



ΟΙΚΟΝΟΜΙΚΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΘΗΝΩΝ



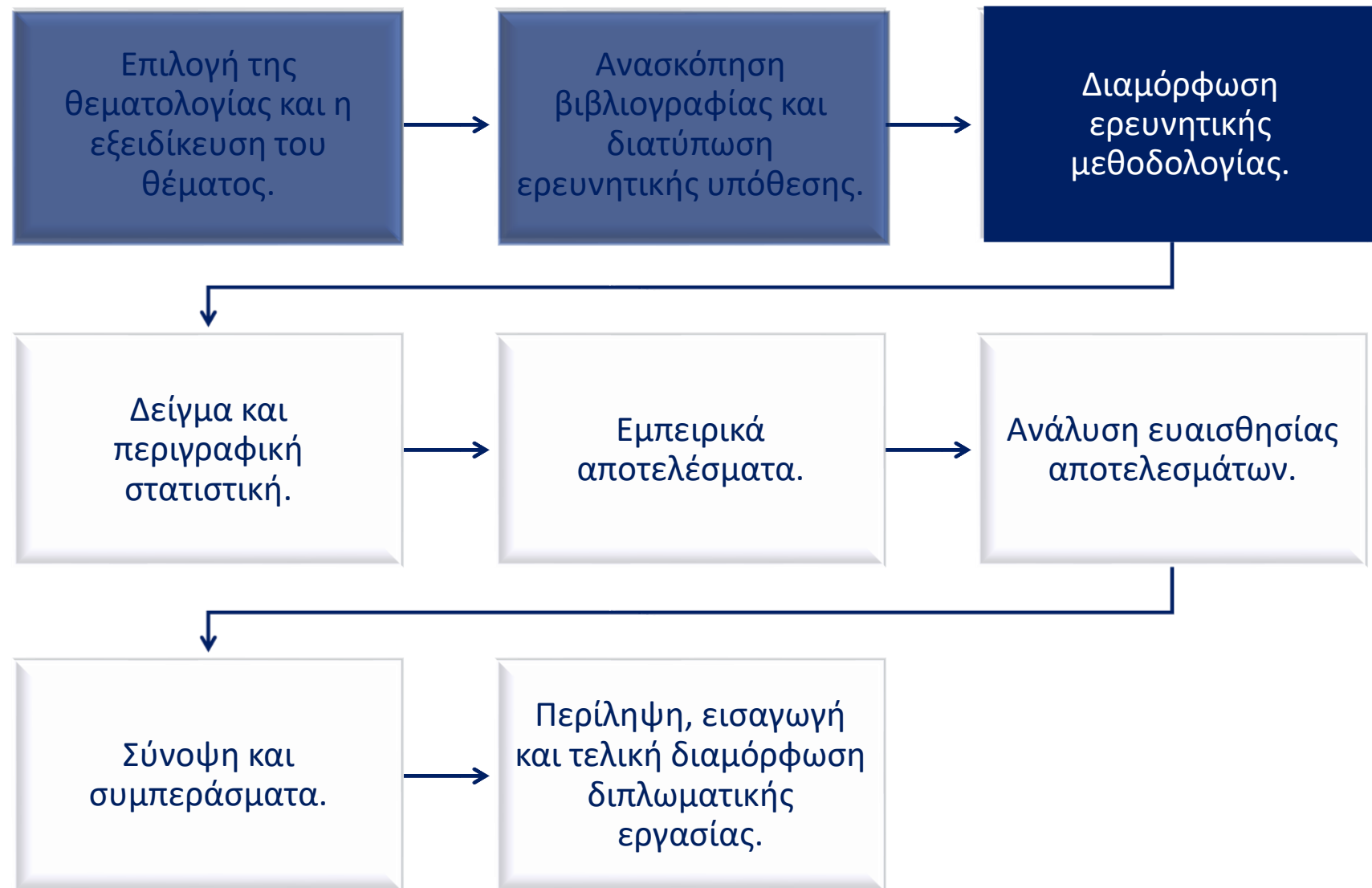
ATHENS UNIVERSITY
OF ECONOMICS
AND BUSINESS

ΣΧΟΛΗ
ΔΙΟΙΚΗΣΗΣ
ΕΠΙΧΕΙΡΗΣΕΩΝ
SCHOOL OF
BUSINESS

ΜΕΤΑΠΤΥΧΙΑΚΟ ΛΟΓΙΣΤΙΚΗΣ &
ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΗΣ
MSc IN ACCOUNTING & FINANCE

Σεμινάρια Εκπόνησης Διπλωματικής Εργασίας: Μέρος Β

Διαμόρφωση Ερευνητικής Μεθοδολογίας



Διαμόρφωση Ερευνητικής Μεθοδολογίας

- Η περιγραφή της θεωρίας αποτελεί το ένα θεμέλιο της επιστημονικής σκέψης.
 - Σε ένα ερευνητικό δοκίμιο ή σε μία διπλωματική εργασία η περιγραφή της θεωρίας εντοπίζεται στη βιβλιογραφική επισκόπηση και στη διατύπωση των ερευνητικών υποθέσεων.
- Τα άλλα δύο θεμέλια είναι (α) η εμπειρική απόδειξη της θεωρίας και (β) η επιβεβαίωση των προβλέψεων που στηρίζονται στη θεωρία.
- Η ερευνητική μεθοδολογία αποτελεί το σύνολο των συστηματικών εργασιών που εστιάζουν στην εμπειρική απόδειξη της θεωρίας.
- Στις κοινωνικές επιστήμες το πρώτο βήμα της ερευνητικής μεθοδολογίας είναι διαμόρφωση μίας μαθηματική έκφρασης της θεωρίας.
- **Γιατί;**
 - Η θεωρία αποτελεί μια περιγραφή οντοτήτων (εννοιών και μεταβλητών) και των σχέσεων τους.
 - Το κύριο ερευνητικό εργαλείο στις κοινωνικές επιστήμες (Λογιστική) είναι οι στατιστικές τεχνικές που εδράζονται στη μαθηματική μοντελοποίηση και την αξιοποίηση ποσοτικών (αλλά και ποιοτικών) δεδομένων) για την εξαγωγή στατιστικών συμπερασμάτων.

Διαμόρφωση Ερευνητικής Μεθοδολογίας: Από τη Θεωρία στις Αλγεβρικές Εκφράσεις

○ Πως μπορούμε να εξάγουμε την αλγεβρική έκφραση μίας θεωρίας;


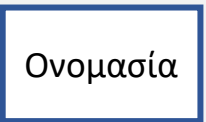




1. Αρχικά εντοπίζονται οι μεταβλητές και οι σχέσεις που αναπτύσσουν μεταξύ τους.

- Στο χώρο των κοινωνικών επιστημών οι περισσότερες μεταβλητές είναι άδηλες (latent) δηλαδή δεν μπορούν να παρατηρηθούν άμεσα και άρα να μετρηθούν με έναν κοινά αποδεκτό και αξιόπιστο τρόπο.
- Μία άδηλη μεταβλητή χαρακτηρίζεται από μια συγκεκριμένη και προβλέψιμη συμπεριφορά ή οποία όμως μπορεί να διαταραχθεί. Όσο μεγαλύτερο είναι το μέγεθος της διαταραχής στη συμπεριφορά της άδηλης μεταβλητής τόσο μεγαλύτερο είναι το σφάλμα διαταραχής (που συμβολίζεται με το γράμμα ζ).
- Για το παραπάνω λόγο ο ερευνητής υιοθετεί, επιλέγει ή κατασκευάζει μία μετρήσιμη μεταβλητή που θεωρεί ότι οι μεταβολές της σχετίζονται ικανοποιητικά με τις μεταβολές της αντίστοιχης άδηλης μεταβλητής. Όσο μεγαλύτερη είναι η συσχέτιση των μεταβολών της μετρήσιμης με τις αντίστοιχες μεταβολές της άδηλης τόσο μικρότερο είναι το σφάλμα μέτρησης (που συμβολίζεται με το γράμμα ε).
- Οι (άδηλες ή μετρήσιμες) μεταβλητές αναπτύσσουν μεταξύ τους σχέσεις (αιτιώδεις, συσχετίσεις, κ.λπ).

Διαμόρφωση Ερευνητικής Μεθοδολογίας: Από τη Θεωρία στις Αλγεβρικές Εκφράσεις


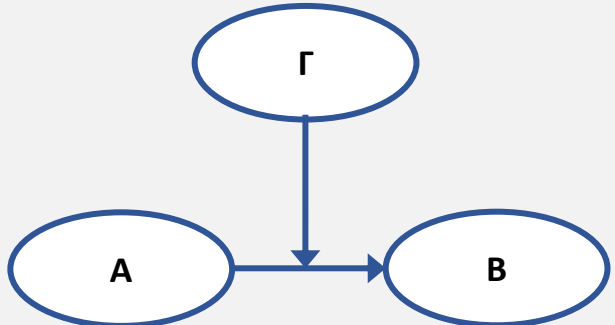

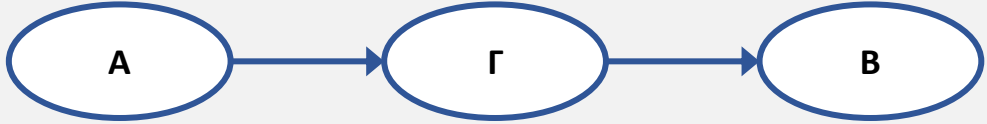
- Πως μπορούμε να εξάγουμε την αλγεβρική έκφραση μίας θεωρίας (συνέχεια);
 2. Εν συνεχεία, αποδίδονται γραφικά οι άδηλες μεταβλητές και οι σχέσεις που αναπτύσσουν μεταξύ τους με τη βοήθεια κοινά αποδεκτών συμβολικών αναπαραστάσεων.
 3. Στο τρίτο βήμα, ενσωματώνονται οι μετρήσιμες μεταβλητές στο προηγούμενο γραφικό μοντέλο.
 4. Με βάση το γραφικό μοντέλο όπως έχει διαμορφωθεί στα προηγούμενα βήματα διατυπώνεται η σχέση μεταξύ των **μετρήσιμων** μεταβλητών με αλγεβρικό τρόπο.
 5. Ορίζεται η εξίσωση παλινδρόμησης για το πληθυσμό με βάση την αλγεβρική έκφραση της θεωρίας.
 6. Ορίζεται η εξίσωση παλινδρόμησης που θα ελεγχθεί στατιστικά με τη αξιοποίηση ενός δείγματος πραγματικών, δηλαδή ιστορικά καταγεγραμμένων, τιμών των μετρήσιμων μεταβλητών.

Διαμόρφωση Ερευνητικής Μεθοδολογίας: Από τη Θεωρία στις Μαθηματικές Εκφράσεις

Structural Equation Modelling (SEM) – Symbols	
 Ονομασία	Το σύμβολο αυτό χρησιμοποιείται για να υποδηλώσει μία άδηλη για το μελετητή μεταβλητή.
 Ονομασία	Το σύμβολο αυτό χρησιμοποιείται για να υποδηλώσει μία μετρήσιμη (με τρόπο με τον οποίο ορίζει ο μελετητής) μεταβλητή.
	Αποδίδει την επίδραση μίας μεταβλητής σε μία άλλη μεταβλητή.
	Αποδίδει τη συσχέτιση ή συν διακύμανση μεταξύ ενός ζεύγους μεταβλητών.
 ε	Όταν έχει το ελληνικό γράμμα ε αναφέρεται σε σφάλμα μέτρησης μίας μετρήσιμης μεταβλητής.
 ζ	Όταν έχει το ελληνικό γράμμα ζ αναφέρεται σε διαταράξεις μίας άδηλης μεταβλητής.

