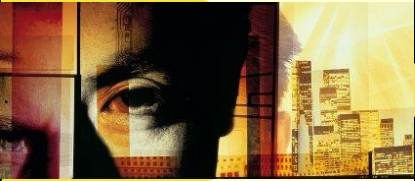
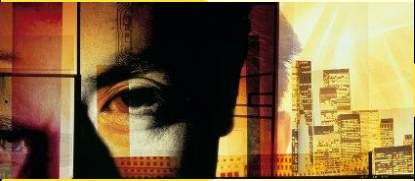


# Ανάλυση Δεδομένων



## Βασικές Αρχές της Ανάλυσης Δεδομένων

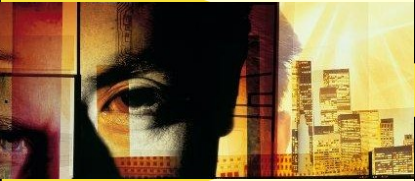
- ❖ Βασικός ρόλος της Στατιστικής είναι η συλλογή, παρουσίαση και ανάλυση στατιστικών στοιχείων.
- ❖ Τα στοιχεία οργανώνονται σε μεταβλητές οι οποίες περιγράφουν χαρακτηριστικά / φαινόμενα προς μελέτη.
- ❖ Ο σκοπός είναι να βγάλουμε συμπεράσματα για τον πληθυσμό, δηλαδή την ομάδα που αντιπροσωπεύουν τα στοιχεία.



# Βασικές Αρχές της Ανάλυσης Δεδομένων

Σκοπός της ανάλυσης είναι:

- Η εδραίωση αιτιωδών σχέσεων (αιτία – αποτέλεσμα).
- Η πρόβλεψη / πρόγνωση γεγονότων του «πραγματικού κόσμου» σύμφωνα με τα ερευνητικά αποτελέσματα.
- Η εύρεση απαντήσεων σε συγκεκριμένα «προβλήματα»
- Η διατύπωση συμπερασμάτων που αφορούν τον «πραγματικό κόσμο».



# Βασικές Αρχές της Ανάλυσης Δεδομένων

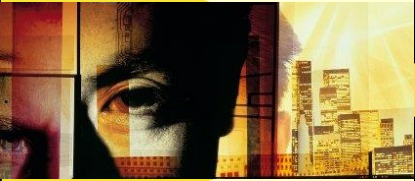
Πρέπει να γνωρίζουμε ότι:

- Η ανάλυση πρέπει απαραίτητα να περιλαμβάνει:
  - **Αιτιολόγηση της επιλογής και περιγραφή της στατιστικής μεθόδου** που χρησιμοποιήθηκε (χρήση βιβλιογραφίας),
  - **Επιστημονική συζήτηση**  
(Ερμηνεία Αποτελεσμάτων) (τι βρέθηκε; τι σημαίνει αυτό που βρέθηκε;)
  - **Συμπεράσματα**  
(πως μπορεί να χρησιμοποιηθεί αυτό που βρέθηκε;)



# Ποσοτική Ανάλυση Δεδομένων

- Επιλογή στατιστικού προγράμματος ανάλυσης δεδομένων.
- Δημιουργία φόρμας εισαγωγής / Κωδικοποίηση δεδομένων (προτείνεται η φόρμα να δημιουργείται μετά το τέλος της κατασκευής του ερωτηματολογίου).
- Καταχώρηση των δεδομένων.



# Βασικές Αρχές της Ανάλυσης Δεδομένων

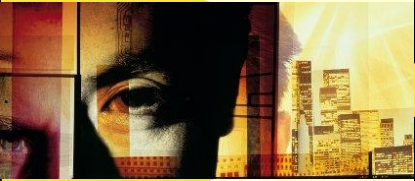
## **Ποσοτική μεταβλητή**

Μπορεί να μετρηθεί (δηλ. οι τιμές της έχει αριθμητικές ιδιότητες και χρησιμοποιούμε κάποια μονάδα μέτρησης για να εκφράσουμε τις τιμές της).

Παραδείγματα ποσοτικών μεταβλητών είναι η ηλικία, το βάρος, το ύψος, η αξία μίας μετοχής, ο δείκτης νοημοσύνης, κάθε μήκος, εμβαδόν ή όγκος, κάθε χρονική περίοδος, κ.λ.π.

## **Ποιοτική μεταβλητή**

Περιγράφει χαρακτηριστικά του πληθυσμού που μεταβάλλονται κατά ποιότητα ή είδος, αλλά όχι κατά μέγεθος.



## Βασικές Αρχές της Ανάλυσης Δεδομένων

Επιπλέον οι ποσοτικές μεταβλητές μπορεί να είναι:

**Συνεχείς**, (μεταβλητές που παίρνουν τιμές στο σύνολο των πραγματικών αριθμών  $\mathbb{R}$ , π.χ. βάρος, ύψος, αξία μετοχής. Από οικονομικής άποψης, κάθε χρηματικό μέγεθος θεωρείται συνεχής μεταβλητή).

**Διακριτές**, (μεταβλητές που παίρνουν τιμές στο σύνολο των ακεραίων  $\mathbb{N}$ , π.χ. δείκτης νοημοσύνης (όταν είναι ακέραιος και όχι δεκαδικός). Επίσης σημαντικό είναι ότι οι τιμές των διακριτών μεταβλητών ανήκουν σε ένα πεπερασμένο εύρος τιμών, δηλ. δεν εκτείνονται στο άπειρο).



# Βασικές Αρχές της Ανάλυσης Δεδομένων

**Μεγάλο ρόλο στις συνεχείς μεταβλητές παίζει η κλίμακα μέτρησης:**

➤ **Κλίμακα λόγου / Αριθμητικές (Ratio):** Είναι η πληρέστερη και χρησιμοποιείται στις περισσότερες ποσοτικές μεταβλητές. Οι τιμές τις επιτρέπουν την διάταξη των στοιχείων που μετρούν (μικρότερο / μεγαλύτερο), το 0 σημαίνει έλλειψη αυτού που μετριέται. Ίσες διαφορές τιμών σημαίνουν ίσες διαφορές μεταξύ των ποσοτήτων που μετρούν. Ο λόγος μεταξύ δυο τιμών δίνει τη σχέση τους (π.χ. διπλάσιο μέγεθος).

➤ **Κλίμακα διαστημάτων (Interval):** Οι τιμές καθορίζονται αυθαίρετα. Εκφράζουν διάταξη αλλά η τιμή 0 δεν σημαίνει έλλειψη ποσότητας. Εδώ ίσες διαφορές τιμών σημαίνουν ίσες διαφορές μεταξύ των ποσοτήτων που μετρούν, όμως ο λόγος μεταξύ δυο τιμών δεν δίνει τη σχέση μεταξύ των αντίστοιχων μεγεθών. Π.χ. η θερμοκρασία: Οι τιμές της επιτρέπουν τη διάταξη, αλλά δεν μπορούμε να πούμε ότι θερμοκρασία 30°C είναι διπλάσια ζέστη από ότι 15°C.





# Βασικές Αρχές της Ανάλυσης Δεδομένων

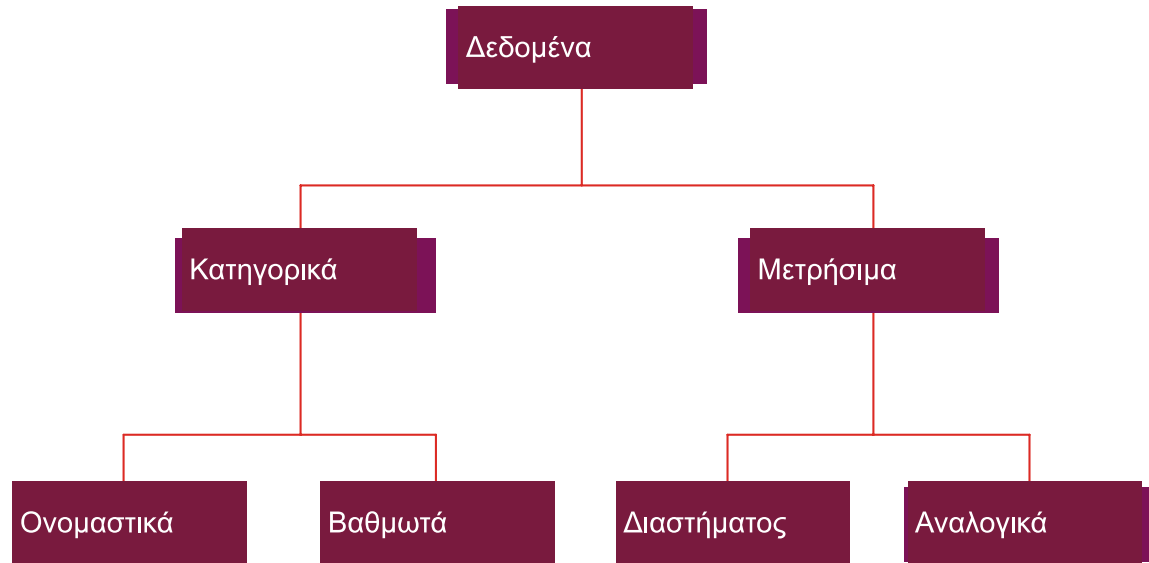
Οι δε ποιοτικές μεταβλητές μπορεί να είναι:

➤ **Διατάξιμες (ή ιεραρχικές)** (Ordinal): Μεταβλητές που παρόλο που δεν μπορούν να μετρηθούν έχουν ξεκάθαρη την έννοια του μεγαλύτερου / μικρότερου, άρα της διάταξης, π.χ. η κατάσταση της υγείας ενός ασθενή (πολύ σοβαρή, σοβαρή, μέτρια, ομαλή), το επίπεδο εκπαίδευσης (πρωτοβάθμια, δευτεροβάθμια, κλπ), ο βαθμός ικανοποίησης κάποιου από ένα προϊόν (πάρα πολύ / πολύ / λίγο).

