)

Θέλω να:	Πρέπει να επιλέξω:
Βρω τη συσχέτιση μεταξύ βάρους (Y) και ύψους σώματος	Απλή γραμμική
Βρω τη συσχέτιση μεταξύ βάρους (Y) και: ύψους σώματος (X _i), ηλικίας (X _{ii})	Πολλαπλή γραμμική
Να μελετήσω πως επηρεάζονται οι πιθανότητες να περάσει κάποιος φοιτητής το μάθημα του Marketing (Y) από : το εξάμηνο του (X _i), το αν το έχει ξαναδώσει (X _{ii}) και το πλήθος των διαλέξεων που παρακολούθησε (X _{iii})	Διχοτομική λογιστική
Να διερευνήσω ποιον κύκλο σπουδών θα επιλέξει ένας μηχανολόγος (Y), βάσει των βαθμολογιών του σε όλα τα μαθήματα των δύο πρώτων ετών (X _i ,...X _n) και του φύλου του (X _{n+1})	Πολυνομαστική λογιστική
Να μελετήσω πως επηρεάζεται η βαθμολογία ενός φοιτητή στο μάθημα του Marketing (Y) από: το εξάμηνο του (X _i), το αν το έχει ξαναδώσει (X _{ii}) και το πλήθος των διαλέξεων που παρακολούθησε (X _{iii})	Διατακτική λογιστική



Τα μοντέλα γραμμικής παλινδρόμησης βασίζονται στον υπολογισμό των βαρυτήτων (B_i) των ανεξάρτητων μεταβλητών, με στόχο τη σύνταξη της συνάρτησης $Y = B_0 + B_i X_i + \dots + B_n X_n$

Τα μοντέλα λογιστικής παλινδρόμησης βασίζονται στον υπολογισμό των σχετικών πιθανοτήτων (odds) εμφάνισης των εναλλακτικών τιμών της εξαρτημένης μεταβλητής (π.χ. Ναι ή Όχι), μέσω της αλλαγής των τιμών των ανεξάρτητων μεταβλητών

Ανάλυση Παλινδρόμησης (5/8)

Παράδειγμα 1: Απλή γραμμική παλινδρόμηση (1/2)

Σε μια πρόσφατη έρευνα του ΕΜΠ, μετρήθηκε το ποσοστό ανεργίας των διπλωματούχων του ΕΜΠ που αποφοίτησαν μεταξύ του 2002 και του 2010. Τα αποτελέσματα ανά έτος παρουσιάζονται στον πίνακα

Έτος (X)	Κωδ/ποίηση	% ανεργίας (Y)
2002	1	4,20%
2003	2	5,10%
2004	3	4,90%
2005	4	7,40%
2006	5	6,80%
2007	6	9,40%
2008	7	8,60%
2009	8	10,10%
2010	9	11,70%

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,964 ^a	,929	,919	,73221

Το μοντέλο εξηγεί το 92,9% της διακύμανσης!!!

Coefficients^a

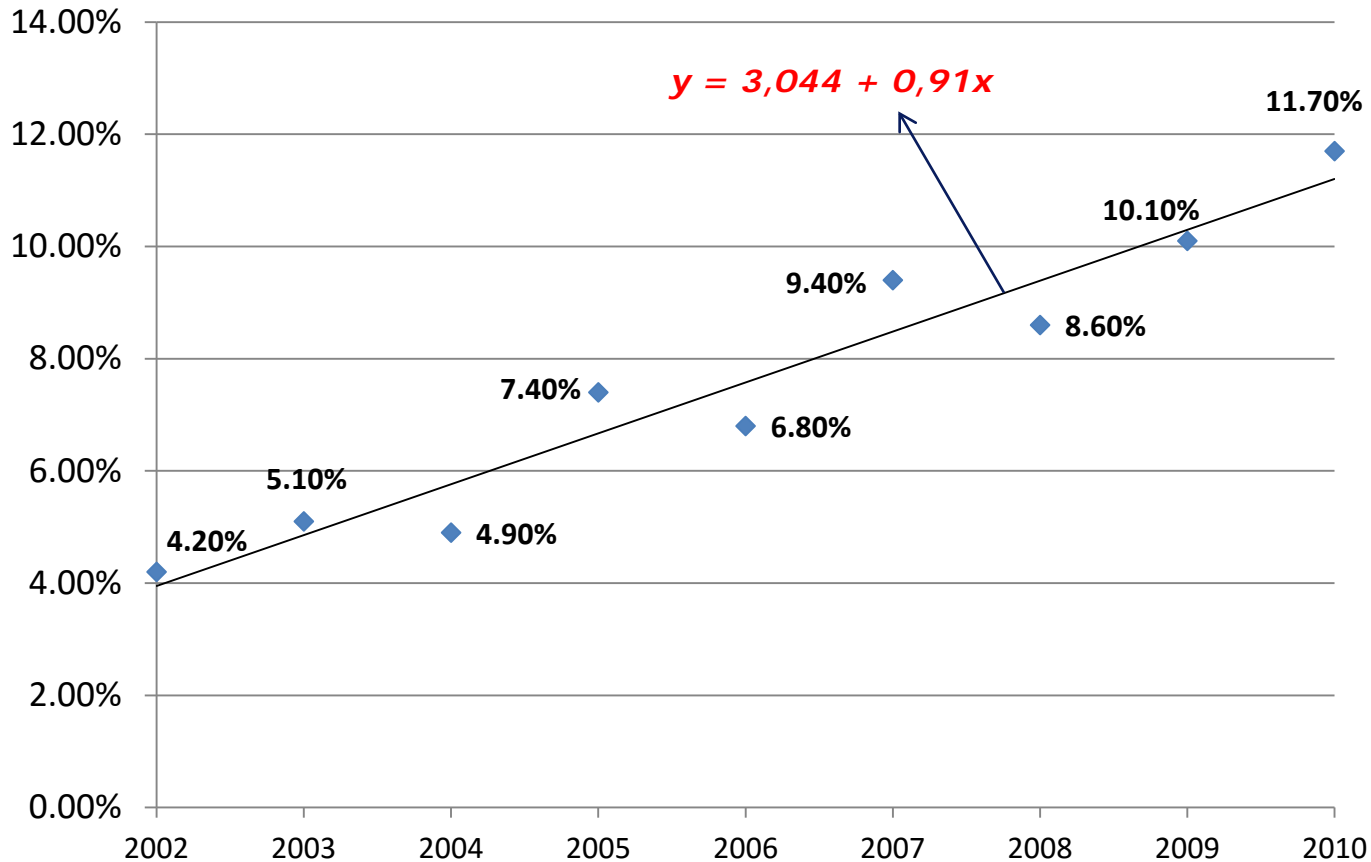
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3,044	,532		5,723	,001
	Year	,907	,095	,964	9,592	,000

$$y = 3,044 + 0,91x$$

Για κάθε έτος από το 2002 έως το 2010 το ποσοστό ανεργίας αυξάνεται κατά 1% περίπου

Ανάλυση Παλινδρόμησης (6/8)

Παράδειγμα 1: Απλή γραμμική παλινδρόμηση (2/2)



Ανάλυση Παλινδρόμησης (7/8)

Παράδειγμα 2: Διχοτομική λογιστική παλινδρόμηση (1/2)

Στην ίδια έρευνα, οι μηχανικοί ρωτήθηκαν κατά πόσο είναι συνεπείς στις ασφαλιστικές τους εισφορές προς το ΤΣΜΕΔΕ.