Διαμόρφωση Ερευνητικής Μεθοδολογίας: Από τη Θεωρία στις Μαθηματικές Εκφράσεις

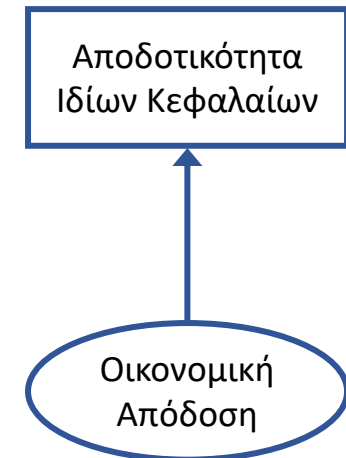
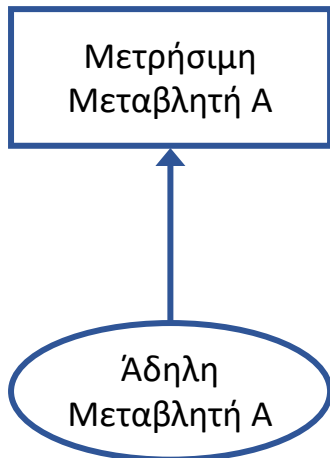
○ Είδη σχέσεων μεταξύ (άδηλων) μεταβλητών:

	<p>Άμεση επίδραση (direct effect): η άδηλη μεταβλητή A ασκεί άμεση επίδραση στην άδηλη μεταβλητή B.</p>	 <p>Μετριαστική επίδραση (moderating effect): η άδηλη μεταβλητή A ασκεί άμεση επίδραση στην άδηλη μεταβλητή B αλλά η ένταση της σχέσης αυτής επηρεάζεται από την άδηλη μεταβλητή Γ.</p>
	<p>Αλληλεπίδραση (covariance): η άδηλη μεταβλητή A ανταλλάσσει επιδράσεις με την άδηλη μεταβλητή B.</p>	
 <p>Μεσολαβητική επίδραση (mediating effect): η άδηλη μεταβλητή Γ δέχεται επιδράσεις από την άδηλη μεταβλητή A και μεσολαβεί ώστε οι επιδράσεις της άδηλης μεταβλητής A να επηρεάσουν την άδηλη μεταβλητή B. Η άδηλη μεταβλητή Γ μεσολαβεί στη σχέση των άδηλων μεταβλητών A και B.</p>		

Διαμόρφωση Ερευνητικής Μεθοδολογίας: Από τη Θεωρία στις Αλγεβρικές Εκφράσεις

Ορισμός μετρήσιμης μεταβλητής:

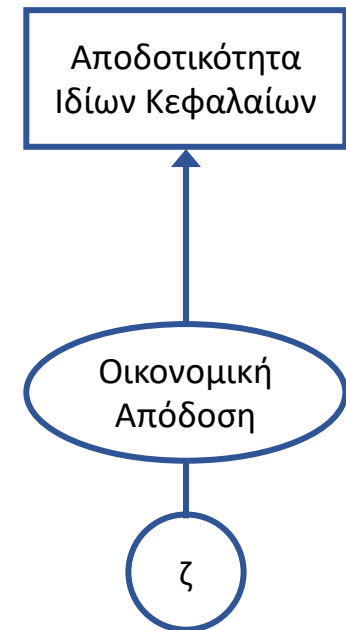
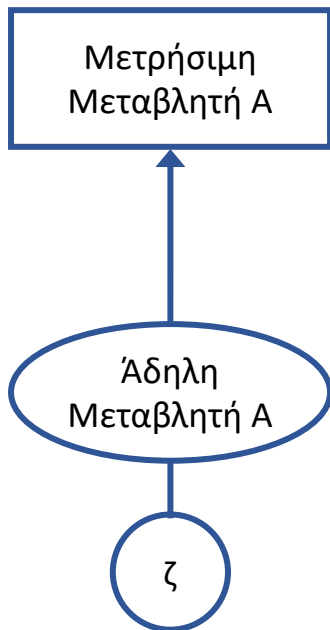
- Για κάθε άδηλη μεταβλητή θα πρέπει να ορισθεί μία μετρήσιμη μεταβλητή:
 - Ο μεταβολές στις αριθμητικές εκφράσεις της μετρήσιμης μεταβλητής πρέπει να αποδίδουν με αναλογικό τρόπο τις (μη παρατηρήσιμες) μεταβολές της άδηλης μεταβλητής.
 - Οι αριθμητικές εκφράσεις της μετρήσιμης μεταβλητής πρέπει να επιτρέπουν την ερμηνεία της συμπεριφοράς της άδηλης μεταβλητής.
 - Παράδειγμα: η αποδοτικότητα ιδίων κεφαλαίων αποτελεί μία μετρήσιμη μεταβλητή που αντιστοιχεί στην άδηλη μεταβλητή της οικονομικής απόδοσης μίας λογιστικής οντότητας.
 - Για την ίδια άδηλη μεταβλητή δύναται να υπάρχει παραπάνω από μία μετρήσιμη μεταβλητή.



Διαμόρφωση Ερευνητικής Μεθοδολογίας: Από τη Θεωρία στις Αλγεβρικές Εκφράσεις

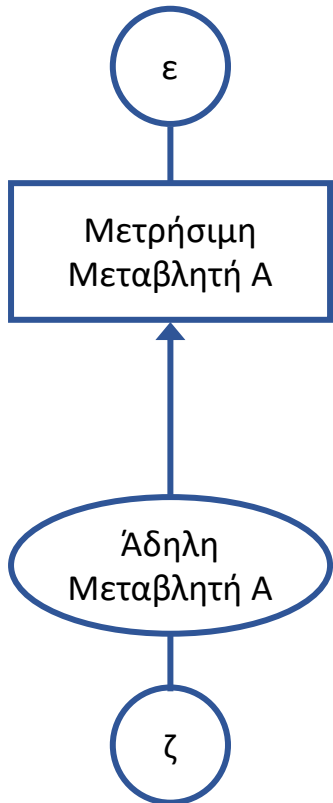
○ Σφάλματα μεταβλητών:

- Τόσο η άδηλη όσο και η αντίστοιχη μετρήσιμη μεταβλητή συνδέονται με σφάλματα.
- Η άδηλη μεταβλητή συνδέεται με **σφάλμα διατάραξης (ζ)** το οποίο οφείλεται (μεταξύ άλλων):
 - Σε σφάλμα θεωρητικής εξειδίκευσης (π.χ., παράλειψη μίας άλλης άδηλης μεταβλητής).
 - Σε τυχαίες επιδράσεις κάποιου εξωγενή παράγοντα (π.χ., ένα μη προβλέψιμο οικονομικό σοκ).

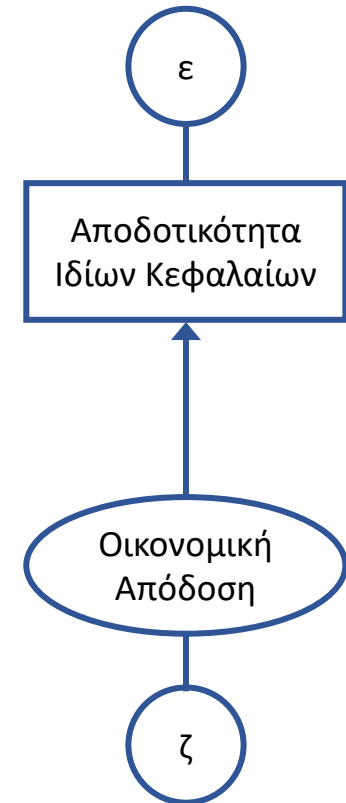


Διαμόρφωση Ερευνητικής Μεθοδολογίας: Από τη Θεωρία στις Αλγεβρικές Εκφράσεις

○ Σφάλματα μεταβλητών:



- Η μετρήσιμη μεταβλητή συνδέεται με **σφάλμα μέτρησης (ε)** το οποίο οφείλεται (μεταξύ άλλων):
 - Το εύρος των αριθμητικών τιμών της μετρήσιμης μεταβλητής δεν καλύπτει το εύρος των μεταβολών της άδηλης μεταβλητής.
 - Οι αριθμητικές τιμές της μετρήσιμης μεταβλητής δεν έχουν το ίδιο επίπεδο ευαισθησίας για όλο το εύρος των τιμών της άδηλης μεταβλητής (π.χ., σε ακραίες μεταβολές της άδηλης μεταβλητής).
 - Η μετρήσιμη μεταβλητή ενσωματώνει τις επιδράσεις από άλλες άδηλες μεταβλητές.
 - Η μετρήσιμη μεταβλητή δεν είναι η καλύτερη δυνατή.



Διαμόρφωση Ερευνητικής Μεθοδολογίας: Από τη Θεωρία στις Μαθηματικές Εκφράσεις

- Το παράδειγμα που ακολουθεί έχει ληφθεί από δημοσιευμένη ερευνητική εργασία όπου πραγματεύεται την έννοια της ποιότητας των λογιστικών κερδών και τους προσδιοριστικούς παράγοντές της.
- Ως παράδειγμα είναι χρήσιμο για την κατανόηση πως η θεωρία αποτυπώνεται σε μεταβλητές και σε σχέσεις μεταξύ τους και, εν συνεχεία, τα παραπάνω μετατρέπονται σε αλγεβρικές εκφράσεις.



ELSEVIER

Contents lists available at ScienceDirect

Journal of Accounting and Economics

journal homepage: www.elsevier.com/locate/jae



Understanding earnings quality: A review of the proxies, their determinants and their consequences[☆]

Patricia Dechow^a, Weili Ge^b, Catherine Schrand^{c,*}

^a University of California, Berkeley, CA 94720, United States

^b University of Washington, Seattle, WA 98195, United States

^c University of Pennsylvania, Philadelphia, PA 19104, United States

A B S T R A C T

Researchers have used various measures as indications of “earnings quality” including persistence, accruals, smoothness, timeliness, loss avoidance, investor responsiveness, and external indicators such as restatements and SEC enforcement releases. For each measure, we discuss causes of variation in the measure as well as consequences. We reach no single conclusion on what earnings quality is because “quality” is contingent on the decision context. We also point out that the “quality” of earnings is a function of the firm’s fundamental performance. The contribution of a firm’s fundamental performance to its earnings quality is suggested as one area for future work.

© 2010 Elsevier B.V. All rights reserved.

Διαμόρφωση Ερευνητικής Μεθοδολογίας: Από τη Θεωρία στις Μαθηματικές Εκφράσεις

ABSTRACT

Researchers have used various measures as indications of “earnings quality” including persistence, accruals, smoothness, timeliness, loss avoidance, investor responsiveness, and external indicators such as restatements and SEC enforcement releases. For each measure, we discuss causes of variation in the measure as well as consequences. **We reach no single conclusion on what earnings quality is because “quality” is contingent on the decision context. We also point out that the “quality” of earnings is a function of the firm’s fundamental performance. The contribution of a firm’s fundamental performance to its earnings quality is suggested as one area for future work.**

© 2010 Elsevier B.V. All rights reserved.

Βασικά σημεία:

- Η ποιότητα των λογιστικών κερδών είναι μία ενδεχομενική έννοια η οποία μπορεί να λάβει διάφορες εννοιολογικές αποχρώσεις ανάλογα με πλαίσιο λήψης απόφασης που αξιοποιεί την λογιστική πληροφορία αναφορικά με τα κέρδη.
- Η ποιότητα των λογιστικών κερδών εξαρτάται από την πραγματική λειτουργική απόδοση της οποίας τα λογιστικά κέρδη αποτελούν μία μόνο διάστασή της.

Διαμόρφωση Ερευνητικής Μεθοδολογίας: Από τη Θεωρία στις Μαθηματικές Εκφράσεις

In order to provide a commentary on the state of the literature, we begin by presenting a framework for thinking about earnings quality. For expositional convenience, we define reported earnings as follows:

Reported Earnings $\equiv f(X)$.

X is the “...enterprise’s financial performance during a reporting period,” which SFAC No. 1 states is what earnings, a primary focus of financial reporting, should represent.⁶ **The function f represents the accounting system that converts the unobservable X into observable earnings. One implication of this definition is that earnings, and the decision usefulness of earnings, is a function of performance itself, and not just the measurement of X ,** which is an important point that we will return to later.

Οι συγγραφείς εξειδικεύουν την θέση τους: Τα αναφερόμενα από τις οικονομικές καταστάσεις λογιστικά κέρδη είναι συνάρτηση της πραγματικής λειτουργικής απόδοσης.

Διαμόρφωση Ερευνητικής Μεθοδολογίας: Από τη Θεωρία στις Μαθηματικές Εκφράσεις

The second noteworthy feature of our definition of reported earnings is that reported earnings does not equal X ; it is a function of X . We distinguish three explanations for why an accounting measurement system (f) would not perfectly measure performance:

- (1) **Multiple decision models:** An accounting system that produces a single reported earnings number cannot produce a representation of X that is equally relevant in all decision models. In U.S. GAAP: “The objectives are directed toward the common interests of many users in the ability of an enterprise to generate favorable cash flows but are phrased using investment and credit decisions as a reference to give them a focus” (SFAC No. 1). Ultimately, the standard setters make trade-offs in setting standards across anticipated users’ needs, and in the end *no* individual decision-maker gets a representation of firm performance that is perfectly relevant for his or her decision.⁷
- (2) **Variation in X :** Firms choose among a limited set of pre-determined measurement principles (e.g., accounting standards) to measure X . No single standard will perfectly measure X for any given firm.⁸ Consider, for example, cost of goods sold (COGS), which represents the reportable measure of a firm’s unobservable inventory production performance during the period. GAAP defines the costs to be included in COGS and the timing of the recognition of the costs. However, the resulting “standardized” measure of COGS will not be an equally good measure of decision-relevant performance across all X s (e.g., retail chains versus oil producing companies, to use the Graham and Dodd example), and it will not be a perfect representation of any X .
- (3) **Implementation:** An accounting system that measures an unobservable construct (X) inherently involves estimations and judgment, and thus has the potential for unintentional errors and intentional bias (i.e., earnings management).