➤ **Κατηγορικές** (μη διατάξιμες) (Nominal), δηλαδή μεταβλητές που περιγράφουν αποκλειστικά μη μετρήσιμα, μη διατάξιμα μεγέθη π.χ. επάγγελμα, το φύλο, η υπηκοότητα, το θρήσκευμα, ομάδα, είδος μουσικής, κ.λ.π.

Τα ποσοτικά ή ποιοτικά στοιχεία λέγονται **διαστρωματικά** αν αναφέρονται σε συγκεκριμένη χρονική περίοδο και **χρονολογικά** στη διαχρονική εξέλιξη ενός χαρακτηριστικού (π.χ. τιμή μετοχής, τιμή σπιτιών, δείκτης τιμών καταναλωτή, και γενικά μακροοικονομικά μεγέθη).

# Τύποι ποσοτικών δεδομένων

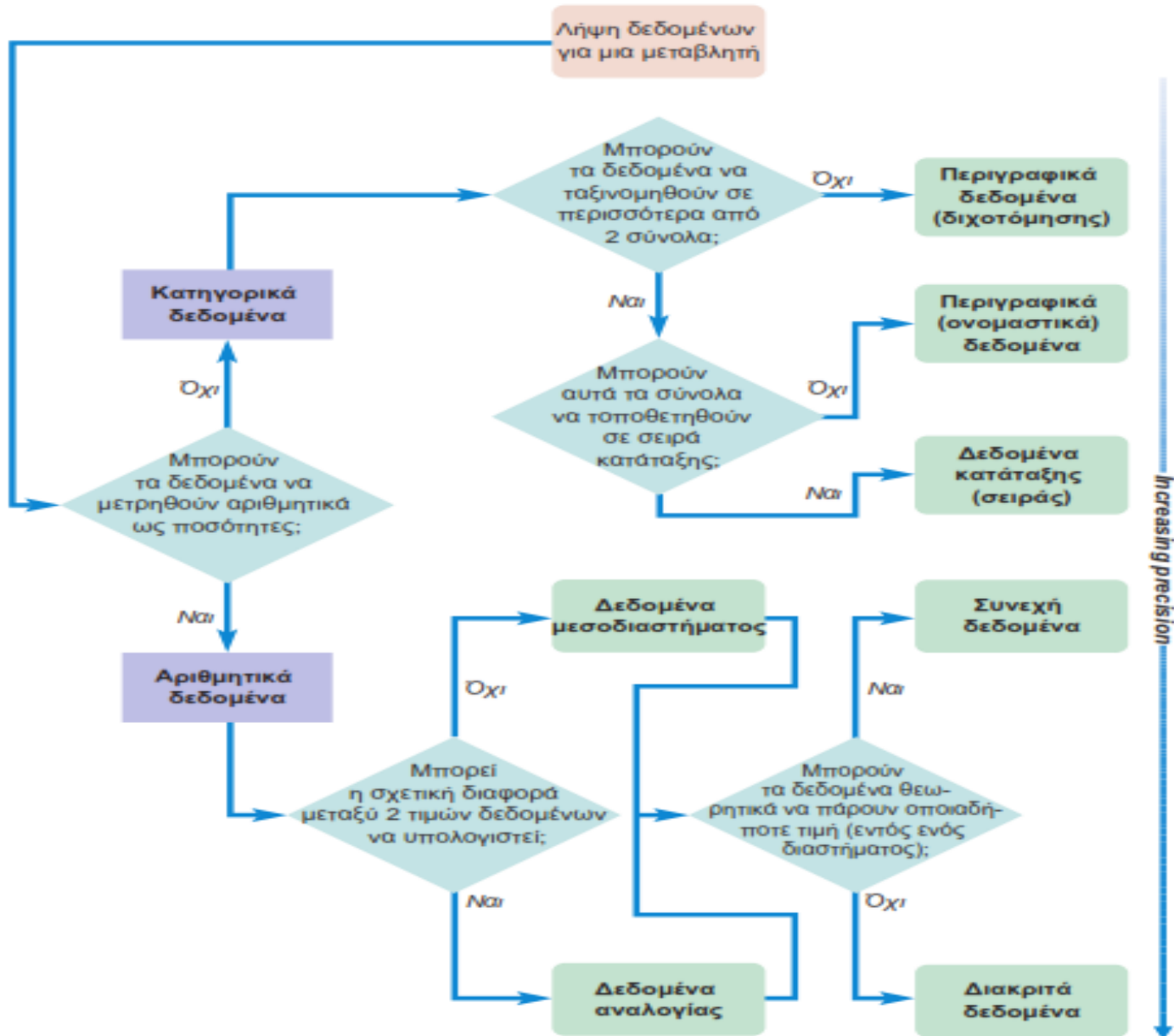


Βαθμός ακρίβειας



	<b>Ονομαστικά</b>	<b>Βαθμωτά</b>	<b>Διαστήματος</b>	<b>Αναλογικά</b>
Παράδειγμα χρήσης	Είδος επιχείρησης	Προτιμήσεις καταναλωτών	Θερμοκρασία	Πωλήσεις
	Είδος προϊόντος	Ιεραρχία οργάνωσης	Αρτηριακή πίεση	Κόστη
	Τοποθεσία οργανισμού	Στάσεις		Αριθμός πελατών
Στατιστικοί έλεγχοι	Μη-παραμετρικοί έλεγχοι			Παραμετρικοί έλεγχοι

# Τύποι ποσοτικών δεδομένων





# Ποσοτική Ανάλυση Δεδομένων

- **Μονομεταβλητή Ανάλυση  
(Univariate analysis)**

Εξέταση διαφόρων στατιστικών μέτρων που αφορούν μόνο μία μεταβλητή (περιγραφική στατιστική - μέτρα κεντρικής τάσης, μέτρα διασποράς κλπ)

- **Διμεταβλητή Ανάλυση  
(Bivariate analysis)**

Εξέταση των πιθανών σχέσεων μεταξύ δύο μόνο μεταβλητών (ανάλυση συσχέτισης, ANOVA κλπ)

- **Πολυμεταβλητή Ανάλυση  
(Multivariate analysis)**

Εξέταση των σχέσεων μεταξύ πολλών μεταβλητών (παραγοντική ανάλυση, ανάλυση παλινδρόμησης κλπ)



# Ποσοτική Ανάλυση Δεδομένων

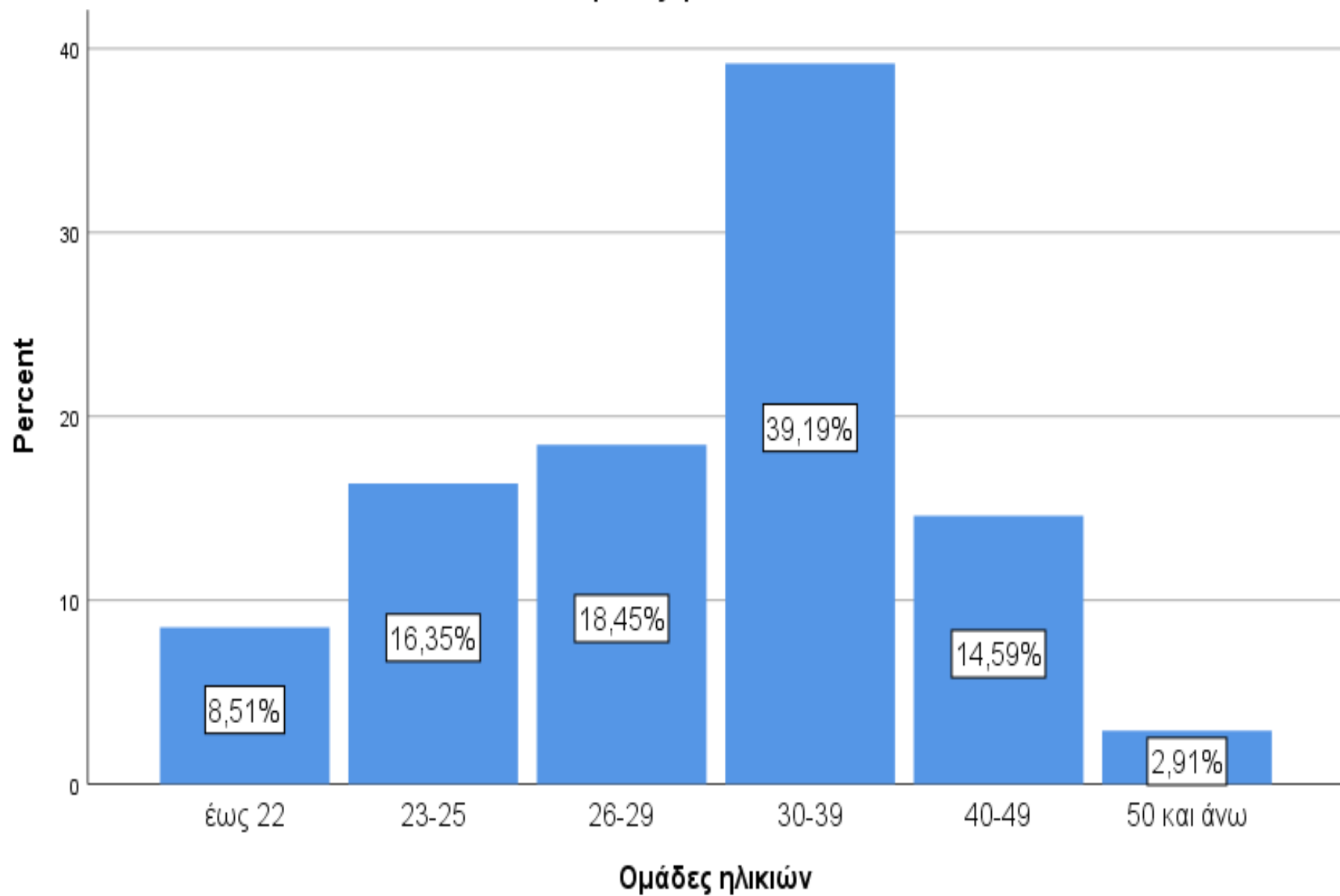
## Μονομεταβλητή Ανάλυση (Univariate analysis)