Με δεδομένο ότι οι εταιρείες των μισθωτών πληρώνουν το Ταμείο, αναζητήθηκαν οι παράγοντες που επηρεάζουν την πληρωμή όσον αφορά τους ελεύθερους επαγγελματίες.

Προκειμένου να διερευνηθούν οι παράγοντες που επηρεάζουν την πληρωμή του Ταμείου, χρησιμοποιείται η διχοτομική παλινδρόμηση με τιμές εξαρτημένης μεταβλητής 0: δεν πληρώνω & 1: πληρώνω

Εξετάζεται ταυτόχρονα η επίδραση:

- Της ειδικότητας (9 Σχολές του ΕΜΠ)
- Των ετήσιων εισοδημάτων (μετρημένα σε κλίμακα 0-5.000€/ χρόνο, 5.000 € – 10.000€/ χρόνο κλπ)

Ανάλυση Παλινδρόμησης (8/8)

Παράδειγμα 2: Διχοτομική λογιστική παλινδρόμηση (2/2)

Παράγοντες	B	Exp(B)	Wald	Sig.
Αρχιτέκτονες Μηχανικοί	-1,670	,188	13,898	,000*
Εισοδήματα 0-€5.000	-1,328	,265	23,231	,000*
Εισοδήματα €5.000 - €10.000	-,827	,438	8,089	,004*

Πως διαβάζω τον Πίνακα;

- Διακρίνω τις μεταβλητές με τιμή sig < 0.05 (σημαντική συνάφεια). Οι μεταβλητές με sig > 0,05 δεν επηρεάζουν σημαντικά την τιμή της εξαρτημένης. Για παράδειγμα, όλες οι ειδικότητες πλην Αρχιτεκτόνων δεν επηρεάζουν το αποτέλεσμα της πληρωμής των ασφαλιστικών εισφορών

- Για κάθε γραμμή του Πίνακα, παρατηρώ τα B και Exp(B), τα οποία με βοηθούν να καταλάβω πως επηρεάζει η τιμή το αποτέλεσμα.

- Για $B = 0$ ή $Exp(B) = 1$, οι πιθανότητες δεν αλλάζουν

- Για $B < 0$ ή $Exp(B) < 1$, οι πιθανότητες μειώνονται

- Για $B > 0$ ή $Exp(B) > 1$, οι πιθανότητες αυξάνονται

Π.χ.: Οι Αρχιτέκτονες έχουν περίπου πέντε (1/0,188) φορές λιγότερες πιθανότητες να πληρώνουν τις ασφαλιστικές εισφορές τους, εφόσον οι μεταβλητές των υπόλοιπων σημαντικών μεταβλητών παραμείνουν σταθερές!

Περιεχόμενα

1. Έλεγχοι Συσχετίσεων Δύο Μεταβλητών
2. Ανάλυση Παλινδρόμησης
3. Ανάλυση Παραγόντων
4. Ανάλυση Συστάδων
5. Ανάλυση Αντιστοιχιών

Ανάλυση Παραγόντων (1/9)

Η ανάλυση παραγόντων χρησιμοποιείται για την:

- Αναγνώριση των διαστάσεων (παραγόντων) που εξηγούν τη συσχέτιση μεταξύ ενός σετ μεταβλητών. Με άλλα λόγια, η σύνθεση ενός παράγοντα από συγκεκριμένες μεταβλητές με μεγάλη επίδραση μας δίνει να καταλάβουμε τη συσχέτιση αυτών των μεταβλητών.
- Αναγνώριση ενός μικρότερου αριθμού μη συσχετιζόμενων μεταβλητών ώστε να χρησιμοποιηθούν σε μία επόμενη πολυμεταβλητή ανάλυση (π.χ. ανάλυση παλινδρόμησης, ανάλυση κατά συστάδες, διαχωριστική ανάλυση).
- Αναγνώριση ενός σετ σημαντικών μεταβλητών από ένα σύνολο περισσότερων μεταβλητών ώστε να διακριθούν οι μεταβλητές που προσφέρουν αξία και πρέπει να χρησιμοποιηθούν στις επόμενες αναλύσεις.

Η παραγοντική ανάλυση μπορεί να είναι:

- Διερευνητική (exploratory), δηλαδή να βοηθάει να ανακαλύψουμε και να ταυτοποιήσουμε μη παρατηρούμενους ή λανθάνοντες (latent) παράγοντες, ή
- Επιβεβαιωτική (confirmatory), όπου ελέγχεται αν ένα σύνολο μεταβλητών που χρησιμοποιούμε επιβεβαιώνουν μια θεωρητική συσχέτιση (προϋποθέτει υφιστάμενη γνώση)

Ανάλυση Παραγόντων (2/9)

Παράδειγμα:

Υποθέτουμε μια έρευνα αγοράς η οποία στοχεύει στην αξιολόγηση της ποιότητας των ολοκληρωμένων πληροφοριακών συστημάτων τύπου ERP. Στο ερωτηματολόγιο χρησιμοποιήθηκαν 7 μεταβλητές (χαρακτηριστικά ποιότητας) οι οποίες βαθμονομήθηκαν σε κλίμακες από το 1 έως το 10. Οι αρχικές μεταβλητές (ερωτήσεις) είναι:

- Είναι εύκολο να δοθεί στο σύστημα η εντολή που θέλουμε
- Η εκμάθηση χειρισμού του συστήματος είναι εύκολη
- Οι χρήστες μπορούν να βασιστούν στο σύστημα
- Το σύστημα είναι εύκολο στη χρήση
- Το σύστημα παρέχει διαθεσιμότητα και σαφήνεια πληροφοριών
- Το σύστημα παρέχει πληροφορίες σε σύντομο χρονικό διάστημα
- Το σύστημα υποστηρίζει τη λήψη Διοικητικών αποφάσεων