Το λογιστικό σύστημα της λογιστικής οντότητας είναι η συνάρτηση που μετασχηματίζει τη πραγματική λειτουργική απόδοση στα αναφερόμενα από τις οικονομικές καταστάσεις λογιστικά κέρδη.

Το λογιστικό σύστημα της λογιστικής οντότητας δεν αποτελεί το τέλειο σύστημα μέτρησης της πραγματικής λειτουργικής απόδοσης:

- Δεν είναι εύκολο να ικανοποιήσει ετερογενή μοντέλα αποφάσεων.
- Η πραγματική λειτουργική απόδοση έχει πολλές πτυχές.
- Ακούσια ή εκούσια σφάλματα μέτρησης.

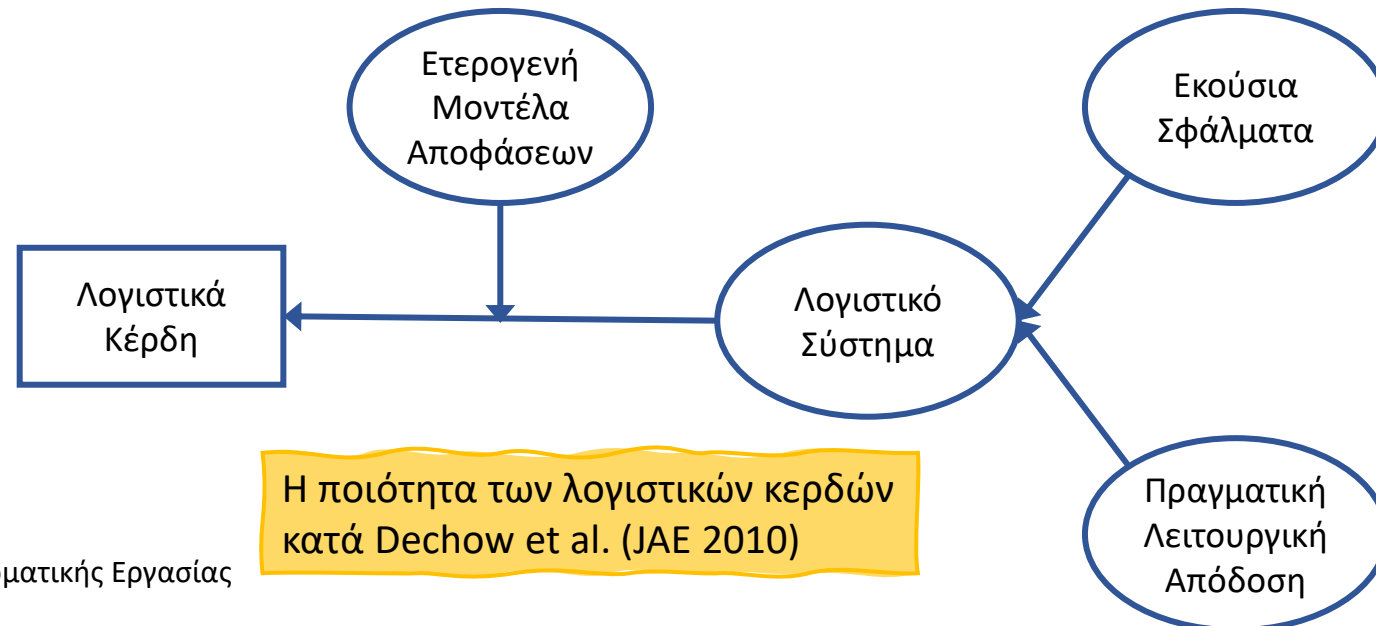
Διαμόρφωση Ερευνητικής Μεθοδολογίας: Από τη Θεωρία στις Μαθηματικές Εκφράσεις

Το λογιστικό σύστημα της λογιστικής οντότητας είναι η συνάρτηση που μετασχηματίζει τη πραγματική λειτουργική απόδοση στα αναφερόμενα από τις οικονομικές καταστάσεις λογιστικά κέρδη.

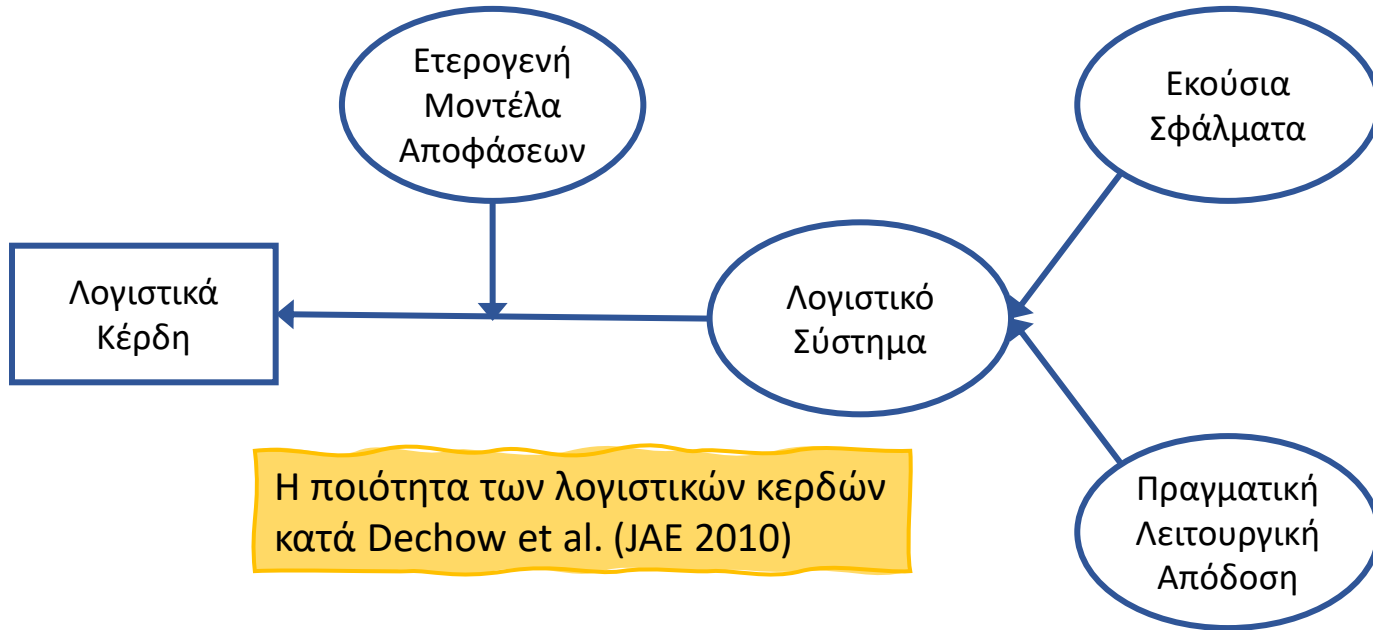
Το λογιστικό σύστημα της λογιστικής οντότητας δεν αποτελεί το τέλειο σύστημα μέτρησης της πραγματικής λειτουργικής απόδοσης:

- Δεν είναι εύκολο να ικανοποιήσει ετερογενή μοντέλα αποφάσεων.
- Η πραγματική λειτουργική απόδοση έχει πολλές πτυχές.
- Ακούσια ή εκούσια σφάλματα μέτρησης.

Τα παραπάνω αποδίδονται γραφικά ως εξής:



Διαμόρφωση Ερευνητικής Μεθοδολογίας: Από τη Θεωρία στις Μαθηματικές Εκφράσεις



Η θεωρία των Dechow et al. (JAE 2010) μπορεί να θεωρηθεί ένα γενικευμένο πλαίσιο επί της βάση του οποίου θα δομηθούν επιμέρους ερευνητικές υποθέσεις από άλλους ερευνητές.

Ας θυμηθούμε την ερευνητική εργασία του Roychowdhury (2006). Αν και δημοσιεύθηκε νωρίτερα από τη θεωρία των Dechow et al. (2015) είναι δυνατόν να την τοποθετήσουμε στην θεωρία αναφορικά με την ποιότητα των λογιστικών κερδών;



Available online at www.sciencedirect.com



Journal of Accounting and Economics 42 (2006) 335–370



www.elsevier.com/locate/jae

Earnings management through real activities manipulation[☆]

Sugata Roychowdhury*

Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02142, USA

Διαμόρφωση Ερευνητικής Μεθοδολογίας: Από τη Θεωρία στις Μαθηματικές Εκφράσεις

Ας θυμηθούμε την ερευνητική εργασία του Roychowdhury (2006).

Αν και δημοσιεύθηκε νωρίτερα από τη θεωρία των Dechow et al. (2015) είναι δυνατόν να την τοποθετήσουμε στην θεωρία αναφορικά με την ποιότητα των λογιστικών κερδών;



Available online at www.sciencedirect.com



Journal of Accounting and Economics 42 (2006) 335–370



www.elsevier.com/locate/jae

Earnings management through real activities manipulation ☆

Sugata Roychowdhury*

Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02142, USA

Υπό τη συνθήκη ότι η λογιστική οντότητα είναι οριακά κερδοφόρα (=υψηλό κίνητρο για χειραγωγήση κερδών), τα διοικητικά στελέχη είναι δυνατόν να χειραγωγήσουν τα λογιστικά κέρδη μέσω λειτουργικών αποφάσεων (real earnings management).

Πτυχές χειραγωγής λογιστικών κερδών μέσω λειτουργικών αποφάσεων (real earnings management) :

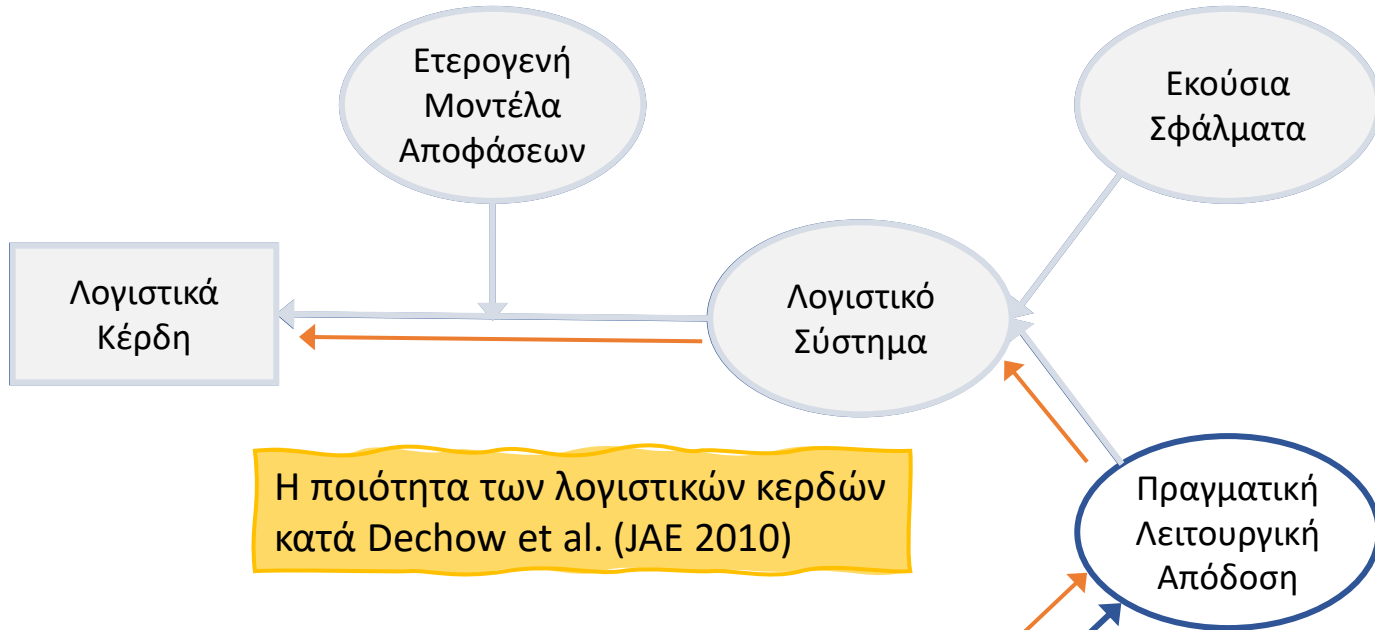
- Καθαρές ταμειακές ροές από λειτουργικές δραστηριότητες εξαιτίας της μόχλευσης των πωλήσεων.
- Η μείωση των δαπανών που υπόκεινται στη διακριτική ευχέρεια των διοικητικών στελεχών.
- Η μείωση του ανά μονάδα κόστους εξαιτίας της αδικαιολόγητης αύξησης της παραγωγής.