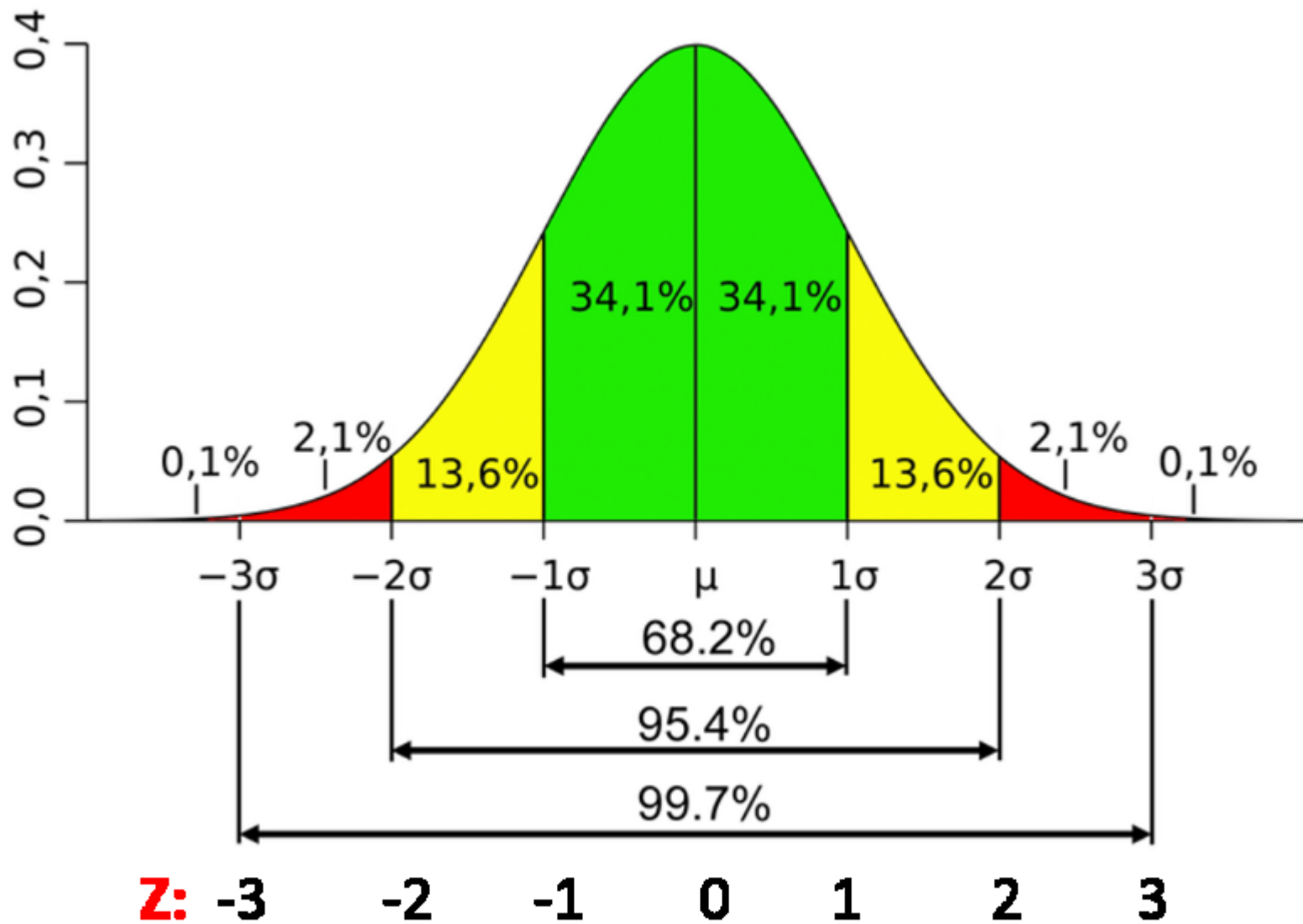
- **Πίνακες Συχνοτήτων** (άνδρες, γυναίκες, μόρφωση, κλάδοι δραστηριοποίησης, κτλ)
- **Μέτρα κεντρικής τάσης:** (Αυτά προσδιορίζουν την θέση των τιμών της μεταβλητής πάνω στον οριζόντιο άξονα) αριθμητικός μέσος όρος (mean), mode (επικρατούσα τιμή), διάμεσος (median).
- **Μέτρα διασποράς:** (Εκφράζουν το πόσο μακριά είναι οι τιμές της  $X$  από κάποιο μέτρο θέσης) διακύμανση (variance), τυπική απόκλιση (standard deviation), κλπ.
- **Μέτρα μορφής μιας κατανομής:** ασυμμετρία (skewness) (δεξιά ή αριστερά), κύρτωση (kurtosis) (πάνω ή κάτω), κλπ.

## Ομάδες ηλικιών

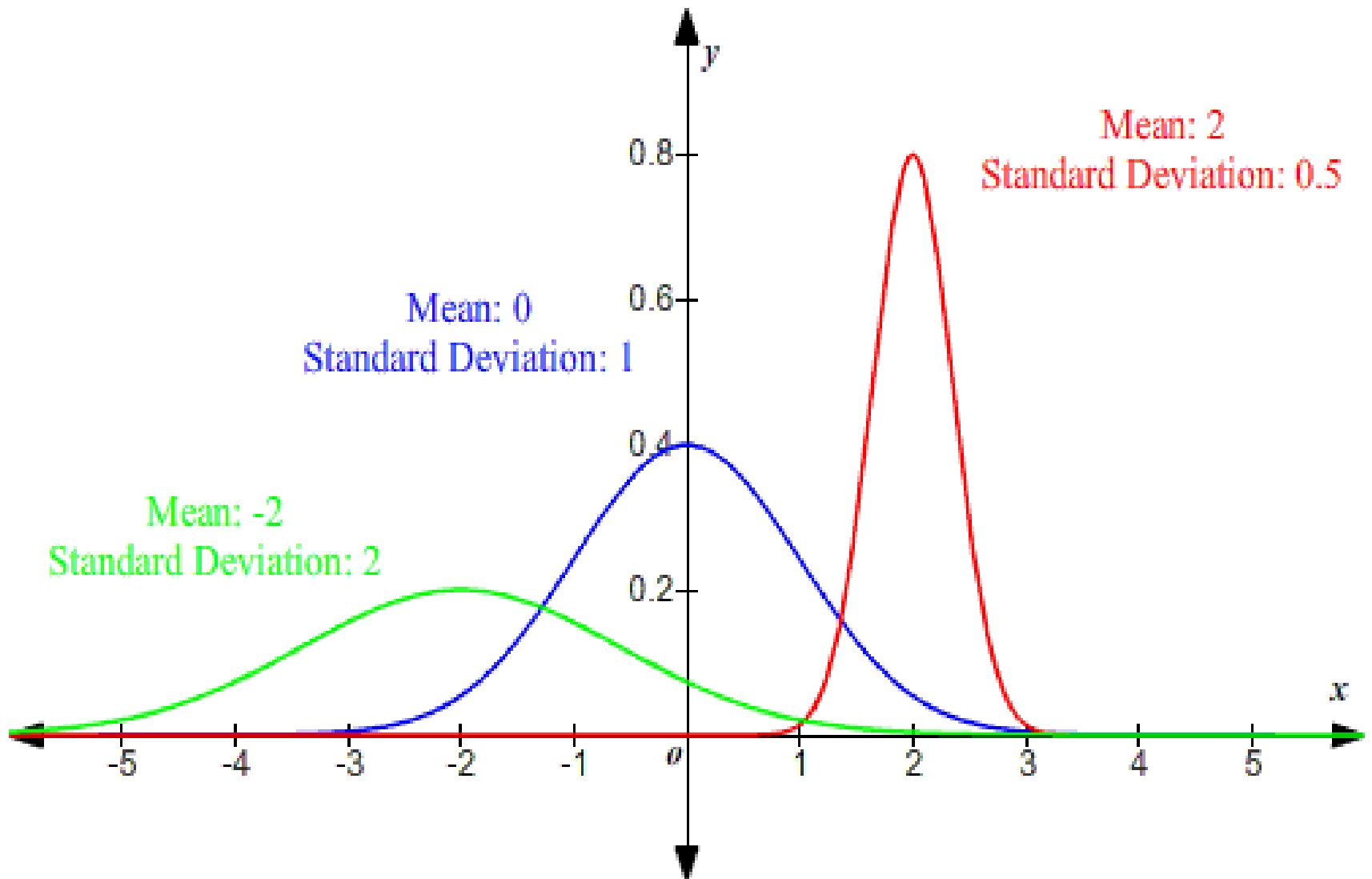
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	έως 22	126	8,5	8,5	8,5
	23-25	242	16,3	16,4	24,9
	26-29	273	18,4	18,4	43,3
	30-39	580	39,0	39,2	82,5
	40-49	216	14,5	14,6	97,1
	50 και άνω	43	2,9	2,9	100,0
	Total	1480	99,5	100,0	
Missing	System	7	,5		
Total		1487	100,0		