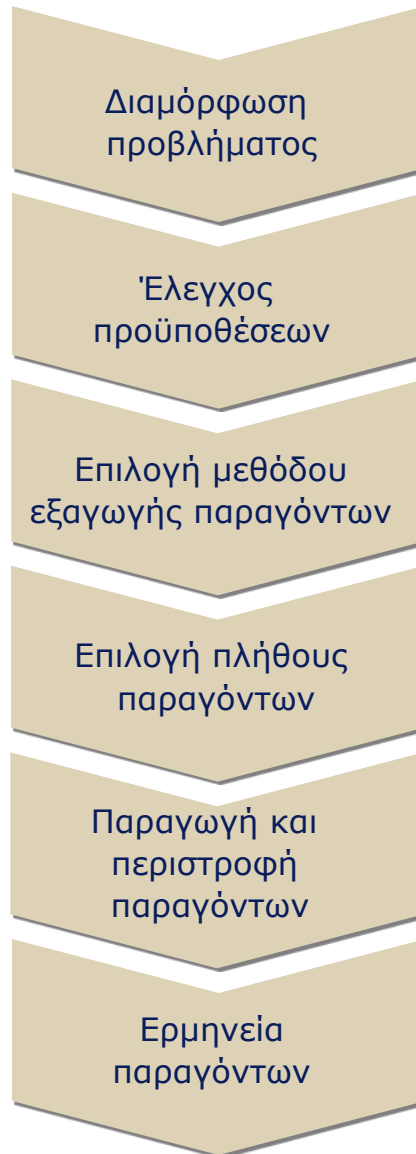
Ανάλυση Παραγόντων (3/9)

Μετά την εκτέλεση της ανάλυσης παραγόντων, διαμορφώθηκαν 2 νέες μεταβλητές (παράγοντες), κάθε ένας εκ των οποίων συνοψίζει την πληροφορία κάποιων εκ των αρχικών μεταβλητών

Αρχικές μεταβλητές	Παράγοντας 1	Παράγοντας 2
Είναι εύκολο να δοθεί στο σύστημα η εντολή που θέλουμε	✓	
Η εκμάθηση χειρισμού του συστήματος είναι μια εύκολη υπόθεση	✓	
Οι χρήστες μπορούν να βασιστούν στο σύστημα		✓
Το σύστημα είναι εύκολο στη χρήση	✓	
Το σύστημα παρέχει διαθεσιμότητα και σαφήνεια πληροφοριών		✓
Το σύστημα παρέχει πληροφορίες σε σύντομο χρονικό διάστημα		✓
Το σύστημα υποστηρίζει τη λήψη Διοικητικών αποφάσεων		✓

Αναγνωρίζοντας τις μεταβλητές που φορτίζουν κάθε παράγοντα, μπορούμε να αναγνωρίσουμε ότι οι 2 διαστάσεις ποιότητας των πληροφοριακών συστημάτων παραπέμπουν στην **ευκολία χρήσης** (Παράγοντας 1) και στη **διαχείριση πληροφορίας** (Παράγοντας 2)

Ανάλυση Παραγόντων (4/9)



Ανάλυση Παραγόντων (5/9)

Διαμόρφωση
προβλήματος

Έλεγχος
προϋποθέσεων

- Οι μεταβλητές επιλέγονται με βάση παλαιότερη έρευνα, ή υφιστάμενες θεωρίες, ή κρίση του ερευνητή
 - Οι μεταβλητές θα πρέπει να είναι ποσοτικές ή διατακτικές
 - Το δείγμα πρέπει να αποτελείται από τουλάχιστον 100 παρατηρήσεις, ενώ τυπικά απαιτούνται 10 – 20 παρατηρήσεις για κάθε εισερχόμενη μεταβλητή
-
- Έλεγχος γραμμικής συσχέτισης κατά Pearson
 - Bartlett's Test of Sphericity: διερεύνηση του κατά πόσο οι δυαδικές συσχετίσεις που παρουσιάζονται στον πίνακα συσχετίσεων επαρκούν
 - Έλεγχος δείκτη **Keiser-Meyer-Olkin** (ΚΜΟ): Ο έλεγχος αξιολογεί την επάρκεια του δείγματος ($KMO > 0,6$)

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,728
Bartlett's Test of Sphericity	Approx. Chi-Square	188,928
	df	21
	Sig.	,000

Ζητάμε sig ≈ 0 , προκειμένου να απορριπτεται η υπόθεση περί ανεξαρτησία των μεταβλητών

Ανάλυση Παραγόντων (6/9)

Επιλογή μεθόδου
εξαγωγής παραγόντων

Η Ανάλυση Παραγόντων προσφέρει μια σειρά από εναλλακτικές μεθόδους για την εκτίμηση ή εξαγωγή (extraction) των παραγόντων, μεταξύ των οποίων

- η μέθοδος των **κυρίων συνιστωσών** (principal components) και,
- η **ανάλυση παραγόντων** (common factor analysis)

Ανάλυση Κυρίων Συνιστωσών	Παραγοντική Ανάλυση
Οι κύριες συνιστώσες διατηρούν τη μέγιστη διακύμανση (πληροφορία) των αρχικών μεταβλητών	Οι παράγοντες διατηρούν τη βασική διακύμανση των αρχικών μεταβλητών
Η ανάλυση αποσυνθέτει τον πίνακα συσχετίσεων	Η ανάλυση αποσυνθέτει τον προσαρμοσμένο πίνακα συσχετίσεων
Μοναδιαίες τιμές στους διαγώνιους του πίνακα συσχετίσεων	Οι διαγώνιοι του πίνακα συσχετίσεων προσαρμόζονται σε μοναδικούς παράγοντες
Ελαχιστοποιείται το άθροισμα των τετραγώνων των καθέτων αποστάσεων στους άξονες των συστατικών	Υπολογίζει τους παράγοντες που επηρεάζουν τις απαντήσεις ως προς τις αρχικές μεταβλητές
Οι τιμές των παραγόμενων συστατικών διαμορφώνονται ως γραμμικός συνδυασμός των αρχικών μεταβλητών αφού σταθμιστούν από τα ιδιοδιανύσματα (eigenvectors)	Οι παρατηρούμενες (αρχικές) μεταβλητές αποτελούνται από γραμμικούς συνδυασμούς των παραγόντων

Ανάλυση Παραγόντων (7/9)