For my primary tests, *suspect firm-years* are firm-years reporting small annual profits. My two main hypotheses, stated in alternate form, are the following:

H1A. *After controlling for sales levels, suspect firm-years exhibit at least one of the following: unusually low cash flow from operations (CFO) OR unusually low discretionary expenses.*

H2A. *After controlling for sales levels, suspect firm-years exhibit unusually high production costs.*

Διαμόρφωση Ερευνητικής Μεθοδολογίας: Από τη Θεωρία στις Μαθηματικές Εκφράσεις



For my primary tests, *suspect firm-years* are firm-years reporting small annual profits. My two main hypotheses, stated in alternate form, are the following:

H1A. After controlling for sales levels, *suspect firm-years* exhibit at least one of the following: unusually low cash flow from operations (CFO) OR unusually low discretionary expenses.

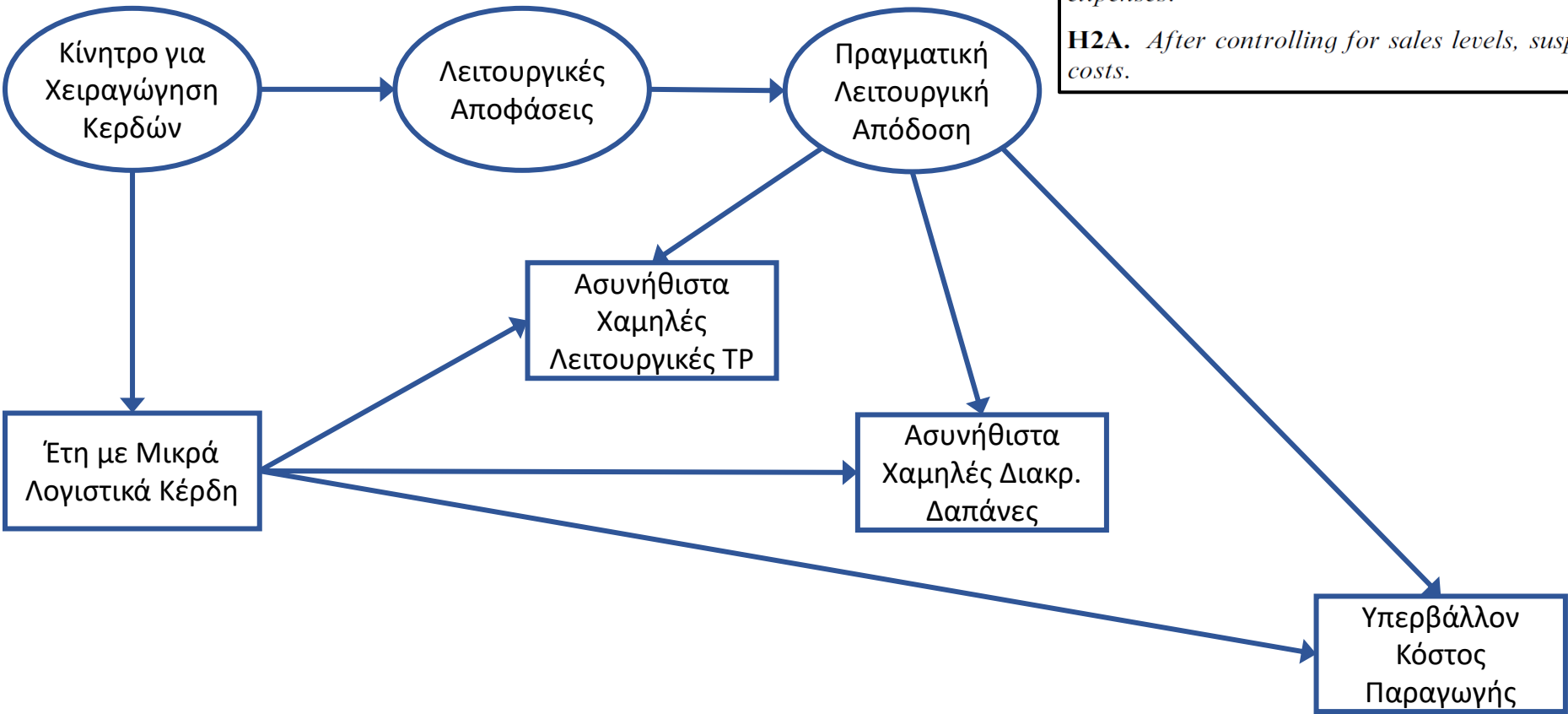
H2A. After controlling for sales levels, *suspect firm-years* exhibit unusually high production costs.

Διαμόρφωση Ερευνητικής Μεθοδολογίας: Από τη Θεωρία στις Μαθηματικές Εκφράσεις

For my primary tests, *suspect firm-years* are firm-years reporting small annual profits. My two main hypotheses, stated in alternate form, are the following:

H1A. After controlling for sales levels, suspect firm-years exhibit at least one of the following: unusually low cash flow from operations (CFO) OR unusually low discretionary expenses.

H2A. After controlling for sales levels, suspect firm-years exhibit unusually high production costs.



Διαμόρφωση Ερευνητικής Μεθοδολογίας: Από τη Θεωρία στις Μαθηματικές Εκφράσεις

Έτη με Μικρά Λογιστικά Κέρδη

Ψευδομεταβλητή: λαμβάνει την τιμή 1 αν τα λογιστικά κέρδη είναι μικρά, διαφορετικά 0

Ασυνήθιστα Χαμηλές Λειτουργικές TP

Πραγματικές Λειτουργικές TP μείον Προβλεπόμενες Λειτουργικές TP

Ασυνήθιστα Χαμηλές Διακρ. Δαπάνες

Πραγματικές Διακριτές Δαπάνες μείον Προβλεπόμενες Διακριτές Δαπάνες

Υπερβάλλον Κόστος Παραγωγής

Πραγματικό Κόστος Παραγωγής μείον Προβλεπόμενο Κόστος Παραγωγής

Παρατήρηση: ο τρόπος μέτρησης των μετρήσιμων μεταβλητών ενέχει σφάλμα μέτρησης (ϵ).

research methods - σεμινάρια εκπόνησης διπλωματικής εργασίας

To increase the power of my tests to detect real activities manipulation, I concentrate on firm-years in the interval to the immediate right of zero, the *suspect firm-years*. Suspect firm-years have net income scaled by total assets that is greater than or equal to zero but less than 0.005 (interval 16 in the figure). There are 503 suspect firm-years, including 450 unique firms.

$$CFO_t/A_{t-1} = \alpha_0 + \alpha_1(1/A_{t-1}) + \beta_1(S_t/A_{t-1}) + \beta_2(\Delta S_t/A_{t-1}) + \epsilon_t, \quad (1)$$

where A_t is the total assets at the end of period t , S_t the sales during period t and $\Delta S_t = S_t - S_{t-1}$.

$$DISEXP_t/A_{t-1} = \alpha_0 + \alpha_1(1/A_{t-1}) + \beta(S_{t-1}/A_{t-1}) + \epsilon_t. \quad (5)$$

I define production costs as $PROD_t = COGS_t + \Delta INV_t$. Using (2) and (3), I estimate normal production costs from the following industry-year regression.¹⁹

$$PROD_t/A_{t-1} = \alpha_0 + \alpha_1(1/A_{t-1}) + \beta_1(S_t/A_{t-1}) + \beta_2(\Delta S_t/A_{t-1}) + \beta_3(\Delta S_{t-1}/A_{t-1}) + \epsilon_t. \quad (4)$$

$$COGS_t/A_{t-1} = \alpha_0 + \alpha_1(1/A_{t-1}) + \beta(S_t/A_{t-1}) + \epsilon_t. \quad (2)$$

$$\Delta INV_t/A_{t-1} = \alpha_0 + \alpha_1(1/A_{t-1}) + \beta_1(\Delta S_t/A_{t-1}) + \beta_2(\Delta S_{t-1}/A_{t-1}) + \epsilon_t, \quad (3)$$

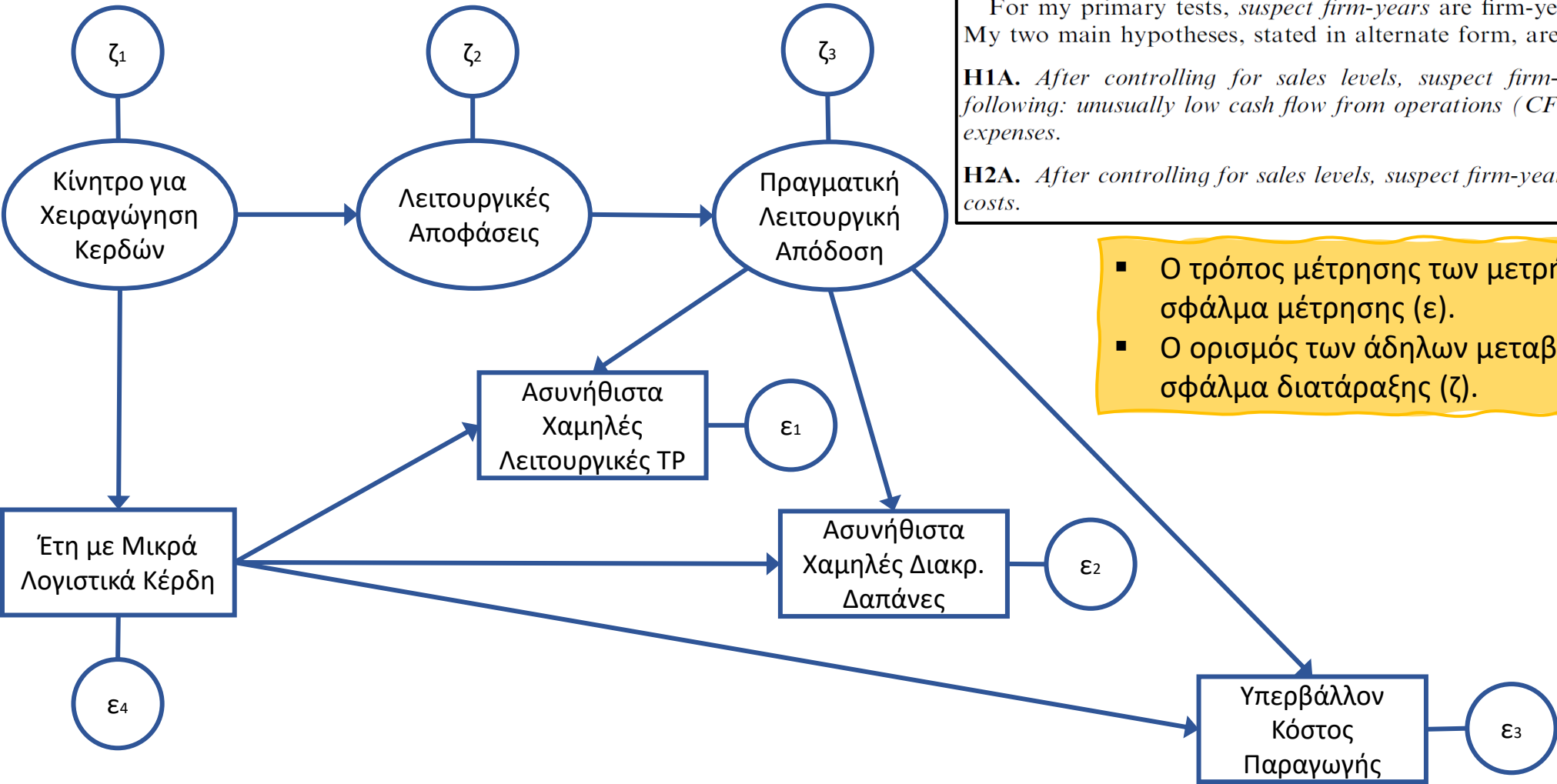
Διαμόρφωση Ερευνητικής Μεθοδολογίας: Από τη Θεωρία στις Μαθηματικές Εκφράσεις