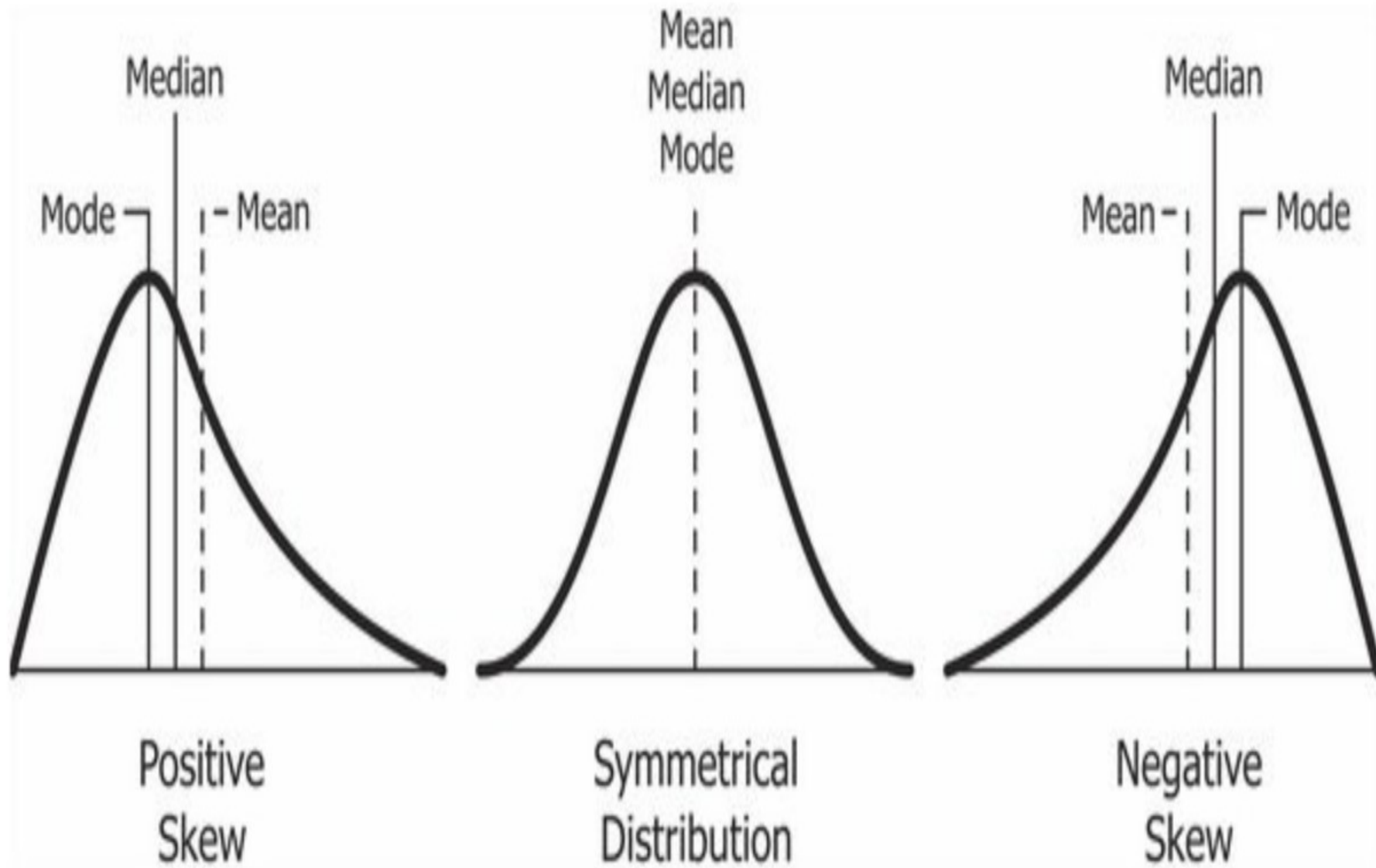
### Ομάδες ηλικιών











# Έλεγχος κανονικότητας

Οι παραμετρικοί έλεγχοι θεωρούν ότι τα δείγματα προέρχονται από κανονικά κατανομημένους πληθυσμούς που σημαίνει ότι οι τιμές δεδομένων για καθεμία από τις ποσοτικές μεταβλητές θα πρέπει επίσης να είναι κανονικά κατανομημένες.

## Έλεγχος με χρήση γραφημάτων

- (ιστογράμματα, διαγράμματα πλαισίου απολήξεων, πολύγωνα συχνοτήτων κλπ)
- για κανονικά κατανομημένα δεδομένα, οι τιμές του αριθμητικού μέσου, της διάμεσου και της επικρατούσας τιμής είναι ίδιες.

## Έλεγχος με χρήση στατιστικών ελέγχων

- έλεγχος Kolmogorov–Smirnov
- έλεγχος Shapiro–Wilks
- Με πολύ μεγάλα δείγματα, είναι εύκολο να πάρετε σημαντικές διαφορές ανάμεσα σε μια μεταβλητή δείγματος και σε μια συγκρίσιμη κανονική κατανομή ακόμα και όταν οι πραγματικές διαφορές είναι σχετικά μικρές. Γι' αυτό το λόγο, είναι χρήσιμο να χρησιμοποιήσετε επίσης ένα διάγραμμα προκειμένου η απόφασή σας να είναι εμπεριστατωμένη.

# Διμεταβλητή Ανάλυση (Bivariate analysis)

- Αναζητά σχέσεις μεταξύ δύο μεταβλητών
- Ελέγχει την συνδιακύμανση (co-variance) και τις συσχετίσεις (correlations)
- **Δεν μπορεί να προσδιορίσει σχέσεις αιτίας-αιτιατού (causality)**
- Contingency tables (Πίνακες Διπλής Εισόδου)
  - Συνδέει τις συχνότητες (frequencies) δύο μεταβλητών
  - Βοηθά στην αναγνώριση προτύπων σχέσης (patterns of association) μεταξύ δύο μεταβλητών

# The chi-square test

## A.1 \* A.3 Crosstabulation

		A.3					
		Λύκειο	ΤΕΙ	ΑΕΙ	Μεταπτυχικά	Total	
A.1	Άνδρας	Count	164	107	213	57	541
		% within A.1	30,3%	19,8%	39,4%	10,5%	100,0%
		% within A.3	57,7%	54,3%	49,2%	57,0%	53,4%
		% of Total	16,2%	10,6%	21,0%	5,6%	53,4%
	Γυναίκα	Count	120	90	220	43	473
		% within A.1	25,4%	19,0%	46,5%	9,1%	100,0%
		% within A.3	42,3%	45,7%	50,8%	43,0%	46,6%
		% of Total	11,8%	8,9%	21,7%	4,2%	46,6%
Total	Count	284	197	433	100	1014	
	% within A.1	28,0%	19,4%	42,7%	9,9%	100,0%	
	% within A.3	100,0%	100,0%	100,0%	100,0%	100,0%	
	% of Total	28,0%	19,4%	42,7%	9,9%	100,0%	

### Chi-Square Tests

	Value	df	Asymptotic Significance (2- sided)
Pearson Chi-Square	5,823 <sup>a</sup>	3	,121
Likelihood Ratio	5,829	3	,120
Linear-by-Linear Association	2,168	1	,141
N of Valid Cases	1014		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 46,65.

### Directional Measures

		Value
Nominal by Interval	Eta	,076
	A.1 Dependent	,076
	A.3 Dependent	,046

### Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Contingency Coefficient	,076	,121
N of Valid Cases		1014	

# Έλεγχοι Στατιστικής Σημαντικότητας

## Το **chi-square** test

Προσδιορίζει το πόσο σίγουροι μπορούμε να είμαστε πως υπάρχει μία σχέση (relationship) μεταξύ δύο μεταβλητών στον πληθυσμό.

## Συσχέτιση και στατιστική σημαντικότητα

Παρέχει πληροφορίες σχετικά με την πιθανότητα να υπάρχει συσχέτιση μεταξύ δύο μεταβλητών στο πληθυσμό

Σύγκριση των μέσων και στατιστική σημαντικότητα – ο δείκτης **F**

# Έλεγχος σημαντικότητας σχέσεων και διαφορών

Ο έλεγχος για την πιθανότητα μια παρατηρούμενη σχέση ή διαφορά ανάμεσα σε μεταβλητές να είναι συμπτωματική και μόνο.

**Μηδενική υπόθεση ( $H_0$ ):** «δεν υπάρχει σημαντική σχέση ή διαφορά ανάμεσα στις μεταβλητές ...»

**Εναλλακτική υπόθεση ( $H_1$ ):** «υπάρχει σημαντική σχέση ή διαφορά ανάμεσα στις μεταβλητές ...»



- η στατιστική σημαντικότητα μιας σχέσης που υποδεικνύεται από ένα στατιστικό έλεγχο καθορίζεται εν μέρει από το μέγεθος του δείγματος
- είναι πολύ δύσκολο να προκύψει στατιστικά σημαντικός έλεγχος από ένα πολύ μικρό δείγμα.
- όσο αυξάνεται το μέγεθος του δείγματος, λιγότερο προφανείς σχέσεις και διαφορές διαπιστώνεται ότι είναι στατιστικά σημαντικές
- μικρά δείγματα μπορούν να μειώσουν στο ελάχιστο την ευαισθησία των στατιστικών ελέγχων, ενώ τα πολύ μεγάλα δείγματα μπορούν να κάνουν το αντίθετο

# Στατιστική Σημαντικότητα

- Πόσο σίγουροι μπορεί να είμαστε πως τα αποτελέσματα που προέκυψαν από την ανάλυση των στοιχείων ενός δείγματος μπορούν να γενικευθούν (ισχύουν για όλο τον πληθυσμό);
- Πόσο επικίνδυνο είναι να κάνουμε αυτή την αναγωγή;
- Ουσιαστικά ισχύει μόνο για πιθανολογικά δείγματα.