Επιλογή πλήθους παραγόντων

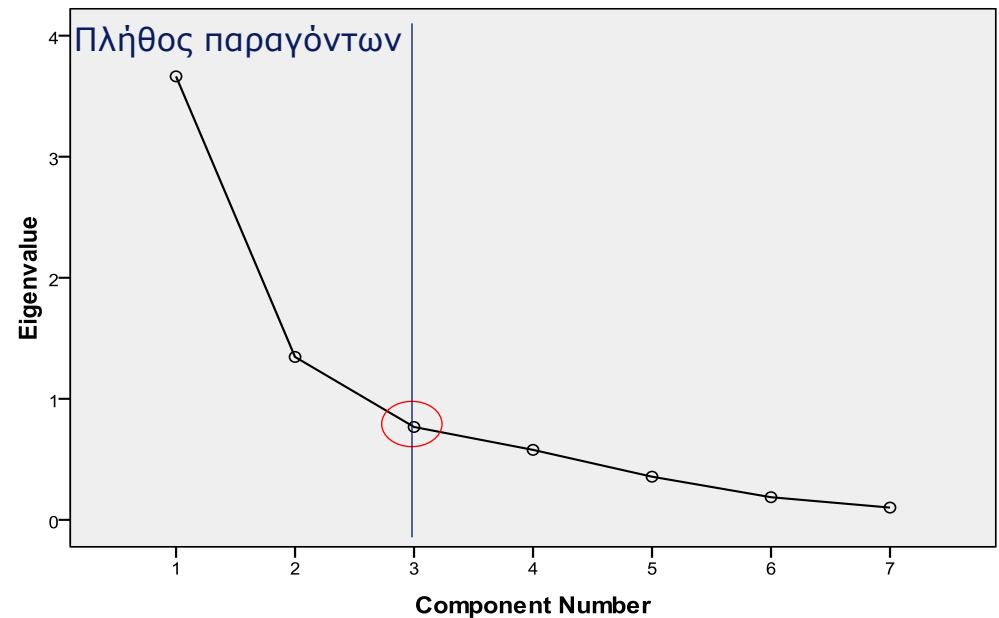
Η εκτίμηση του πλήθους των παραγόντων μπορεί να πραγματοποιηθεί είτε:

- Αριθμητικά, μέσω της διατήρησης των παραγόντων με ιδιοτιμή (eigenvalue) μεγαλύτερη της μονάδας. Η ιδιοτιμή εκφράζει το πλήθος των αρχικών μεταβλητών, των οποίων την πληροφορία εξηγεί ο νέος παράγοντας
- Γραφικά, μέσω ενός γραφήματος ιδιοτιμών (Y) / συνιστωσών (X) εντοπίζοντας το σημείο στο οποίο η γραμμή αρχίζει να παραλληλίζεται με τον άξονα των X

Total Variance Explained			
Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	3,664	52,338	52,338
2	1,346	19,227	71,565
3	,767	10,961	82,527
4	,578	8,259	90,786
5	,357	5,095	95,880
6	,187	2,672	98,552
7	,101	1,448	100,000

Extraction Method: Principal Component Analysis.

Scree Plot



Ανάλυση Παραγόντων (8/9)

Παραγωγή και περιστροφή παραγόντων

- Παράγεται ο αρχικός πίνακας φορτίσεων της ανάλυσης. Κάθε μία από τις αρχικές μεταβλητές «φορτίζει» με μία τιμή έναν από τους νέους παράγοντες. Όσο η φόρτιση τείνει προς τη μονάδα, τόσο πιο έντονη είναι η επίδραση της μεταβλητής σε αυτόν
- Μέσω της περιστροφής (rotation) του αρχικού πίνακα, οι αρχικοί παράγοντες γίνονται πιο ερμηνεύσιμοι.
- Θεωρούμε ως σημαντική φόρτιση, κάθε τιμή μεγαλύτερη του 0,5

Αρχικός πίνακας παραγόντων

Component Matrix ^a		
	Component	
	1	2
Είναι εύκολο να δοθεί στο σύστημα η εντολή που θέλουμε	,827	-,137
Η εκμάθηση χειρισμού του συστήματος είναι εύκολη	,732	-,617
Οι χρήστες μπορούν να βασιστούν στο σύστημα	,739	,141
Το σύστημα είναι εύκολο στη χρήση	,794	-,524
Το σύστημα παρέχει διαθεσιμότητα και σαφήνεια πληροφοριών	,728	,469
Το σύστημα παρέχει πληροφορίες σε σύντομο χρονικό διάστημα	,733	,427
Το σύστημα υποστηρίζει τη λήψη Διοικητικών αποφάσεων	,446	,499
Extraction Method: Principal Component Analysis.		
a. 2 components extracted.		

Πίνακας παραγόντων μετά την περιστροφή

Rotated Component Matrix ^a		
	Component	
	1	2
Είναι εύκολο να δοθεί στο σύστημα η εντολή που θέλουμε	,696	,468
Η εκμάθηση χειρισμού του συστήματος είναι εύκολη	,956	,052
Οι χρήστες μπορούν να βασιστούν στο σύστημα	,442	,609
Το σύστημα είναι εύκολο στη χρήση	,938	,163
Το σύστημα παρέχει διαθεσιμότητα και σαφήνεια πληροφοριών	,209	,841
Το σύστημα παρέχει πληροφορίες σε σύντομο χρονικό διάστημα	,241	,813
Το σύστημα υποστηρίζει τη λήψη Διοικητικών αποφάσεων	-,017	,669
Extraction Method: Principal Component Analysis.		
Rotation Method: Varimax with Kaiser Normalization.		
a. Rotation converged in 3 iterations.		

Ανάλυση Παραγόντων (9/9)

Ερμηνεία νέων παραγόντων

- Στο τελευταίο βήμα, προσπαθούμε να αντιληφθούμε το φυσικό νόημα των παραγόντων που διαμορφώθηκαν
- Πρακτικά, αναγνωρίζουμε την κοινή διάσταση των αρχικών μεταβλητών που φορτίζουν τον ίδιο παράγοντα

Rotated Component Matrix ^a		
	Component	
	1	2
Είναι εύκολο να δοθεί στο σύστημα η εντολή που θέλουμε	,696	,468
Η εκμάθηση χειρισμού του συστήματος είναι εύκολη	,956	,052
Οι χρήστες μπορούν να βασιστούν στο σύστημα	,442	,609
Το σύστημα είναι εύκολο στη χρήση	,938	,163
Το σύστημα παρέχει διαθεσιμότητα και σαφήνεια πληροφοριών	,209	,841
Το σύστημα παρέχει πληροφορίες σε σύντομο χρονικό διάστημα	,241	,813
Το σύστημα υποστηρίζει τη λήψη Διοικητικών αποφάσεων	-,017	,669

Ευκολία χρήσης

Διαχείριση πληροφορίας

Περιεχόμενα

1. Έλεγχοι Συσχετίσεων Δύο Μεταβλητών
2. Ανάλυση Παλινδρόμησης
3. Ανάλυση Παραγόντων
4. Ανάλυση Συστάδων
5. Ανάλυση Αντιστοιχιών