For my primary tests, *suspect firm-years* are firm-years reporting small annual profits. My two main hypotheses, stated in alternate form, are the following:

H1A. After controlling for sales levels, *suspect firm-years* exhibit at least one of the following: *unusually low cash flow from operations (CFO) OR unusually low discretionary expenses.*

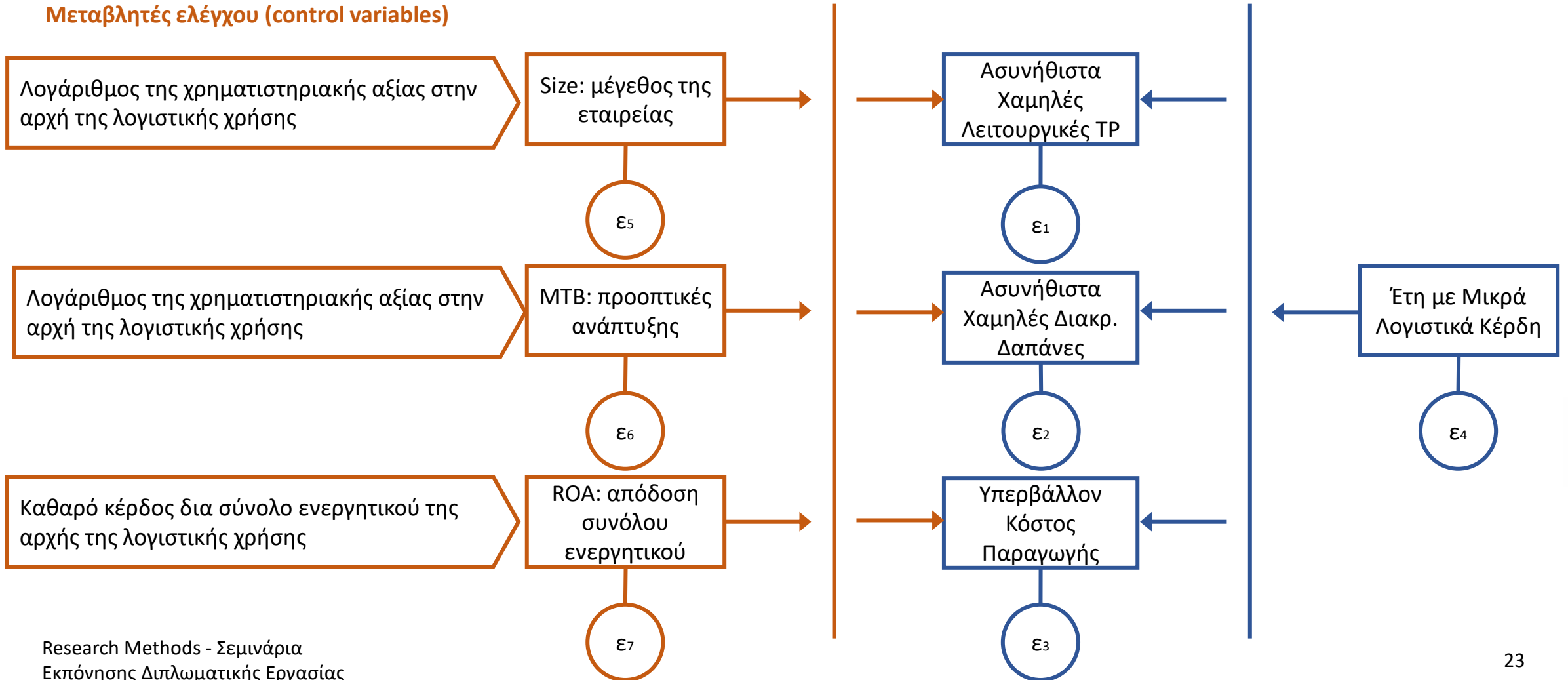
H2A. After controlling for sales levels, *suspect firm-years* exhibit *unusually high production costs.*

- Ο τρόπος μέτρησης των μετρήσιμων μεταβλητών ενέχει σφάλμα μέτρησης (ϵ).
- Ο ορισμός των άδηλων μεταβλητών συνδέεται με σφάλμα διατάραξης (ζ).



Διαμόρφωση Ερευνητικής Μεθοδολογίας: Από τη Θεωρία στις Μαθηματικές Εκφράσεις

Μεταβλητές ελέγχου (control variables)

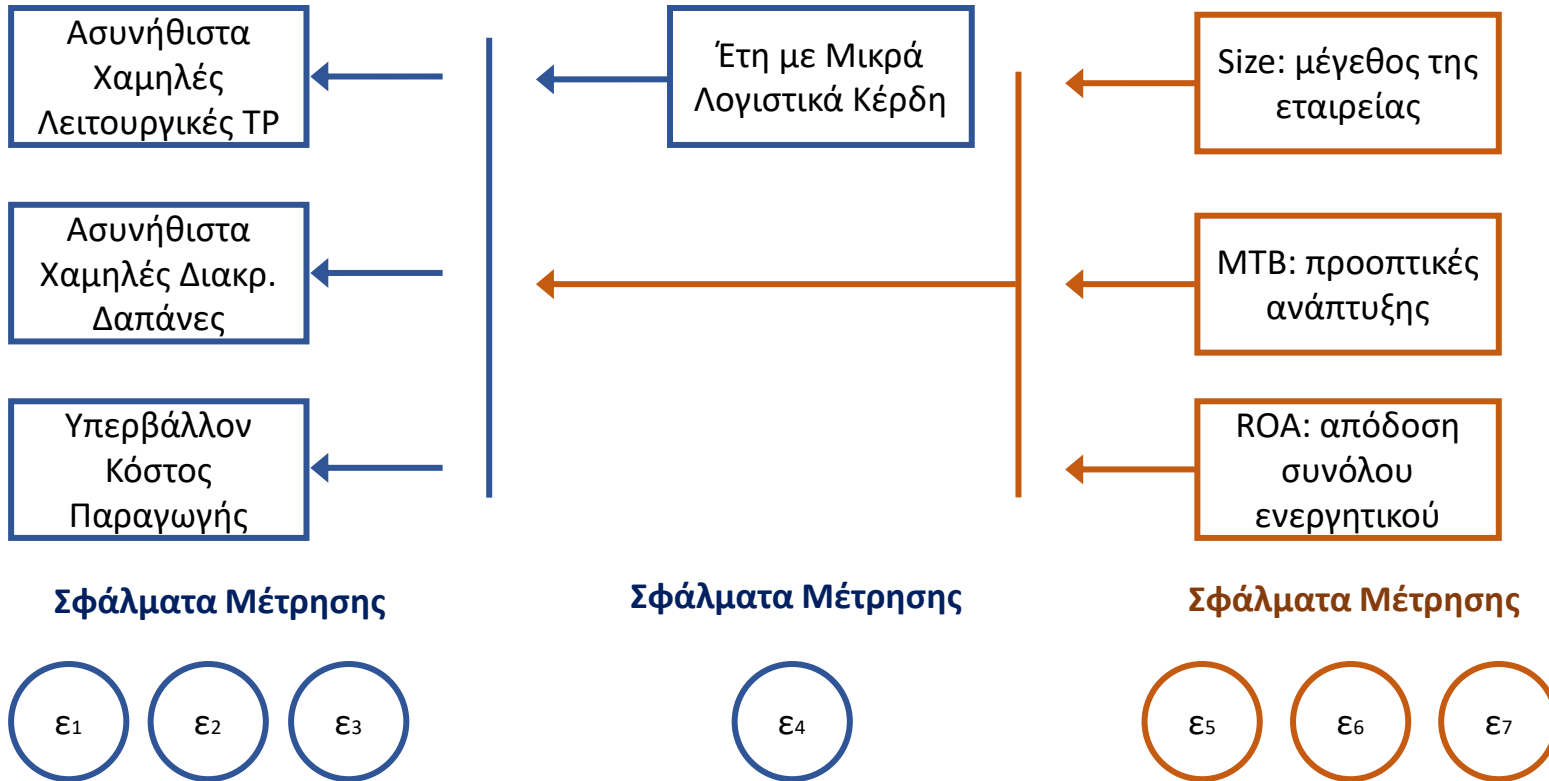


Διαμόρφωση Ερευνητικής Μεθοδολογίας: Από τη Θεωρία στις Μαθηματικές Εκφράσεις

Εξαρτημένες Μεταβλητές

Ανεξάρτητη Μεταβλητή Κύριου Ερευνητικού Ενδιαφέροντος

Ανεξάρτητες Μεταβλητές Ελέγχου



Διαμόρφωση Ερευνητικής Μεθοδολογίας: Από τη Θεωρία στις Μαθηματικές Εκφράσεις

Εξαρτημένες Μεταβλητές

- Ασυνήθιστα Χαμηλές Λειτουργικές TP
- Ασυνήθιστα Χαμηλές Διακρ. Δαπάνες
- Υπερβάλλον Κόστος Παραγωγής

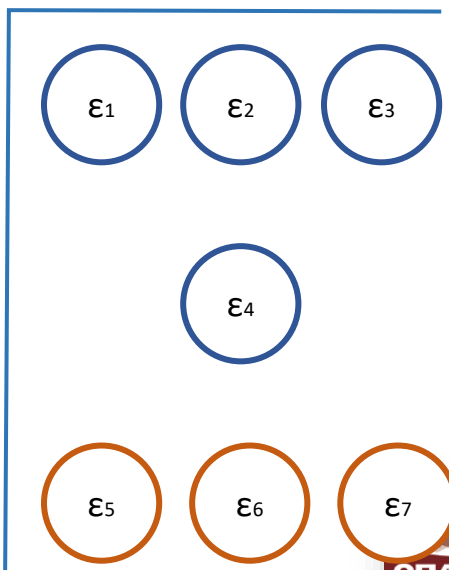
Ανεξάρτητες Μεταβλητές Ελέγχου

- Size: μέγεθος της εταιρείας
- MTB: προοπτικές ανάπτυξης
- ROA: απόδοση συνόλου ενεργητικού