# Στατιστική Σημαντικότητα

Κατάσταση στον πληθυσμό		
Απόφαση για τη μηδενική υπόθεση	Η υπόθεση ισχύει	Η υπόθεση δεν ισχύει
Η υπόθεση απορρίπτεται	Σφάλμα τύπου I  <p>Σφάλμα τύπου I εσφαλμένα θετικό</p>	Σωστή απόφαση
Η υπόθεση δεν απορρίπτεται	Σωστή απόφαση	Σφάλμα τύπου II  <p>Σφάλμα τύπου II (εσφαλμένα αρνητικό)</p>



# Ποσοτική Ανάλυση Δεδομένων

## ΑΝΑΛΥΣΗ ΔΙΑΣΠΟΡΑΣ (ANOVA - ANalysis Of VAriance)

Η ανάλυση διακύμανσης χρησιμοποιείται για τον έλεγχο της στατιστικής σημαντικότητας των διαφορών των μέσων όρων περισσότερων από δύο ομάδων-δειγμάτων.

Η ANOVA ελέγχει την υπόθεση ότι όλες οι μέσες τιμές των ομάδων που ελέγχουμε είναι ίσες.

Η ύπαρξη έστω και μιας διαφοράς ερμηνεύεται ότι ο παράγοντας (κατηγορική μεταβλητή) επηρεάζει σημαντικά την ποσοτική μεταβλητή

# Ποσοτική Ανάλυση Δεδομένων

## ΑΝΑΛΥΣΗ ΔΙΑΣΠΟΡΑΣ (ANOVA - ANalysis Of VAriance)

### Descriptives

Αντιλαμβανόμενη αυτεπάρκεια (perceived self efficacy)

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Λύκειο	284	3,3930	1,08687	,06449	3,2660	3,5199	1,00	5,00
TEI	197	3,3635	1,01553	,07235	3,2208	3,5061	1,00	5,00
AEI	433	3,5376	,97918	,04706	3,4452	3,6301	1,00	5,00
Μεταπτυχικά	100	3,7460	,86403	,08640	3,5746	3,9174	1,00	5,00
Total	1014	3,4838	1,01220	,03179	3,4215	3,5462	1,00	5,00

### Test of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
Αντιλαμβανόμενη αυτεπάρκεια (perceived self efficacy)	Based on Mean	5,027	3	1010	,002
	Based on Median	5,352	3	1010	,001
	Based on Median and with adjusted df	5,352	3	1008,254	,001
	Based on trimmed mean	5,355	3	1010	,001

### ANOVA

Αντιλαμβανόμενη αυτεπάρκεια (perceived self efficacy)

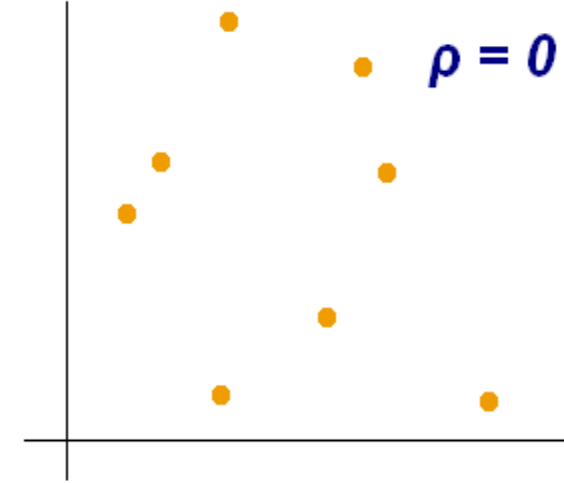
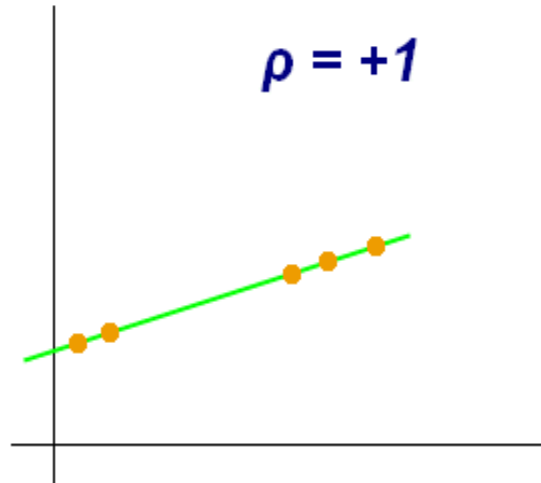
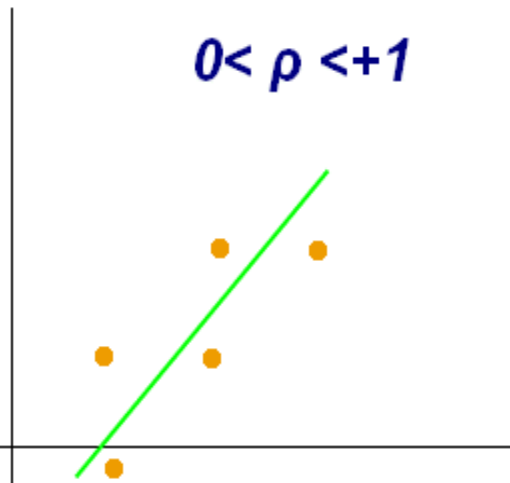
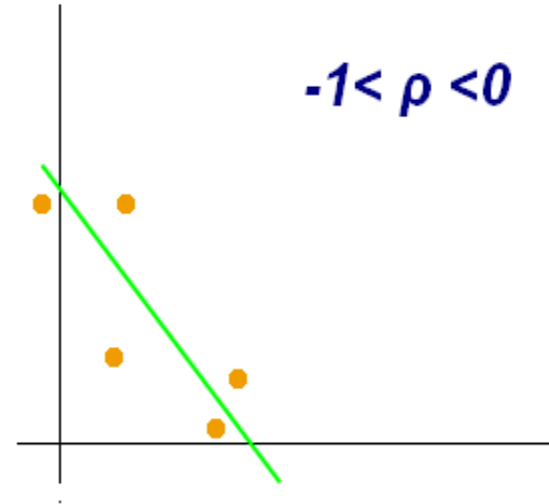
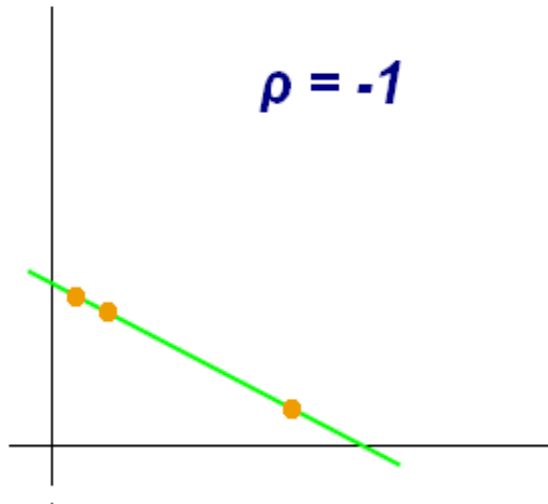
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	13,327	3	4,442	4,379	,005
Within Groups	1024,548	1010	1,014		
Total	1037,875	1013			

# Αξιολόγηση της έντασης της σχέσης μεταξύ δύο μεταβλητών

## Συντελεστής συσχέτισης:

- δίνει τη δυνατότητα να ποσοτικοποίησης της έντασης της γραμμικής σχέσης ανάμεσα σε δύο μεταβλητές κατάταξης ή αριθμητικές.
- συμβολίζεται με το γράμμα  **$r$**
- μπορεί να πάρει οποιαδήποτε τιμή ανάμεσα στο +1 που αναπαριστά μια πλήρως θετική συσχέτιση και -1 που αναπαριστά μια πλήρως αρνητική συσχέτιση
- συντελεστές συσχέτισης ανάμεσα στο +1 και το -1 αναπαριστούν ασθενέστερες θετικές και αρνητικές συσχετίσεις, ενώ η τιμή 0 σημαίνει ότι οι μεταβλητές είναι απόλυτα ανεξάρτητες
- στην έρευνα στους τομείς των επιχειρήσεων και της διοίκησης, οι τέλει ή μηδενικές συσχετίσεις είναι εξαιρετικά ασυνήθιστες.

# Αξιολόγηση της έντασης της σχέσης μεταξύ δύο μεταβλητών



# Αξιολόγηση της έντασης της σχέσης μεταξύ δύο μεταβλητών

## Συντελεστής γραμμικής συσχέτισης του Pearson (PMCC)

- και οι δύο μεταβλητές είναι αριθμητικές
- τα δείγματα είναι τυχαία

## Αν η μία ή και οι δύο μεταβλητές περιέχουν δεδομένα κατάταξης

- συντελεστής συσχέτισης κατάταξης του Spearman (rho [ $\rho$ ] του Spearman)
- συντελεστής συσχέτισης κατάταξης του Kendall (tau [ $\tau$ ] του Kendall)
- ο συντελεστής του Kendall θεωρείται πιο κατάλληλος αν τα δεδομένα μιας εκ των μεταβλητών περιέχουν συνδεδεμένες βαθμίδες (tied ranks).



# Ποσοτική Ανάλυση Δεδομένων

## ΣΥΣΧΕΤΙΣΗ

Για να καταλήξουμε στην υποστήριξη ή απόρριψη μιας υπόθεσης εξετάζουμε: (α) το πρόσημο του **συντελεστή συσχέτισης**, και (β) το **επίπεδο εμπιστοσύνης** ( $p = \text{significance level}$ ).

Αν το επίπεδο εμπιστοσύνης είναι μικρότερο του επιπέδου εμπιστοσύνης που ορίσαμε (συνήθως 0,01 ή 0,05) και το πρόσημο του συντελεστή συμφωνεί με την υπόθεση (θετικό για θετική σχέση *και αρνητικό για αρνητική σχέση*), τότε καταλήγουμε στο ότι η υπόθεση υποστηρίζεται από τα δεδομένα. Σε αντίθετη περίπτωση το συμπέρασμα είναι ότι η υπόθεση δεν υποστηρίζεται από τα δεδομένα.