Ανάλυση Συστάδων (1/8)

Η ανάλυση συστάδων χρησιμοποιείται για την:

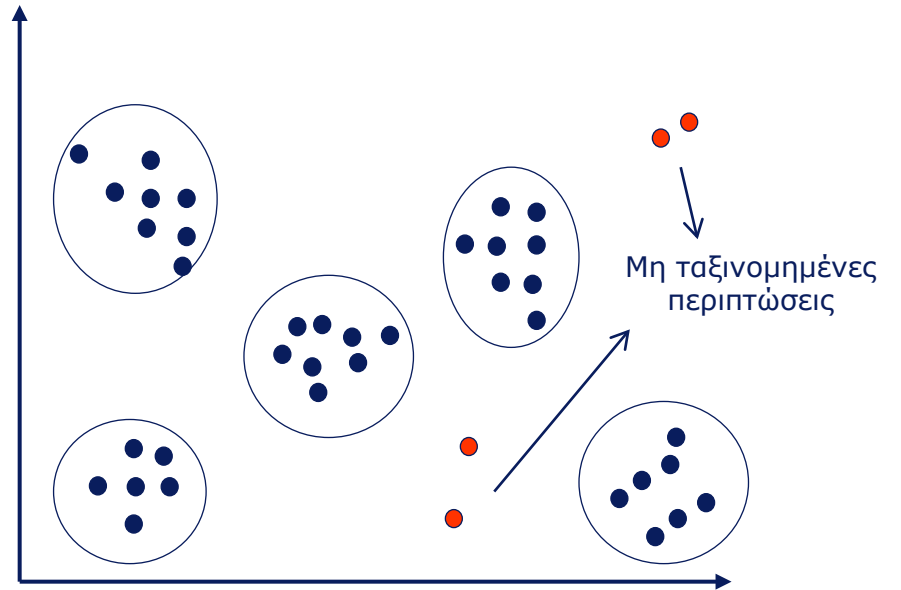
- Ταξινόμηση παρατηρήσεων σε σχετικά ομογενείς ομάδες με βάση επιλεγμένες μεταβλητές. Οι παρατηρήσεις στην ίδια ομάδα είναι σχετικά όμοιες και διαφορετικές σε σχέση με αυτές σε άλλες ομάδες.
- Το αποτέλεσμα της ανάλυσης συστάδων συναντάται στην επιστήμη του Marketing ως τμηματοποίηση (segmentation). Τα τμήματα που διαμορφώνονται συνήθως χρησιμοποιούνται:

Σε αντίθεση με την ανάλυση παραγόντων:

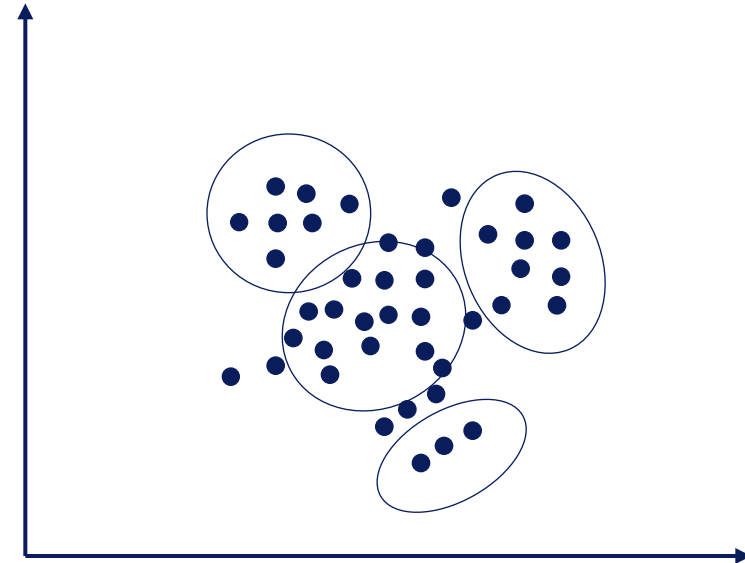
- Οι μεταβλητές δεν μειώνονται, αλλά μας βοηθάνε να καταλάβουμε πως ταξινομούνται οι παρατηρήσεις του δείγματος με βάση τις ομοιότητες και διαφορές τους.
- Δεν βασίζεται σε κάποια αυστηρή στατιστική θεωρία αλλά σε ευρετικούς αλγορίθμους ομαδοποίησης.

Ανάλυση Συστάδων (2/8)

Ιδεατή περίπτωση



Πιθανή περίπτωση



Ανάλυση Συστάδων (3/8)

Παράδειγμα:

Μια εταιρεία παραγωγής κινητών τηλεφώνων (smartphones) ετοιμάζεται να λανσάρει το νέο της προϊόν, το οποίο χαρακτηρίζεται από υψηλή ποιότητα αλλά και αρκετά ακριβή τιμή. Προκειμένου να εντοπίσει τους υποψήφιους πελάτες στους οποίους οφείλει να εστιάσει τη διαφημιστική της καμπάνια, ανάθεσε σε μια εξειδικευμένη εταιρεία την υλοποίηση μιας έρευνας αγοράς.

Η εταιρεία υλοποίησε μια έρευνα αγοράς, συλλέγοντας δεδομένα σε μια σειρά δημογραφικών χαρακτηριστικών και αγοραστικών προτιμήσεων, μεταξύ των οποίων:

Αγοραστικές προτιμήσεις	Δημογραφικά
Πόσο πιθανό είναι να αγοράσετε μια νέα συσκευή το προσεχές διάστημα*	Φύλλο
Πόσο σημαντικός παράγοντας είναι για εσάς:	Ηλικία
Η ποιότητα της συσκευής*	Οικογενειακή κατάσταση
Το κόστος της συσκευής*	
<i>(από το 1 – 10, όπου 1 καθόλου σημαντικό, 10 πολύ σημαντικό)</i>	

Ανάλυση Συστάδων (4/8)

Παράδειγμα:

Με το τέλος της έρευνας, η εταιρεία παρέδωσε τα παρακάτω αποτελέσματα στην εταιρεία:

Συστάδες	Πιθανότητα αγοράς συσκευής	Αξιολόγηση παράγοντα ποιότητας	Αξιολόγηση παράγοντα κόστους	Προφίλ
1	Υψηλή	Υψηλή	Χαμηλή	Γυναίκες, 35 ετών, ελεύθερες
2	Μέτρια	Υψηλή	Χαμηλή	Άνδρες, 35 ετών, παντρεμένοι ή ελεύθεροι
3	Μέτρια	Μέτρια	Μέτρια	Άνδρες ή Γυναίκες 25 ετών (ελεύθεροι)
3	Χαμηλή	Μέτρια	Μέτρια	Γυναίκες 30 ετών παντρεμένες
4	Υψηλή	Χαμηλή	Υψηλή	Άνδρες ή γυναίκες 20 ετών

Αξιολογώντας τα παραπάνω αποτελέσματα, η εταιρεία αποφάσισε να διαμορφώσει την καμπάνια της στοχεύοντας γυναίκες περίπου 35 ετών, αναπτύσσοντας διαφημίσεις που αναδεικνύουν το προφίλ της ανεξάρτητης γυναίκας που αφιερώνει το χρόνο της στη δουλειά και τη διασκέδαση...