Ανεξάρτητη Μεταβλητή Κύριου Ερευνητικού Ενδιαφέροντος




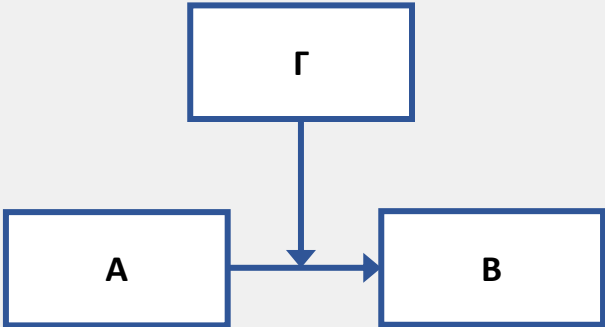
- Έτη με Μικρά Λογιστικά Κέρδη

Σφάλματα Μέτρησης



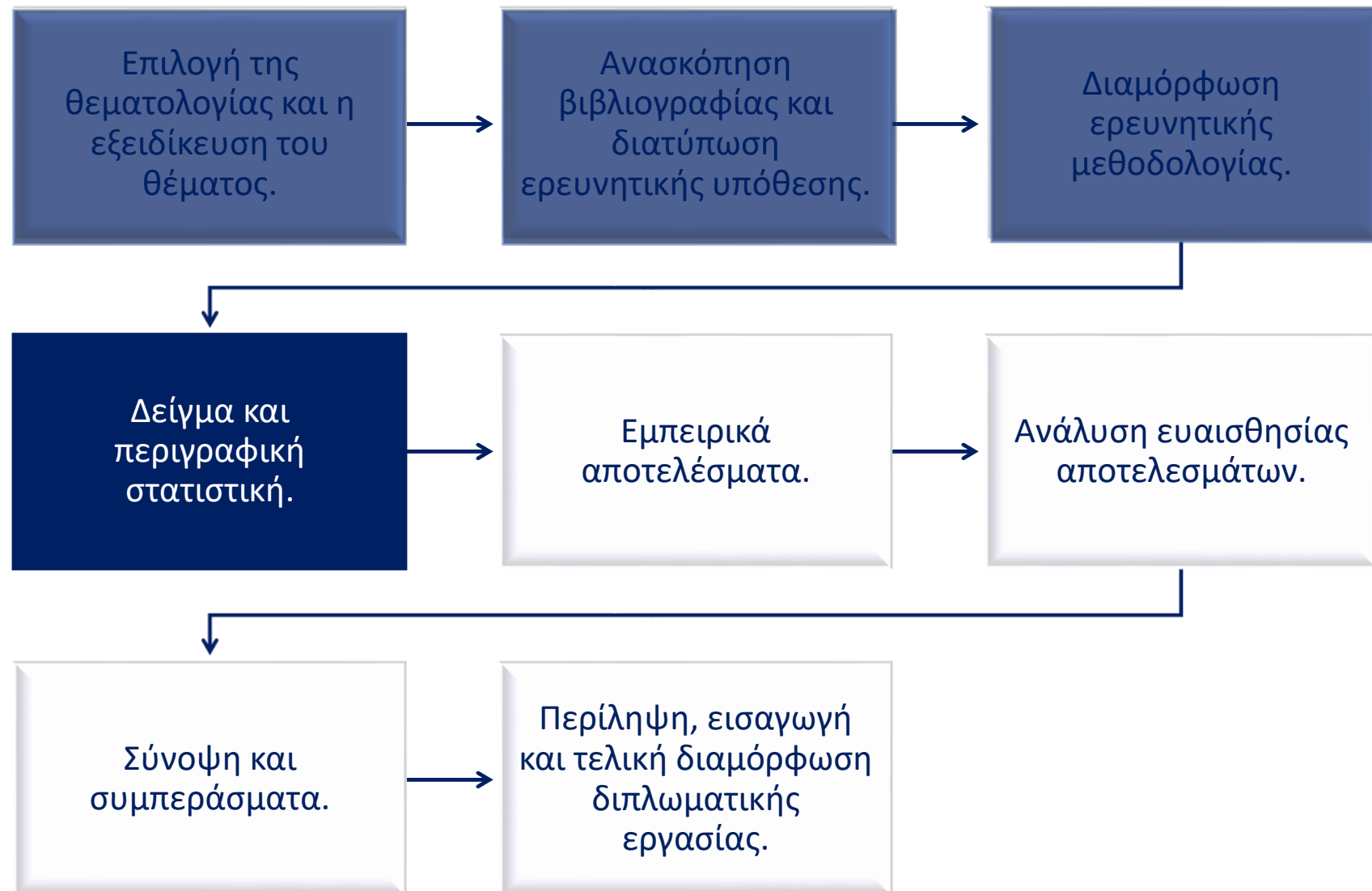
$$Y_t = \alpha + \beta_1(\text{SIZE})_{t-1} + \beta_2(\text{MTB})_{t-1} + \beta_3(\text{Net income})_t + \beta_4(\text{SUSPECT_NI})_t + \epsilon_t.$$

Τυπολογία Πιθανών Οικονομετρικών Εξειδικεύσεων

	Είδος σχέσης	Πιθανή Οικονομετρική Εξειδίκευση
	Άμεση επίδραση (direct effect)	$B_t = a_1 + b_1 A_t + e_t$
	Αλληλεπίδραση (covariance)	$B_t = a_1 + b_1 A_{t-1} + e_t$ $A_t = c_1 + d_1 B_{t-1} + u_t$
	Μεσολαβητική επίδραση (mediating effect):	$\Gamma_t = a_1 + b_1 A_t + e_t$ $B_t = c_1 + d_1 \Gamma_t + u_t$
	Μετριαστική επίδραση (moderating effect):	$B_t = a_1 + b_1 A_t + c_1 A_t \times \Gamma_t + e_t$



Δείγμα και Περιγραφική Στατιστική



Δείγμα Παρατηρήσεων

- Με τον όρο δείγμα αναφερόμαστε σε ένα σύνολο τιμών των μετρήσιμων μεταβλητών και λοιπών μεγεθών που είναι απαραίτητα για την εμπειρική εφαρμογή της ερευνητικής μεθοδολογίας (π.χ., εκτίμηση οικονομετρικού υποδείγματος, κ.λπ).
- Τα δεδομένα του δείγματος δύναται να συλλέγονται πρωτογενώς (π.χ., μέσω ερωτηματολογίων, συνεντεύξεων, μελετών περιπτώσεων, κ.λπ) αλλά κυρίως είναι δευτερογενή στη φύση τους, δηλαδή εκμαιεύονται από μία βάση δεδομένων.
- Το Τμήμα Λογιστικής & Χρηματοοικονομικής παρέχει πρόσβαση σε αρκετές βάσεις δεδομένων (Bloomberg, Datastream, κ.ά.) μέσω των εργαστηρίων του. Έχει γίνει σχετική επίδειξη και στο e-class μπορείτε να βρείτε ένα πληροφοριακό έντυπο από τις διαθέσιμες βάσεις δεδομένων.

Λογικά Βήματα για τη Διαμόρφωση του Δείγματος

- Τα λογιστικά βήματα για τη διαμόρφωση του δείγματος είναι τα ακόλουθα:
- **Βήμα 1^ο** : Καθορισμός των μεταβλητών και της χρονικής περιόδου. Συνίσταται η καταγραφή της λίστας των εταιρειών (κάθε εταιρεία αποτελεί μία περίπτωση), των μεταβλητών (συνήθως αποτελούν διακριτά λογιστικά ή οικονομικά μεγέθη) και της χρονικής περιόδου (συνήθως με αναφορά σε ετήσια, τριμηνιαία, μηνιαία ή καθημερινά δεδομένα) με ιδιαίτερη προσοχή.
- **Βήμα 2^ο** : Κατέβασμα δεδομένων. Με βάση την προεργασία του βήματος 1 εντοπίζονται οι κωδικές ονομασίες των μεταβλητών σύμφωνα με την κωδικοποίηση της βάση δεδομένων (στο e-class σας παρέχεται ο αναλυτικός κατάλογος κωδικών για όλες τις μεταβλητές που περιέχονται στην βάση δεδομένων Worldscor). Ένα παράδειγμα μεταβλητών και αντίστοιχων κωδικών παρουσιάζονται στον ακόλουθο πίνακα:

Λογιστικό μέγεθος	Κωδικός REUTERS
Sales per Share	WC05508
Return on Equity – Total (%)	WC08301
Cash	WC02003

Λογικά Βήματα για τη Διαμόρφωση του Δείγματος

Εφαρμογή 2^η

Λογικά Βήματα για τη Διαμόρφωση του Δείγματος

- **Βήμα 3^ο** : Εκκαθάριση δεδομένων. Τα δεδομένα θα πρέπει να εκκαθαρισθούν λαμβάνοντας υπόψη τα παρακάτω:
 - Διαγραφή τιμών που δηλώνονται ως μη διαθέσιμες N/A. Δηλαδή αντικατάσταση της ένδειξης N/A σε κενό.
 - Διαγραφή τιμών που είναι πρόδηλα εσφαλμένες. Ως πρόδηλα εσφαλμένο θεωρείται οτιδήποτε αντιβαίνει θεμελιώδεις αρχές. Παραδείγματα: πωλήσεις με αρνητικές τιμές, κυκλοφορούν ενεργητικό με αρνητικές τιμές, κ.λπ.
 - Αντιμετώπιση ακραίων τιμών (outliers). Πιθανές ακραίες τιμές μίας ή περισσότερων μεταβλητών είναι δυνατόν να επηρεάσουν τα αποτελέσματα της ανάλυσης. Έχουν διαμορφωθεί δύο αντιλήψεις επί του τρόπου αντιμετώπισης των ακραίων τιμών. Η πρώτη διατείνεται ότι ο ερευνητής δεν πρέπει να πειράξει τις ακραίες τιμές διότι εμπεριέχουν πληροφόρηση που πρέπει να ληφθεί υπόψη από την εμπειρική ανάλυση. Η δεύτερη θεωρεί ότι οι ακραίες τιμές αποτελούν μία πηγή σφάλματος και άρα πρέπει να αντιμετωπισθούν με έναν από τους ακόλουθους τρόπους:
 - Διαγραφή όλων των τιμών που ευρίσκονται στο χαμηλότερο ή υψηλότερο 5% (ή 1%) του εύρους τιμών της μεταβλητής (trimming).
 - Αντικατάσταση όλων των τιμών που ευρίσκονται στο χαμηλότερο 5% (ή 1%) του εύρους τιμών της μεταβλητής με την τιμή της που αντιστοιχεί στο 5% (ή 1%) του εύρους τιμών της ΚΑΙ αντικατάσταση όλων των τιμών που ευρίσκονται στο υψηλότερο 5% (ή 1%) του εύρους τιμών της μεταβλητής με την τιμή της που αντιστοιχεί στο 95% (ή 99%) του εύρους τιμών της (winsorization).
 - Οποιοδήποτε άλλο τρόπο υποδεικνύει η εξειδικευμένη βιβλιογραφία.

Λογικά Βήματα για τη Διαμόρφωση του Δείγματος

- **Βήμα 4^ο** : Μετατροπή μεταβλητών συμβολοσειρών σε αριθμητικές μεταβλητές. Το STATA αναγνωρίζει κατά κύριο λόγο μία μεταβλητή ως συμβολοσειρά (string variable) ή ως αριθμητική (numeric). Οι μεταβλητές συμβολοσειράς μπορεί να περιέχουν και χαρακτήρες και αριθμούς, οι αριθμητικές μεταβλητές μόνο αριθμούς. Στο πλαίσιο της οικονομετρικής στατιστικής ανάλυσης, ο ερευνητής θα πρέπει να βεβαιωθεί ότι όλες οι μεταβλητές που θα αξιοποιήσει είναι αριθμητικές. Αν υπάρχουν μεταβλητές ως συμβολοσειρές θα πρέπει να μετατραπούν σε αριθμητικές.
- **Βήμα 5^ο** : Διαμόρφωση πίνακα αρχικού και τελικού δείγματος. Συντάσσετε πίνακα που περιγράφει τον αρχικό αριθμό των παρατηρήσεων, τις αιτίες διαγραφών παρατηρήσεων, τον αριθμό των παρατηρήσεων που διαγράφηκαν για κάθε αιτία διαγραφής και ο τελικός αριθμός των παρατηρήσεων που χρησιμοποιήθηκε για εμπειρική ανάλυση.

Λογικά Βήματα για τη Διαμόρφωση του Δείγματος

Εφαρμογή 3^η

Κατηγορίες Μεταβλητών

- **Διάκριση μεταβλητών με κριτήριο το χρόνο και το χώρο:**

- **Χρονολογική σειρά:** μία χρονολογική ακολουθία τιμών για μία περίπτωση. Παραδείγματα: η αξία των πωλήσεων μίας εταιρείας της τελευταίας δεκαετίας, η αξία των μηνιαίων πωλήσεων μίας εταιρείας για το χρονικό διάστημα από το 1990 έως και το 2020, το μηνιαίο βάρος ενός αθλητή τους τελευταίους 36 μήνες.
- **Διαστρωματικά δεδομένα:** μία ακολουθία τιμών για όλες τις περιπτώσεις σε ένα χρονικό σημείο. Παραδείγματα: το ύψος των πωλήσεων ενός αριθμού εταιρειών για τη λογιστική χρήση 2024, η αξία των μηνιαίων πωλήσεων ενός αριθμού εταιρειών για το μήνα Ιανουάριο 2024, το βάρος ενός αριθμού αθλητών για το μήνα Ιανουάριο 2024.
- **Δεδομένα τύπου πάνελ:** συνδυασμός χρονολογικής σειράς και διαστρωματικών δεδομένων, δηλαδή ένας αριθμός χρονολογικών ακολουθιών τιμών και έναν αριθμό περιπτώσεων. Παραδείγματα: το ύψος των πωλήσεων ενός αριθμού εταιρειών της τελευταίας δεκαετίας, η αξία των μηνιαίων πωλήσεων ενός αριθμού εταιρειών για το χρονικό διάστημα από το 1990 έως και το 2020, το βάρος ενός αριθμού αθλητών τους τελευταίους 36 μήνες.