Κωδικοποίηση  
Δεδομένων

Έλεγχος  
εγκυρότητας  
της δομής

Βασικά  
στατιστικά  
μέτρα

Έλεγχος  
υποθέσεων

**Πίνακας 2:** Συσχετίσεις ανάμεσα στις μεταβλητές του μοντέλου της έρευνας μας

Αριθμός Μεταβλητής		1	2	3	4	5	6	7
1	Pearson Correlation	1,00						
2	Pearson Correlation	<b>-0,23*</b>	1,00					
3	Pearson Correlation	<b>0,29**</b>		1,00				
4	Pearson Correlation	<b>0,48**</b>	-0,16		1,00			
5	Pearson Correlation	<b>-0,35**</b>	-0,19			1,00		
6	Pearson Correlation	<b>0,36*</b>		0,13		<b>0,41*</b>	1,00	
7	Pearson Correlation	<b>0,44**</b>		<b>0,34**</b>	<b>0,33*</b>			1,00

\* Correlation is significant at the 0,05 level (2-tailed)

\*\* Correlation is significant at the 0,01 level (2-tailed)



# Ανάλυση της σχέσης μεταξύ μεταβλητών διαφορετικού τύπου

**Spearman's *rho***: για την εξέταση της σχέσης μεταξύ ιεραρχικών (ordinal) ή μίας ιεραρχικής και μίας αριθμητικής μεταβλητής

**Phi coefficient**: για την εξέταση της σχέσης μεταξύ δύο διχοτομικών μεταβλητών (*dichotomous variables*) (τιμές από -1 to +1)

**Cramer's *V***: για την εξέταση της σχέσης μεταξύ δύο ονομαστικών (*nominal*), ή μίας ονομαστικής και μίας ιεραρχικής (*ordinal*) μεταβλητής (τιμές μεταξύ 0 και 1)

**Σύγκριση μέσων**: όταν μία ονομαστική μεταβλητή ορίζεται ως η ανεξάρτητη μεταβλητή, οι μέσοι των αριθμητικών / κλίμακας μεταβλητών (interval/ratio) συγκρίνονται για κάθε υποκατηγορία της ονομαστικής μεταβλητής.

***eta***: για την εξέταση σχέσεων (association) μεταξύ διαφορετικού τύπου μεταβλητών, ακόμη και όταν δεν υπάρχει γραμμική σχέση μεταξύ τους.

# Πολυμεταβλητή Ανάλυση (Multivariate analysis)

- Η σχέση (relationship) μεταξύ δύο μεταβλητών να είναι τυχαία (*spurious*)
  - Κάθε μεταβλητή μπορεί να συνδέεται με την ίδια Τρίτη μεταβλητή.
- Μπορεί να υπάρχει μία μεταβλητή που διαμεσολαβεί μεταξύ των δύο μεταβλητών (*intervening variable*)
- Μία Τρίτη μεταβλητή μπορεί να επιδρά πάνω στη σχέση (*moderating variable*)
  - Π.χ. Το Φύλο μπορεί να επιδρά πάνω στη συσχέτιση μεταξύ της Ηλικίας και της Εξάσκησης



# Ποσοτική Ανάλυση Δεδομένων

## ΠΑΡΑΓΟΝΤΙΚΗ ΑΝΑΛΥΣΗ

Η Παραγοντική Ανάλυση (Factor Analysis) έχει σκοπό να βρει την **ύπαρξη κοινών παραγόντων ανάμεσα σε μια ομάδα μεταβλητών**. Έτσι, εκφράζοντας αυτούς τους παράγοντες (οι οποίοι δεν είναι μία υπαρκτή ποσότητα αλλά την κατασκευάζουμε για τις ανάγκες μας) μπορούμε:

**Να μειώσουμε τις διαστάσεις του προβλήματος.** Αντί να δουλεύουμε με τις αρχικές μεταβλητές να δουλέψουμε με λιγότερες, αφού οι παράγοντες είναι έτσι κατασκευασμένοι ώστε να διατηρούν όσο γίνεται την πληροφορία που υπήρχε στις αρχικές μεταβλητές.

**Να δημιουργήσουμε νέες μεταβλητές,** τους παράγοντες, στις οποίες μπορούμε με έναν υποκειμενικό τρόπο να αναγνωρίσουμε ως κάποιες μη μετρήσιμες μεταβλητές, όπως π.χ. η ευφυΐα στην ψυχολογία, ή η ελκυστικότητα ενός προϊόντος.



# Ποσοτική Ανάλυση Δεδομένων

## ΠΑΡΑΓΟΝΤΙΚΗ ΑΝΑΛΥΣΗ

Η παραγοντική ανάλυση μπορεί να είναι **διερευνητική (exploratory)** δηλαδή να μας βοηθάει να ανακαλύψουμε και να ταυτοποιήσουμε μη παρατηρούμενους παράγοντες, **ή επιβεβαιωτική (confirmatory)** όπου ελέγχουμε αν ένα σύνολο μεταβλητών που χρησιμοποιούμε για να μετρήσουμε μη παρατηρούμενους παράγοντες είναι ικανοποιητικό.

Οι πλέον διαδεδομένες μέθοδοι για την εξαγωγή παραγόντων είναι η ανάλυση σε **κύριες συνιστώσες** και η μέθοδος **μέγιστης πιθανοφάνειας**.



# Ποσοτική Ανάλυση Δεδομένων

## ΠΑΡΑΓΟΝΤΙΚΗ ΑΝΑΛΥΣΗ

Δείκτης **Kaiser-Meyer-Olkin (ΚΜΟ)**, που συγκρίνει τα μεγέθη των παρατηρούμενων συντελεστών συσχέτισης προς τους συντελεστές μερικής συσχέτισης. Μικρές τιμές του δείκτη δηλώνουν ότι η παραγοντική ανάλυση δεν είναι **κατάλληλη τεχνική για τα δεδομένα**.

Οι συντελεστές συσχέτισης μεταξύ των μεταβλητών θα πρέπει να είναι υψηλοί. Εάν οι συσχετίσεις είναι χαμηλές είναι σχεδόν αδύνατο οι μεταβλητές να μοιράζονται κοινούς παράγοντες. Για τον σκοπό αυτό χρησιμοποιείται ο **έλεγχος σφαιρικότητας του Bartlett**.

**Φορτίσεις Μεταβλητών (Factor Loadings):** κατά κανόνα εκφράζουν το βαθμό συσχέτισης των παραγόντων με τις αρχικές μεταβλητές.

# Ποσοτική Ανάλυση Δεδομένων

## ΠΑΡΑΓΟΝΤΙΚΗ ΑΝΑΛΥΣΗ

### Correlation Matrix

		B3.1	B3.2	B3.3	B3.4	B3.5	B3.6
Correlation	B3.1	1,000	,815	,730	,622	,055	,079
	B3.2	,815	1,000	,772	,640	,054	,089
	B3.3	,730	,772	1,000	,668	,083	,103
	B3.4	,622	,640	,668	1,000	,107	,116
	B3.5	,055	,054	,083	,107	1,000	,642
	B3.6	,079	,089	,103	,116	,642	1,000

### KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,772
Bartlett's Test of Sphericity	Approx. Chi-Square	3340,316
	df	15
	Sig.	,000

### Anti-image Matrices

		B3.1	B3.2	B3.3	B3.4	B3.5	B3.6
Anti-image Covariance	B3.1	,305	-,155	-,068	-,056	,001	,004
	B3.2	-,155	,264	-,112	-,051	,015	-,012
	B3.3	-,068	-,112	,337	-,128	-,009	-,004
	B3.4	-,056	-,051	-,128	,501	-,026	-,009
	B3.5	,001	,015	-,009	-,026	,585	-,373
	B3.6	,004	-,012	-,004	-,009	-,373	,584
Anti-image Correlation	B3.1	,814 <sup>a</sup>	-,546	-,213	-,143	,001	,010
	B3.2	-,546	,784 <sup>a</sup>	-,376	-,141	,038	-,029
	B3.3	-,213	-,376	,849 <sup>a</sup>	-,311	-,020	-,010
	B3.4	-,143	-,141	-,311	,901 <sup>a</sup>	-,048	-,018
	B3.5	,001	,038	-,020	-,048	,515 <sup>a</sup>	-,638
	B3.6	,010	-,029	-,010	-,018	-,638	,524 <sup>a</sup>

a. Measures of Sampling Adequacy(MSA)

# Ποσοτική Ανάλυση Δεδομένων

## ΠΑΡΑΓΟΝΤΙΚΗ ΑΝΑΛΥΣΗ

Component	Total Variance Explained								
	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,166	52,769	52,769	3,166	52,769	52,769	3,125	52,079	52,079
2	1,607	26,790	79,559	1,607	26,790	79,559	1,649	27,480	79,559
3	,426	7,098	86,657						
4	,356	5,934	92,591						
5	,266	4,436	97,027						
6	,178	2,973	100,000						

Extraction Method: Principal Component Analysis.

### Component Matrix<sup>a</sup>

	Component	
	1	2
B3.1	,890	-,129
B3.2	,909	-,126
B3.3	,893	-,092
B3.4	,820	-,042
B3.5	,182	,888
B3.6	,210	,881

Extraction Method: Principal Component Analysis.

a. 2 components extracted.

### Rotated Component Matrix<sup>a</sup>

	Component	
	1	2
B3.1	,899	,018
B3.2	,917	,024
B3.3	,896	,055
B3.4	,816	,092
B3.5	,034	,906
B3.6	,064	,903

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

### Component Transformation Matrix

Component	1	2
1	,987	,163
2	-,163	,987

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.