Ανάλυση Συστάδων (5/8)

Διαμόρφωση
προβλήματος

Επιλογή
μέτρου απόστασης

Επιλογή μεθόδου
(αλγορίθμου)
συσταδοποίησης

Επιλογή πλήθους
συστάδων

Αναγνώριση και ερμηνεία
συστάδων

Ανάλυση Συστάδων (6/8)

Διαμόρφωση προβλήματος

- Οι μεταβλητές επιλέγονται με βάση παλαιότερη έρευνα, ή υφιστάμενες θεωρίες, ή τα αποτελέσματα προηγούμενης ανάλυσης (π.χ. ανάλυση παραγόντων).
- Σύμφωνα με τους πρώτους ερευνητές της μεθόδου, οι μεταβλητές όφειλαν να είναι ποσοτικές. Ωστόσο, τα σύγχρονα στατιστικά πακέτα επιτρέπουν την χρησιμοποίηση κάθε τύπου μεταβλητής χωρίς ιδιαίτερο πρόβλημα.
- Βασικός στόχος της μεθόδου είναι η παραγωγή εύκολα αναγνωρίσιμων ομάδων. Κατά αυτόν τον τρόπο, μπορεί να χρειαστεί η προσθαφαίρεση μεταβλητών με στόχο την παραγωγή «ευανάγνωστων» αποτελεσμάτων

Επιλογή μέτρου απόστασης

- Η μέτρηση της ομοιότητας μεταξύ των παρατηρήσεων πραγματοποιείται με τη μορφή της (γεωμετρικής) απόστασης μεταξύ αυτών.
- Το μέτρο της απόστασης υπολογίζεται είτε με τη μέθοδο της λογαριθμικής πιθανοφάνειας (log-likelihood) για ποιοτικές μεταβλητές είτε με τη μορφή της ευκλείδειας απόστασης (Euclidean distance) για ποσοτικές μεταβλητές.
- Εφόσον οι μεταβλητές χρησιμοποιούν διαφορετική κλίμακα μέτρησης, οφείλουν να τυποποιηθούν (standardization)

Επιλογή μεθόδου (αλγορίθμου) συσταδοποίησης

- Οι μέθοδοι που χρησιμοποιούνται αφορούν στις κατηγορίες των **ιεραρχικών** και των **μη ιεραρχικών** μεθόδων
- Οι ιεραρχικές μέθοδοι ξεκινάνε με αριθμό συστάδων ίσο με εκείνο των παρατηρήσεων και προχωρούν σταδιακά σε ενώσεις των πιο κοντινών παρατηρήσεων.
- Οι μη ιεραρχικές (k-means) υπολογίζουν αλγεβρικά κέντρα (συστάδες) τα οποία έλκουν τις παρατηρήσεις μέσω της απόστασης προς αυτά. Στις μεθόδους της κατηγορίας ο αριθμός των συστάδων πρέπει να είναι εκ των προτέρων γνωστός.
- Αρκετά πακέτα προσφέρουν υβριδικές μεθόδους που δανείζονται χαρακτηριστικά και από τις δύο γενικές κατηγορίες (**διβηματική ανάλυση συστάδων**).

Ανάλυση Συστάδων (7/8)

Επιλογή πλήθους
συστάδων

Αναγνώριση και ερμηνεία
συστάδων

- Η επιλογή του πλήθους των συστάδων πραγματοποιείται με δύο τρόπους:
- Μέσω «αυθαίρετης» επιλογής του αναλυτή ή της υφιστάμενης γνώσης από προηγούμενες θεωρίες.
 - Μέσω της εκτέλεσης μιας ιεραρχικής μεθόδου που οδηγεί στη διάκριση του βέλτιστου πλήθους συστάδων (είτε γραφικά, είτε αριθμητικά)
- Η ερμηνεία των συστάδων πραγματοποιείται μέσω της παραγωγής των περιγραφικών μέτρων των χρησιμοποιούμενων μεταβλητών, για κάθε παραγόμενη συστάδα.
- Για ποσοτικές μεταβλητές παράγονται οι μέσες τιμές και για ποιοτικές μεταβλητές οι ποσοστιαίες συχνότητες.

Μέχρι και πρόσφατα, η μεθοδολογία της ανάλυσης συστάδων προέβλεπε τη χρησιμοποίηση μιας ιεραρχικής μεθόδου για την επιλογή του πλήθους των συστάδων και εν συνεχεία την επανάληψη της ανάλυσης μέσω μιας μη ιεραρχική μεθόδου για την διαμόρφωση των συστάδων, με δεδομένο ότι ο αλγόριθμος k-means δεν μπορεί να υπολογίσει αριθμό συστάδων αλλά είναι σε θέση να παράγει καλύτερης ποιότητας ομαδοποίηση.

Τα στατιστικά πακέτα προσφέρουν πλέον υβριδικούς αλγόριθμους που εκτελούν και τους δύο στόχους παράλληλα (αρχικά αναγνωρίζουν το πλήθος και στη συνέχεια διαμορφώνουν τις συστάδες τοποθετώντας σε αυτές τις παρατηρήσεις βάσει του επιλεγμένου μέτρου απόστασης)



Ανάλυση Συστάδων (8/8)

Επιστροφή στο παράδειγμα:

Η εταιρεία ερευνών αγοράς χρησιμοποίησε για την ανάλυση συστάδων τις τρεις μεταβλητές αξιολόγησης της αγοραστικής συμπεριφοράς (πιθανότητες αγοράς, αξιολόγηση παράγοντα ποιότητας, αξιολόγηση παράγοντα κόστους).

Πρακτικά, ήθελε να διακρίνει ομάδες με πολλές πιθανότητες αγοράς νέου κινητού, οι οποίες ταυτόχρονα να αγοράζουν προϊόντα βασιζόμενες κυρίως στη διάσταση της ποιότητας και όχι του κόστους.