Κατηγορίες Μεταβλητών

- **Διάκριση μεταβλητών με τη φύση τους:**

- **Ποιοτικές (Qualitative):** μεταβλητές οι οποίες μεταβάλλονται σε είδος, χαρακτηρισμούς ή κατηγορίες, διαφοροποιούνται μεταξύ τους ως προς διαφορετικά χαρακτηριστικά. Παραδείγματα: είδος εταιρείας (εισηγμένη/μη εισηγμένη), κλάδος δραστηριότητας (βιομηχανία/εμπόριο), κ.λπ. Περαιτέρω διάκριση:
 - **Κατηγορικές (categorical ή nominal).**
 - **Διατεταγμένες (ordinal).**
- **Ποσοτικές (Quantitative):** μεταβλητές οι οποίες μεταβάλλονται σε ποσό, περιγράφουν ποσότητες. Παραδείγματα: ύψος πωλήσεων, αξίας ενεργητικού, αριθμός μετοχών, κ.λπ. Περαιτέρω διάκριση:
 - **Διακριτές (discrete) ή ασυνεχείς (discontinuous).**
 - **Συνεχείς (continuous).**

Ποιοτικές Μεταβλητές: Κατηγορικές

- **Κατηγορικές (categorical ή nominal):** δεν αντιστοιχούν σε μετρήσιμα μεγέθη, αλλά απλά κατηγοριοποιούν τα στοιχεία ενός πληθυσμού σε ομάδες σαφώς διαφοροποιημένες μεταξύ τους.
 - Στις κατηγορικές μεταβλητές, οι επιμέρους κατηγορίες (ή ομάδες) που ορίζονται, δεν εμπερικλείουν την έννοια της διάταξης.
 - Παραδείγματα: εισηγμένη έναντι μη εισηγμένης, κλάδος δραστηριότητας (πρωτογενής τομέας, εμπόριο, βιομηχανία, τραπεζικός τομέας).
 - Όταν μία κατηγορική μεταβλητή λαμβάνει μόνο δύο τιμές τότε καλείται **δίτιμη (binary)** ή **διχοτομική (dichotomous)**. Οι τιμές που λαμβάνει μία δίτιμη ή διχοτομική είναι 0 (=απουσία ιδιότητας ή χαρακτηριστικού) ή 1 (=παρουσία ιδιότητας ή χαρακτηριστικού).
 - Η χρήση της αριθμητική κωδικοποίησης (π.χ. πρωτογενής τομέας =1, βιομηχανία = 2, τραπεζικός τομέας =3, κ.ο.κ.) εξυπηρετεί μόνο την ταυτοποίηση των διαφόρων κατηγοριών. Ως εκ τούτου δεν έχει λογικό νόημα (με εξαίρεση τις δίτιμες ή διχοτομικές).
 - **Εξαίρεση** αποτελεί η κωδικοποίηση 0 και 1 για τις δίτιμες μεταβλητές. Σε μια τέτοια περίπτωση, το άθροισμα των αριθμητικών τιμών της μεταβλητής, ορίζει τον αριθμό των παρατηρήσεων που έχουν κατηγοριοποιηθεί με τον αριθμό 1, ενώ ο αριθμητικός μέσος δίνει την αναλογία των παρατηρήσεων που έχουν την τιμή 1 στο σύνολο των παρατηρήσεων.

Ποιοτικές Μεταβλητές: Διατεταγμένες

- **Διατεταγμένες (ordinal variables):** οι κατηγορίες ορίζονται βάσει μιας σχέση διάταξης που υφίσταται μεταξύ τους.
 - Παράδειγμα: σε μια έρευνα αγοράς, η ικανοποίηση που εκφράζει ένας καταναλωτής ως προς μία τραπεζική υπηρεσία, μπορεί να δοθεί με μία σειρά απαντήσεων του είδους: ‘πολύ ικανοποιημένος’, ‘ικανοποιημένος’, ‘ουδέτερος’, ‘δυσανεστημένος’, ‘πολύ δυσανεστημένος’.
- Αυτός ο τρόπος διαφοροποίησης των απαντήσεων, ουσιαστικά κατηγοριοποιεί τα άτομα σε πέντε ομάδες, διατεταγμένες ως προς το βαθμό ικανοποίησής τους **ΚΑΙ** η διαφορά (ή απόσταση) του βαθμού ικανοποίησης από τη μία ομάδα στην άλλη, δεν μπορεί να θεωρηθεί ότι είναι ίδια μεταξύ όλων των ομάδων.
- Λόγω των διαφορετικών αποστάσεων που υφίστανται μεταξύ των βαθμίδων μιας διατεταγμένης μεταβλητής, η χρήση αριθμητικής κωδικοποίησης σε αυτές (π.χ. στο προηγούμενο παράδειγμα η κωδικοποίηση από 0 = πολύ δυσανεστημένος μέχρι 4 = πολύ ικανοποιημένος) δεν επιτρέπει κατά κανόνα τον ορισμό αριθμητικών πράξεων επ’ αυτών.

Ποσοτικές Μεταβλητές: Διακριτές ή Ασυνεχείς

- **Διακριτές (discrete) ή ασυνεχείς (discontinuous):** παίρνουν τιμές πεπερασμένου πλήθους, συνήθως ακέραιες, χωρίς να έχουν τη δυνατότητα να πάρουν μεταξύ αυτών των τιμών άλλες ενδιάμεσες.
- Η αριθμητική έκφραση αυτών των μεταβλητών απορρέει ευθέως από την τιμή του μεγέθους στο οποίο αναφέρονται. Η πλέον συνήθης περίπτωση διακριτών μεταβλητών είναι αυτές που απαριθμούν τα στοιχεία ενός συνόλου.
- Στις διακριτές μεταβλητές ισχύει η σχέση της διάταξης των επιμέρους τιμών τους, ενώ επιπλέον είναι αριθμητικά συγκρίσιμες οι διαφορές μεταξύ αυτών των τιμών.
 - Για παράδειγμα η διαφορά μεγέθους δύο οικογενειών με τρία και τέσσερα μέλη είναι ίση με τη διαφορά μεγέθους δύο οικογενειών με πέντε και έξι μέλη.
- Λόγω της δυνατότητας σύγκρισης των διαφορών των επιμέρους τιμών μίας διακριτής μεταβλητής, οποιαδήποτε αριθμητική πράξη έχει νόημα να ορισθεί επ' αυτών.

Ποσοτικές Μεταβλητές: Συνεχείς

- **Συνεχείς (continuous):** μπορούν να πάρουν οποιαδήποτε τιμή σε όλο το εύρος των πραγματικών αριθμών, ενώ η διαφορά μεταξύ δύο δυνατών τιμών τους μπορεί να είναι απεριόριστα μικρή.
- Παραδείγματα: ύψος πωλήσεων, αξίας ενεργητικού, αριθμός μετοχών, κ.λπ.
- Ο μοναδικός περιοριστικός παράγοντας για τις δυνατές τιμές μιας συνεχούς μεταβλητής, είναι η ακρίβεια της μέτρησης. Θεωρητικά, όσο πιο ακριβές είναι το όργανο με το οποίο μετράται μια συνεχής μεταβλητή, τόσο περισσότερες είναι οι τιμές που αυτή μπορεί να πάρει.
- Στις συνεχείς μεταβλητές ορίζεται σχέση διατάξεως μεταξύ των επιμέρους τιμών τους, ενώ και οι αποστάσεις μεταξύ αυτών των τιμών είναι συγκρίσιμες από αριθμητικής απόψεως. Περαιτέρω διάκριση:
 - **Μεταβλητές αναλογίας (ratio scale).**
 - **Μεταβλητές διαστήματος (interval scale).**



Περιγραφική Στατιστική

- Η περιγραφική στατιστική στοχεύει στη σύνοψη και την εμπειριστατωμένη περιγραφή αριθμητικών δεδομένων, με απώτερο σκοπό την απλούστερη παρουσίαση και την ευκολότερη κατανόηση τους.
- Τα δεδομένα αυτά μπορεί να προέρχονται είτε από το πλήρες σύνολο των στοιχείων ενός πληθυσμού είτε από ένα δείγμα αυτού.
- Η σύνοψη και η περιγραφή αριθμητικών δεδομένων στην περιγραφική στατιστική γίνεται με τη χρήση των:
 - Πινάκων συχνοτήτων.
 - Διαγραμμάτων.
 - Περιγραφικών στατιστικών μέτρων.

Περιγραφική Στατιστική

- Οι πίνακες συχνοτήτων χρησιμοποιούνται για την παρουσίαση κατανομών συχνοτήτων μεταβλητών όλων των τύπων.
- Αν πρόκειται για κατανομές κατηγορικών ή διατεταγμένων μεταβλητών, οι πίνακες αυτοί αποτελούνται από το σύνολο των επιμέρους κατηγοριών ή τάξεων που περιλαμβάνει η μεταβλητή, μαζί με τον αριθμό των παρατηρήσεων (ή ατόμων) που αντιστοιχούν σε κάθε κατηγορία ή τάξη.
- Αν η μεταβλητή είναι διακριτή με περιορισμένο πλήθος δυνατών τιμών, τότε η κατανομή συχνότητας, δίδεται με τρόπο αντίστοιχο αυτού των κατανομών των κατηγορικών ή διατεταγμένων μεταβλητών. Δηλαδή για κάθε τιμή της μεταβλητής, αναφέρεται ο αριθμός των παρατηρήσεων που αντιστοιχούν σε αυτή.



Χρονιές απογραφής	Ανδρες	Γυναίκες
1870	754.186	703.718
1879	880.952	798.518
1889	1.133.625	1.053.583
1896	1.266.816	1.166.990
1907	1.324.942	1.307.010
1920	2.495.316	2.521.573
1928	3.076.235	3.128.449
1940	3.658.393	3.686.467
1951	3.721.648	3.911.153
1961	4.091.894	4.296.659
1971	4.286.748	4.481.624
1981	4.779.571	4.960.018
1991	5.055.408	5.204.492
2001	5.424.089	5.515.516
2011	5.303.223	5.513.063



Μέγεθος νοικοκυριού	Αριθμός νοικοκυριών	Μέγεθος νοικοκυριού	Αριθμός νοικοκυριών
1 μέλος	1.061.547	6 μέλη	68.602
2 μέλη	1.218.466	7 μέλη	20.273
3 μέλη	817.921	8 μέλη	7.511
4 μέλη	726.554	9 μέλη	1.881
5 μέλη	209.569	10 και άνω μέλη	2.216

Περιγραφική Στατιστική

- Στην περίπτωση μιας συνεχούς μεταβλητής, η αναλυτική αναφορά όλων των τιμών της, δεν εξυπηρετεί ούτε για την παρουσίαση των δεδομένων, αλλά ούτε και για την εξαγωγή συμπερασμάτων. Σε αυτού του είδους τις μεταβλητές, είναι επιβεβλημένη η σύμπτυξη των τιμών τους, σε διαστήματα σαφώς διαφοροποιημένα και μη επικαλυπτόμενα μεταξύ τους.
- Σε ορισμένες περιπτώσεις, στους πίνακες συχνότητων, είναι αναγκαίο μαζί με τις απόλυτες συχνότητες (δηλαδή με τον αριθμό των παρατηρήσεων) μιας μεταβλητής, να αναφέρονται και οι αντίστοιχες σχετικές συχνότητες.
 - Όταν λέμε σχετική συχνότητα μιας μεταβλητής, εννοούμε το ποσοστό (%) των παρατηρήσεων που αντιστοιχεί σε κάθε κατηγορία ή διάστημα τιμών της μεταβλητής.



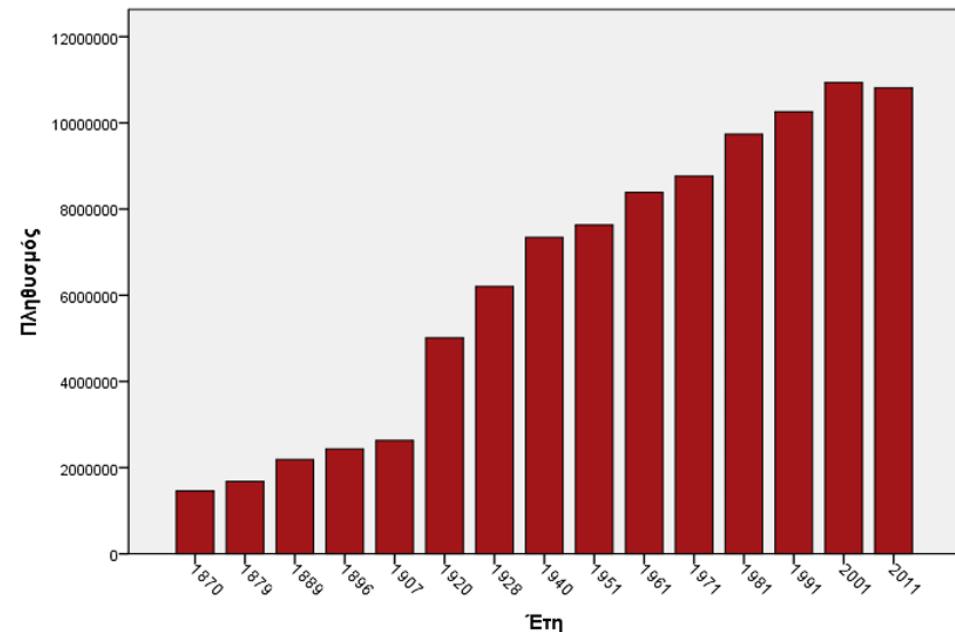
Βάρος σε Kg	Αριθμός ατόμων
40 - 49,9	12
50 - 59,9	102
60 - 69,9	233
70 - 79,9	265
80 - 89,9	176
90 - 99,9	71
100 - 109,9	27
110 - 119,9	9
Σύνολο	895