# Ποσοτική Ανάλυση Δεδομένων

## ΠΑΛΙΝΔΡΟΜΙΣΗ

Η **ανάλυση παλινδρόμησης** (regression) ορίζεται ως η εκτίμηση μιας εξαρτημένης μεταβλητής από μια ανεξάρτητη μεταβλητή (**απλή παλινδρόμηση**) ή από περισσότερες ανεξάρτητες μεταβλητές (**πολλαπλή παλινδρόμηση**).

Στην πράξη **ανεξάρτητη μεταβλητή X** είναι εκείνη την οποία μπορούμε να ελέγξουμε, δηλαδή, να καθορίσουμε τις τιμές της (π.χ. το ύψος της διαφημιστικής δαπάνης ενός προϊόντος, ο αριθμός των λειτουργούντων ταμείων σε ένα υποκατάστημα τραπεζής).

**Εξαρτημένη μεταβλητή Y** είναι εκείνη στην οποία αντανακλάται το αποτέλεσμα των μεταβολών στις ανεξάρτητες μεταβλητές (π.χ. η ζήτηση ενός προϊόντος, ο χρόνος αναμονής των πελατών ενός υποκαταστήματος τραπεζής).





# Ποσοτική Ανάλυση Δεδομένων

## ΠΑΛΙΝΔΡΟΜΙΣΗ

Η παλινδρόμηση ποσοτικοποιεί την εξάρτηση, όταν αυτή υπάρχει.

- Είναι σημαντικό η παλινδρόμηση να εφαρμόζεται αφού έχουμε διαπιστώσει ότι υπάρχει συσχέτιση μεταξύ των μεταβλητών, αλλιώς τα αποτελέσματα της ανάλυσης θα είναι λάθος.
- η διαδικασία υπολογισμού του συντελεστή προσδιορισμού και της εξίσωσης παλινδρόμησης με μία ανεξάρτητη μεταβλητή ορίζεται ως ανάλυση παλινδρόμησης.
- ο υπολογισμός συντελεστή πολλαπλού προσδιορισμού (ή συντελεστής πολλαπλής παλινδρόμησης  $R^2$ ) και η εξίσωση παλινδρόμησης με δύο ή περισσότερες ανεξάρτητες μεταβλητές ορίζεται ως ανάλυση πολλαπλής παλινδρόμησης.

# Αξιολόγηση της έντασης σχέσης αιτιώδους συνάφειας

## Συντελεστής προσδιορισμού ή συντελεστής παλινδρόμησης ( $r^2$ )

- αξιολογεί την ένταση της σχέσης ανάμεσα σε μια αριθμητική εξαρτημένη μεταβλητή και μία ή περισσότερες αριθμητικές ανεξάρτητες μεταβλητές
- τα δεδομένα πρέπει να έχουν επιλεγεί τυχαία
- μπορεί να πάρει οποιαδήποτε τιμή ανάμεσα στο **0** και το **+1**
- μετρά το ποσοστό διακύμανσης μιας εξαρτημένης μεταβλητής που μπορεί να εξηγηθεί στατιστικά από τη διακύμανση μίας ανεξάρτητης μεταβλητής ή πολλών ανεξάρτητων μεταβλητών
  - αν η μεταβολή μπορεί να εξηγηθεί, στο σύνολό της, ο συντελεστής προσδιορισμού θα είναι 1.
  - αν κανένα ποσοστό της μεταβολής δεν μπορεί να εξηγηθεί, ο συντελεστής θα είναι 0
  - στις έρευνές μας, σπάνια παίρνουμε τιμές συντελεστή πάνω από 0,8

$$Y = b_0 + b_1 * X_1 + b_2 * X_2 + b_3 * X_3 + b_4 * X_4 + b_5 * X_5, \text{ όπου:}$$

$Y$  = η εξαρτημένη μεταβλητή «αντιλαμβανόμενη εμπιστοσύνη»,

$X_1$  = η ανεξάρτητη μεταβλητή «αντιλαμβανόμενος κίνδυνος»,

$X_2$  = η ανεξάρτητη μεταβλητή «προσανατολισμός στην αγορά»,

$X_3$  = η ανεξάρτητη μεταβλητή «αξιοπιστία της τεχνολογίας»,

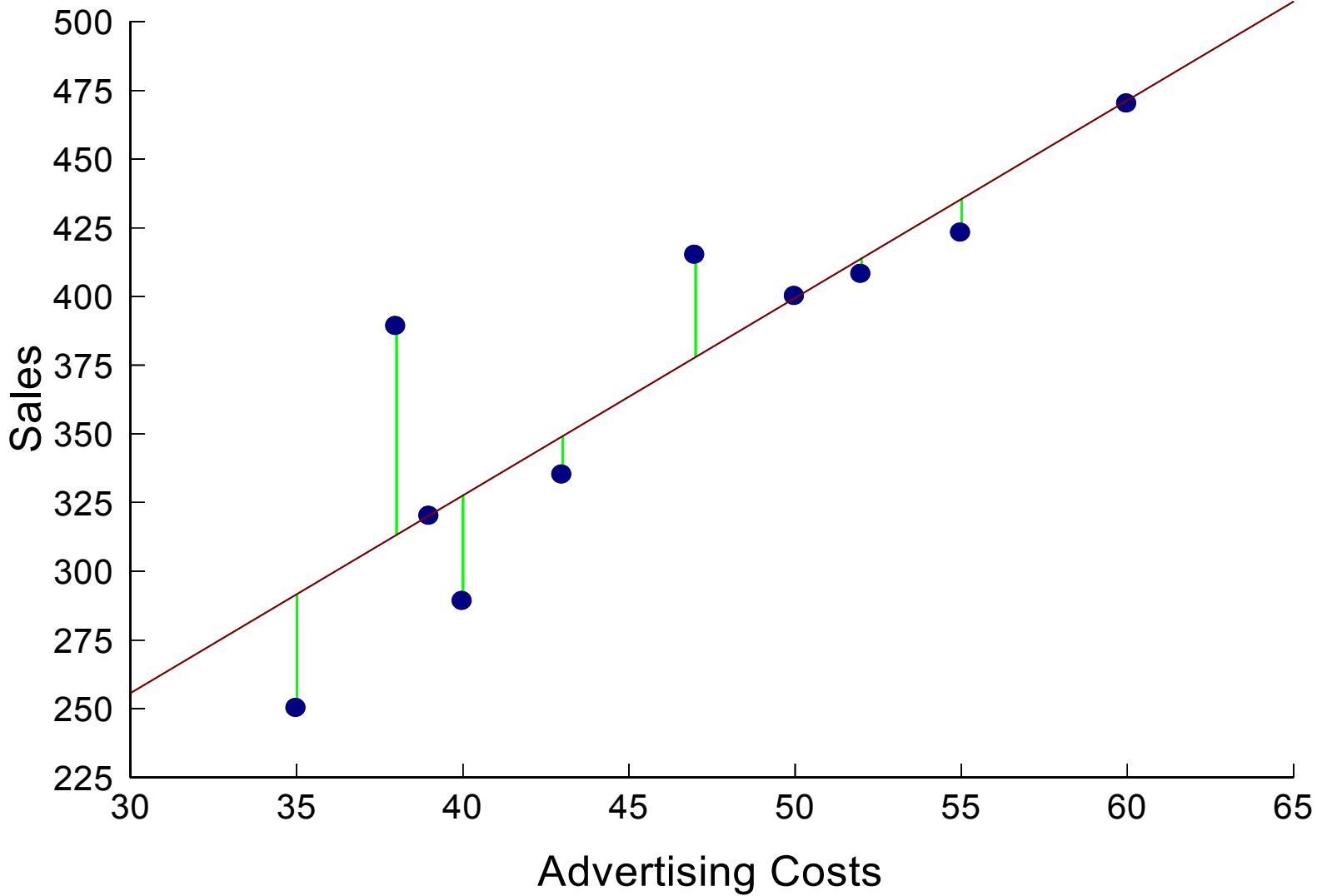
$X_4$  = η ανεξάρτητη μεταβλητή «εμπειρία των χρηστών του Web»,

$X_5$  = η ανεξάρτητη μεταβλητή «ποιότητα του δικτυακού τόπου»,

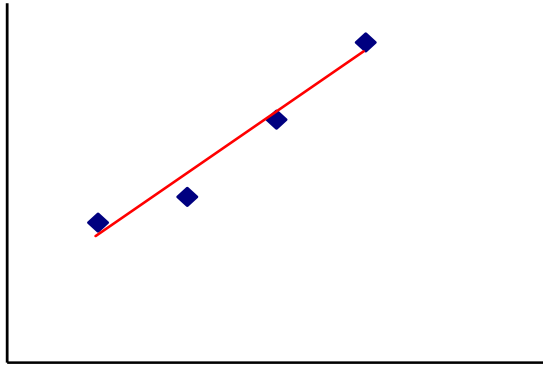
$b_0$  = ο σταθερός όρος,

$b_1, b_2, b_3, b_4, b_5$  = οι συντελεστές μερικής παλινδρόμησης.