Cluster Distribution		
Cluster	N	% of Total
1	285	17,4%
2	300	18,4%
3	389	23,8%
4	342	20,9%
5	306	18,7%

	Πιθανότητα αγοράς	Ποιότητα	Κόστος
Cluster	Mean	Mean	Mean
1	9,09	8,72	5,81
2	7,01	8,22	5,22
3	6,07	5,56	5,46
4	4,72	5,44	5,44
5	8,25	4,94	9,12

Στη συνέχεια, διασταύρωσε τις συστάδες με τα βασικά δημογραφικά χαρακτηριστικά προκειμένου να μπορεί να αναγνωρίσει το προφίλ των ομάδων που παράχθηκαν

Cluster	Ηλικία	Φύλο		Οικογενειακή κατάσταση	
	Μέση τιμή	Άνδρες	Γυναίκες	Άγαμοι	Έγγαμοι
1	34,15	12,1%	87,9%	84%	16%
2	35,28	86,5%	13,5%	51%	49%
3	25,64	48,2%	52,8%	92,2%	7,8%
4	29,52	22,6%	78,4%	6,5%	93,5%
5	20,87	74,9%	25,1%	100%	0%

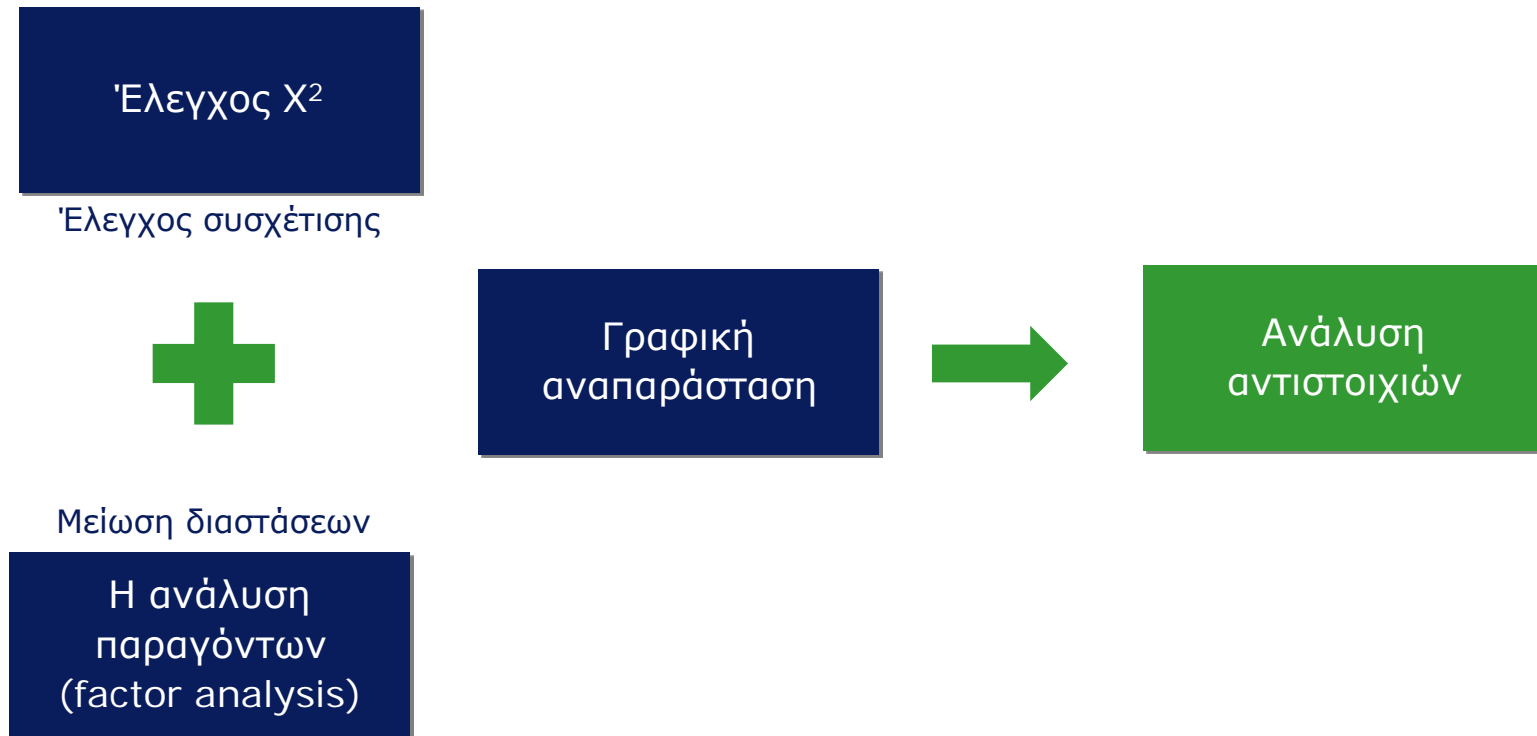
Περιεχόμενα

1. Έλεγχοι Συσχετίσεων Δύο Μεταβλητών
2. Ανάλυση Παλινδρόμησης
3. Ανάλυση Παραγόντων
4. Ανάλυση Συστάδων
5. Ανάλυση Αντιστοιχιών

Correspondence analysis (Ανάλυση Αντιστοιχιών) (1/4)

Η μέθοδος υποστηρίζει:

- τη διερεύνηση της **συσχέτισης μεταξύ δύο κατηγορικών μεταβλητών**,
- τη διευκρίνιση **της αιτίας της συσχέτισης** μέσω της παροχής πληροφοριών για τις επιμέρους τιμές των δύο μεταβλητών που έλκονται ή απωθούνται
- την **μείωση των διαστάσεων** ενός προβλήματος, και τέλος,
- την αναπαράσταση των αποτελεσμάτων με τη μορφή **γραφήματος**



Παράδειγμα:

Ο Δ/ντης Marketing της εταιρεία παραγωγής κινητών τηλεφώνων (smartphones) του προηγούμενου παραδείγματος (ανάλυση συστάδων), προετοιμάζεται για την παρουσίαση των αποτελεσμάτων προς το ΔΣ της εταιρείας του.

Προκειμένου να παρουσιάσει τα αποτελέσματα προς το ΔΣ της εταιρείας του, ζήτησε από την εταιρεία που έκανε την έρευνα να τον βοηθήσει προκειμένου να διατυπώσει τα συμπεράσματα με έναν λιγότερο αριθμητικό τρόπο.

Γνωρίζοντας πως η απαίτηση του πελάτη υποστηρίζεται από την ανάλυση αντιστοιχιών, η εταιρεία υποσχέθηκε να συνοψίσει όλη την χρήσιμη πληροφορία σε ένα γράφημα, μέσω του οποίου οι συσχετίσεις που διαπιστώθηκαν θα γίνονταν κατανοητές από όλους τους συμμετέχοντες.