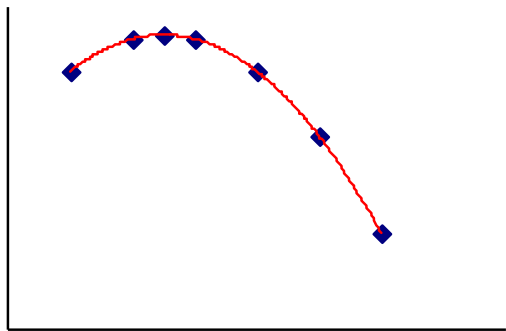
# Γραμμή παλινδρόμησης και κατάλοιπα:



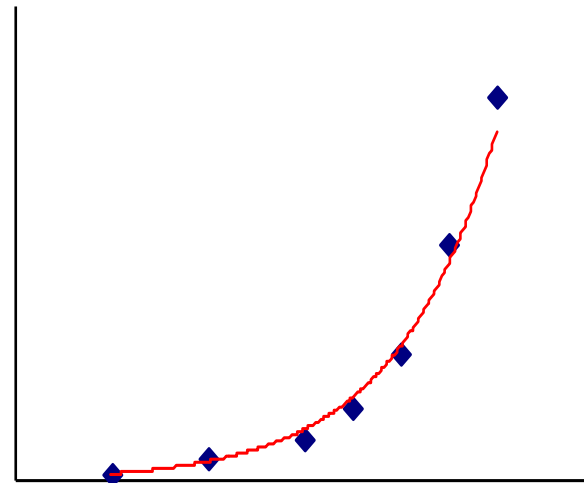
# Τύποι Σχέσης



Γραμμική



Καμπυλόγραμμη



Εκθετική

### Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
Αντιλαμβανόμενη εμπιστοσύνη	,513 <sup>a</sup>	,497	,471	,71

a. Predictors: (Constant), αντιλαμβανόμενος κίνδυνος, αντιλαμβανόμενος προσανατολισμός στην αγορά, αντιλαμβανόμενη αξιοπιστία της τεχνολογίας, εμπειρία των χρηστών του Web, αντιλαμβανόμενη ποιότητα του δικτυακού τύπου

### ANOVA<sup>b</sup>

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	109,899	4	21,980	6,346	,000 <sup>a</sup>
Residual	63,334	127	,499		
Total	173,233	132			

a. Predictors: (Constant), αντιλαμβανόμενος κίνδυνος, αντιλαμβανόμενος προσανατολισμός στην αγορά, αντιλαμβανόμενη αξιοπιστία της τεχνολογίας, εμπειρία των χρηστών του Web, αντιλαμβανόμενη ποιότητα το δικτυακού τύπου

b. Dependent Variable: Αντιλαμβανόμενη εμπιστοσύνη

### Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
	(Constant)	,236	,171		,381	,170
	Αντιλαμβανόμενος κίνδυνος	,350	,089	,014	,171	,036
	Αντιλαμβανόμενος προσανατολισμός στην αγορά	,110	,071	,303	1,077	,051
	Αντιλαμβανόμενη αξιοπιστία της τεχνολογίας	,369	,111	,545	,148	,000
	Εμπειρία των χρηστών του Web	,140	,060	,082	1,268	,207
	Αντιλαμβανόμενη ποιότητα του δικτυακού τύπου	,324	,086	,041	,481	,030

a. Dependent Variable: Αντιλαμβανόμενη εμπιστοσύνη



# Ποσοτική Ανάλυση Δεδομένων

## ΠΑΛΙΝΔΡΟΜΙΣΗ

Από τα αποτελέσματα της παλινδρόμησης παρατηρούμε ότι μοντέλο της παλινδρόμησης είναι στατιστικά σημαντικό ( $df = 4, F = 6.364, P < 0,05$ ) και μπορεί να εξηγήσει το 47% των διακυμάνσεων της μεταβλητής «αντιλαμβανόμενη εμπιστοσύνη» ( $R = 0,47$ ). Με άλλα λόγια, το αποτέλεσμα αυτό σημαίνει ότι η εμπιστοσύνη που έχουν οι χρήστες στο ηλεκτρονικό εμπόριο εξαρτάται κατά 47% από τα  $X_1, X_2, X_3, X_4$  και  $X_5$ . Το υπόλοιπο 53% των μεταβολών της εμπιστοσύνης οφείλεται σε ανερμήνευτους παράγοντες.



# Ποσοτική Ανάλυση Δεδομένων

## ΠΑΛΙΝΔΡΟΜΙΣΗ

Ανάμεσα στις πέντε ανεξάρτητες μεταβλητές του μοντέλου, η «αντιλαμβανόμενη αξιοπιστία της τεχνολογίας», η «αντιλαμβανόμενη ποιότητα του δικτυακού τόπου» και ο «αντιλαμβανόμενος κίνδυνος», έχουν την ισχυρότερη ερμηνευτική δύναμη. Συγκεκριμένα, ο συντελεστής μερικής παλινδρόμησης της «αντιλαμβανόμενης αξιοπιστίας της τεχνολογίας» είναι 0,369 ( $P < 0,01$ ), της «αντιλαμβανόμενης ποιότητας του δικτυακού τόπου» 0,324 ( $P < 0,05$ ) και του «αντιλαμβανόμενου κινδύνου» 0,35 ( $P < 0,05$ ).



# Προϋποθέσεις εφαρμογής ανάλυσης παλινδρόμησης

- η σχέση ανάμεσα στις εξαρτημένες και ανεξάρτητες μεταβλητές είναι **γραμμική**.
- οι τιμές δεδομένων των εξαρτημένων και ανεξάρτητων μεταβλητών έχουν ίσες διακυμανσεις – **ομοσκεδαστικότητα** (έλεγχος Levene για την ομογένεια της διακύμανσης)
- εάν υπάρχει **ετεροσκεδαστικότητα** (δηλαδή άνισες διακυμανσεις), πιθανόν να μπορείτε να εκτελέσετε την ανάλυσή σας κάτω από προϋποθέσεις
- απουσία συσχέτισης ανάμεσα σε δύο ή περισσότερες ανεξάρτητες μεταβλητές (**συγγραμμικότητα ή πολυσυγγραμμικότητα**), καθώς έτσι δυσχεραίνεται ο προσδιορισμός των επιμέρους επιδράσεων των μεμονωμένων μεταβλητών (παράγοντας διογκώσης διακύμανσης - VIF)
- τα δεδομένα για τις ανεξάρτητες μεταβλητές και την εξαρτημένη μεταβλητή **κατανέμονται κανονικά**

	Είδη δεδομένων		
Στόχοι δημοσκόπησης	Ανεξάρτητη μεταβλητή	Εξαρτημένη μεταβλητή	Πιθανός στατιστικός έλεγχος
Για στόχους με μία εξαρτημένη και μία ανεξάρτητη μεταβλητή			
Σύγκριση τμημάτων για τη συχνότητα έγγραφων προειδοποιήσεων	Ονομαστική: ομάδες (τμήματα)	Ονομαστική: (αριθμός έγγραφων προειδοποιήσεων)	Χ-τετράγωνο, <b>ακριβής έλεγχος Fisher</b>
Σύγκριση μιας πειραματικής ομάδας και μιας ομάδας ελέγχου ως προς τις στάσεις τους μετά από μία αντικαπνιστική εκστρατεία	Ονομαστική (διχοτομική): ομάδες (πειραματική και ελέγχου)	Μετρήσιμη (βαθμολογίες στάσεων)	<b>t-test ενός δείγματος</b> , εξαρτημένο t-test και ανεξάρτητο t-test, Wilcoxon signed-ranks test, Wilcoxon rank-sum test
Σύγκριση στάσεων ως προς τις νέες εργασιακές πρακτικές για πέντε τμήματα της εταιρείας	Ονομαστική: περισσότερες από δύο τιμές	Μετρήσιμη (βαθμολογίες στάσεων)	<b>Ανάλυση Διασποράς κατά μία κατεύθυνση</b> (χρησιμοποιώντας F-test)
Να καθορίσουμε αν οι υψηλές βαθμολογίες μέτρησης της αυτοπεποίθησης προβλέπουν υψηλές βαθμολογίες στα τεστ ικανότητας	Μετρήσιμη (βαθμολογίες στάσεων)	Μετρήσιμη (βαθμολογίες γνώσεων)	Παλινδρόμηση (όταν καμία μεταβλητή δεν είναι εξαρτημένη ή ανεξάρτητη, χρησιμοποιείστε συσχέτιση)

	Είδη δεδομένων		
Στόχοι δημοσκόπησης	Ανεξάρτητη μεταβλητή	Εξαρτημένη μεταβλητή	Πιθανός στατιστικός έλεγχος
Για στόχους με δύο ή περισσότερες ανεξάρτητες μεταβλητές			
Σύγκριση εργατών και υπαλλήλων γραφείου σε πειραματικές ομάδες και ομάδες ελέγχου ως προς τις στάσεις τους	Ονομαστικές (εργάτες και υπάλληλοι γραφείου)	Μετρήσιμες (βαθμολογίες στάσεων)	Ανάλυση διακύμανσης (ANOVA)
Να καθορίσουμε αν η διάρκεια υπηρεσίας και το επίπεδο μισθού σχετίζονται με στάσεις	Μετρήσιμες (διάρκεια υπηρεσίας και επίπεδο μισθού)	Μετρήσιμες (βαθμολογίες στάσεων)	Πολλαπλή παλινδρόμηση
Σύγκριση ανδρών και γυναικών σε πειραματικές ομάδες και ομάδες ελέγχου για τις στάσεις τους όταν ελέγχουμε το επίπεδο μισθού	Ονομαστικές (φύλο και ομάδα) με συγχυτικούς παράγοντες (επίπεδο μισθού)	Μετρήσιμες (βαθμολογίες στάσεων)	Ανάλυση συνδιακύμανσης (ANCOVA)
Για στόχους με δύο ή περισσότερες ανεξάρτητες και εξαρτημένες μεταβλητές			
Σύγκριση ανδρών και γυναικών σε πειραματικές ομάδες και ομάδες ελέγχου για τις στάσεις τους και για τις βαθμολογίες σε τεστ γνώσης	Ονομαστικές (φύλο και ομάδα)	Μετρήσιμες (βαθμολογίες σε δύο μέτρα: στάσεις και γνώσεις)	Πολυμεταβλητή ανάλυση διακύμανσης (MANOVA)