Correspondence analysis (Ανάλυση Αντιστοιχιών) (3/4)

Προκειμένου να χρησιμοποιηθεί η ανάλυση αντιστοιχιών, ο αναλυτής οφείλει να έχει 2 κατηγορικές μεταβλητές, με περισσότερες από δύο τιμές η κάθε μία.

Η πρώτη μεταβλητή είχε ήδη παραχθεί και αφορούσε στις συστάδες που παράχθηκαν μέσω της αξιολόγησης των τριών αγοραστικών προτιμήσεων. Στις πέντε ομάδες απέδωσε τίτλους προκειμένου να είναι εύκολα αναγνωρίσιμες

	Πιθανότητα αγοράς	Ποιότητα	Κόστος
Cluster	Mean	Mean	Mean
1	9,09	8,72	5,81
2	7,01	8,22	5,22
3	6,07	5,56	5,46
4	4,72	5,44	5,44
5	8,25	4,94	9,12

Target 1

Target 2

Target 3

Target 4

Target 5

Έχοντας διακρίνει τα δημογραφικά χαρακτηριστικά που συσχετίζονται με τις παραπάνω ομάδες, προχώρησε στη δημιουργία μιας δεύτερης μεταβλητής η οποία ουσιαστικά διέκρινε το δείγμα βάσει ηλικίας, φύλου και οικογενειακής κατάστασης προσπαθώντας να προσομοιάσει τις συσχετίσεις. Η ονομαστική μεταβλητή περιελάμβανε τις ακόλουθες πέντε κατηγορίες:

1	>32,5	Γυναίκες	Άγαμοι
2	>32,5	Άνδρες	-
3	22,5< X < 27,5	-	Άγαμοι
4	27,5> X >32,5	Γυναίκες	Έγγαμοι
5	<22,5	Άνδρες	Άγαμοι

Demo 1

Demo 2

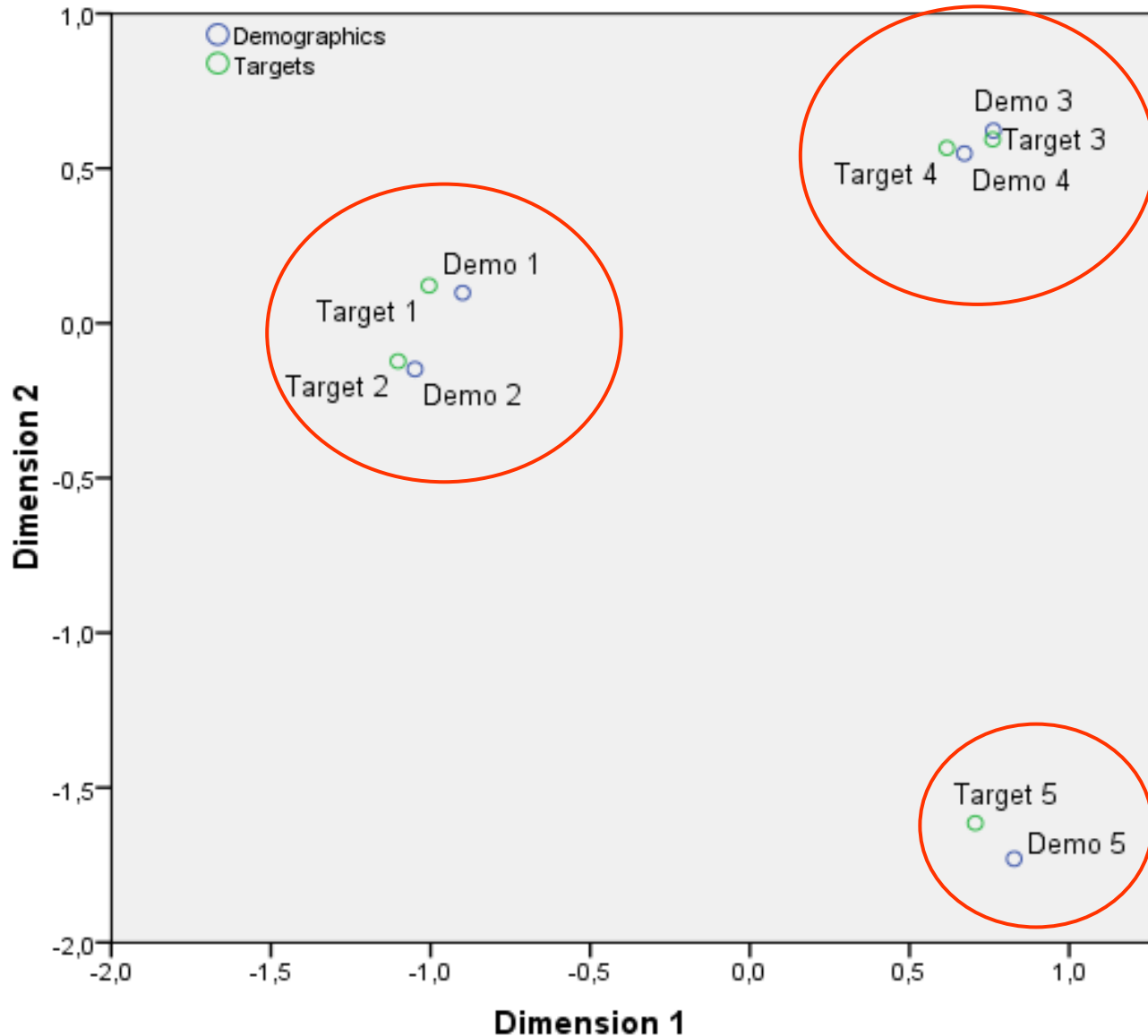
Demo 3

Demo 4

Demo 5

Correspondence analysis (Ανάλυση Αντιστοιχιών) (4/4)

Στη συνέχεια, υπέβαλλε τις δύο μεταβλητές στην ανάλυση...



Targets	Πιθανότητα αγοράς	Ποιότητα	Κόστος
1	9,09	8,72	5,81
2	7,01	8,22	5,22
3	6,07	5,56	5,46
4	4,72	5,44	5,44
5	8,25	4,94	9,12

Demos	Ηλικία	Φύλο	Οικ.
1	>32,5	Γυναίκες	Άγαμοι
2	>32,5	Άνδρες	-
3	22,5 < X < 27,5	-	Άγαμοι
4	27,5 > X > 32,5	Γυναίκες	Έγγαμοι
5	<22,5	Άνδρες	Άγαμοι

...and here's a chart that shows what you might see if you looked at a mountain range through a tennis racket.