Βάρος σε Kg	Αριθμός ατόμων	Σχετική συχνότητα (%)
40 - 49,9	12	1,3
50 - 59,9	102	11,4
60 - 69,9	233	26,0
70 - 79,9	265	29,7
80 - 89,9	176	19,7
90 - 99,9	71	7,9
100 - 109,9	27	3,0
110 - 119,9	9	1,0
Σύνολο	895	100,0

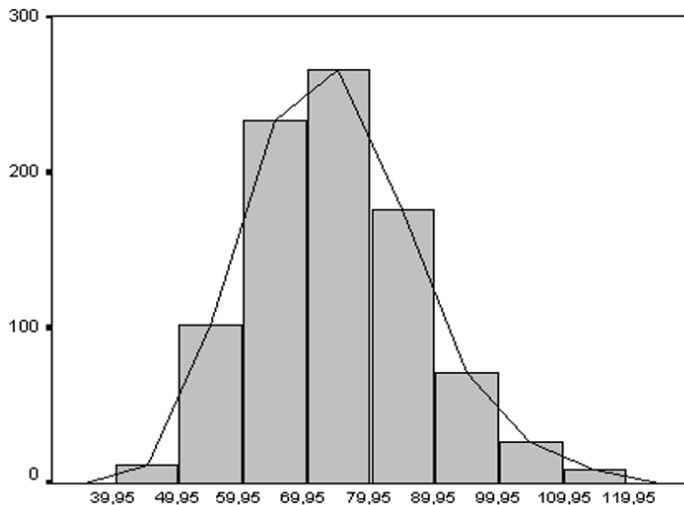
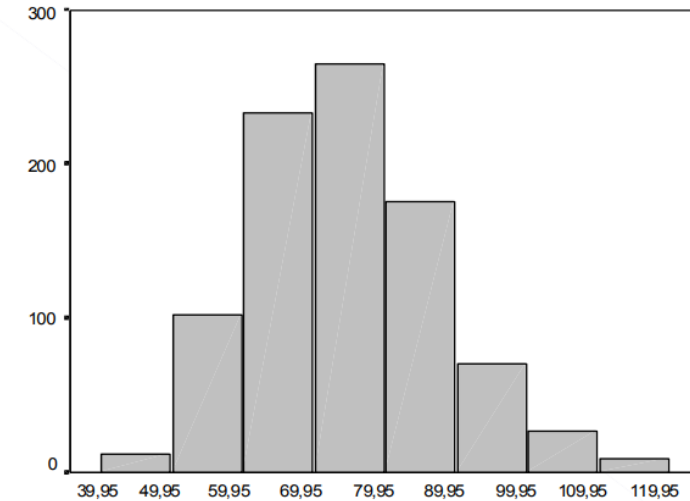
Διαγράμματα

- Τα διαγράμματα αποτελούν γραφικές απεικονίσεις των δεδομένων με συστηματικό τρόπο που επιτρέπουν την ευκολότερη αντίληψη της συμπεριφοράς των δεδομένων (σε σχέση με τους πίνακες) αλλά υστερούν όμως έναντι αυτών, ως προς το βαθμό λεπτομέρειας που διασφαλίζουν κατά την παρουσίαση των δεδομένων.
- Τα **ραβδογράμματα (bar charts)** είναι τύποι διαγραμμάτων, που χρησιμοποιούνται για την απεικόνιση κατανομών συχνοτήτων κατηγορικών και διατεταγμένων μεταβλητών.



Διαγράμματα

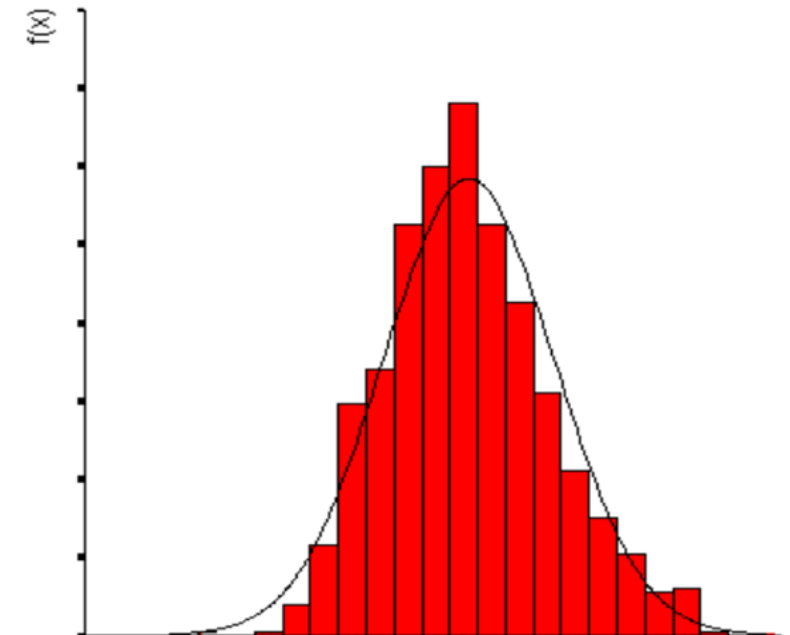
- Τα **ιστογράμματα (histograms)** είναι κατασκευές αντίστοιχες των ραβδογραμμάτων, μόνο που η χρήση τους επιβάλλεται σε περιπτώσεις κατανομών συχνοτήτων ποσοτικών μεταβλητών, διακριτών ή συνεχών. Σε αυτού του είδους τις μεταβλητές, όπως ήδη αναφέρθηκε, απαιτείται η σύμπτυξη των τιμών τους σε διαστήματα.



- Το **πολύγωνο συχνοτήτων (frequency polygon)** είναι ένα παράγωγο διάγραμμα που απορρέει από την κατασκευή ενός ιστογράμματος. Κατασκευάζεται, αν σε ένα ιστογράμμα ενώσουμε τα κέντρα των κορυφών των στηλών του. Η πολυγωνική γραμμή που θα προκύψει, ορίζει το αντίστοιχο πολύγωνο συχνοτήτων.

Διαγράμματα

- Σε ένα ιστόγραμμα όταν ο αριθμός των παρατηρήσεων αυξάνει απεριόριστα και το εύρος των διαστημάτων ελαττώνεται, τότε το ιστόγραμμα και το πολύγωνο συχνοτήτων, τείνουν να συμπέσουν σε μία συνεχή καμπύλη η οποία ονομάζεται **καμπύλη συχνοτήτων (frequency curve)**. Τόσο το ιστόγραμμα όσο και το πολύγωνο αλλά και η καμπύλη των συχνοτήτων, ουσιαστικά μεταφέρουν την ίδια πληροφορία για την κατανομή της μεταβλητής στην οποία αναφέρονται.



Στατιστικά Μέτρα Κεντρικής Τάσης

- **Κεντρική τάση (central tendency)** μιας κατανομής είναι η τάση που εμφανίζουν οι τιμές της κατανομής να συσσωρεύονται γύρω από κάποιο κεντρικό σημείο της. Τα μέτρα κεντρικής τάσης στοχεύουν στον προσδιορισμό αυτής της τάσης:
- **Μέση τιμή (mean):** είναι το πηλίκο του αθροίσματος των μετρήσεων διαιρούμενο δια του πλήθους τους.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

- Δεν έχει νόημα η χρησιμοποίηση της μέσης τιμής σε περιπτώσεις κατηγορικών μεταβλητών, ενώ σε περιπτώσεις διατεταγμένων μεταβλητών μόνο ενδεικτικά μπορεί να χρησιμοποιείται.
 - Εξαίρεση αποτελεί η χρήση της σε δίτιμες μεταβλητές, εφόσον η κωδικοποίηση των δύο δυνατών τιμών της δίτιμης μεταβλητής είναι 0 και 1. Ο αριθμητικός μέσος δίνει την αναλογία των παρατηρήσεων που έχουν την τιμή 1 στο σύνολο των παρατηρήσεων.

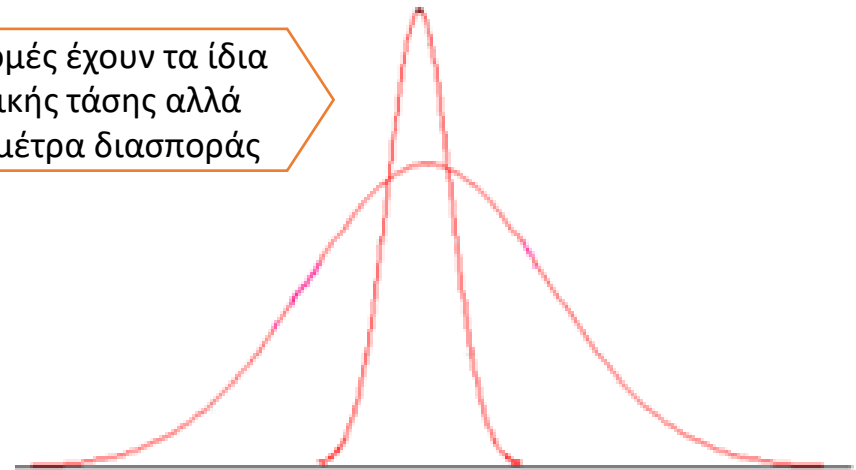
Στατιστικά Μέτρα Κεντρικής Τάσης

- **Διάμεσος (median):** είναι η τιμή η οποία, όταν οι μετρήσεις διαταχθούν κατά αύξουσα σειρά, βρίσκεται ακριβώς στο μέσον τους, έχει δηλαδή από αριστερά της το 50% του συνόλου των μετρήσεων και από δεξιά της το υπόλοιπο 50%. Είναι η τιμή που καταλαμβάνει τη $(n+1)/2$ θέση στη διάταξη των n τιμών.
 - Κατά τον υπολογισμό της διαμέσου λαμβάνονται υπόψη μόνο οι σχετικές θέσεις των τιμών και όχι οι τιμές αυτές καθ' αυτές. Για το λόγο αυτό, η διάμεσος ελάχιστα επηρεάζεται από τη ύπαρξη ακραίων τιμών στο σύνολο των μετρήσεων και χαρακτηρίζεται ως ανθεκτικό (robust) μέτρο κεντρικής τάσης.
 - Η διάμεσος, μπορεί να υπολογιστεί τόσο σε ποσοτικές μεταβλητές (διακριτές ή συνεχείς) όσο και σε διατεταγμένες.
- **Επικρατούσα τιμή (mode):** είναι εκείνη η τιμή η οποία έχει τη μεγαλύτερη συχνότητα εμφάνισης.
 - Η επικρατούσα τιμή μπορεί να υπολογιστεί σε μεταβλητές όλων των τύπων: κατηγορικές, διατεταγμένες, ποσοτικές. Όπως προκύπτει από τον ορισμό, η επικρατούσα τιμή ενός συνόλου μετρήσεων μπορεί να μην είναι μοναδική.

Στατιστικά Μέτρα Διασποράς

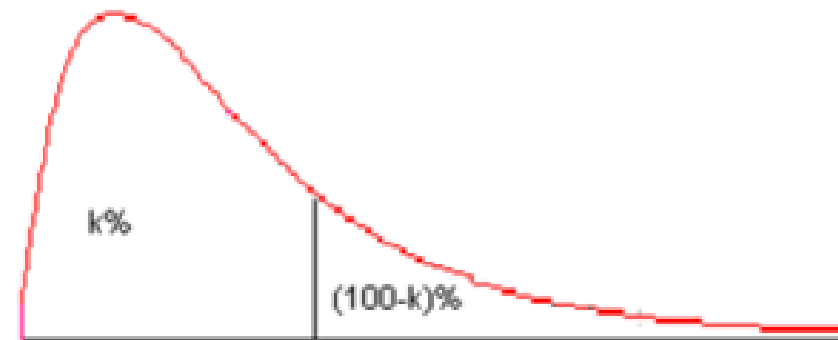
- Τα μέτρα διασποράς στοχεύουν στο προσδιορισμό της μεταβλητότητας (ή ετερογένειας) που παρουσιάζει ένα σύνολο μετρήσεων. Τα μέτρα αυτά χρησιμοποιούνται σε συνδυασμό με τα μέτρα θέσης και από κοινού περιγράφουν τις κατανομές δεδομένων με τρόπο συμπληρωματικό.
- Ενδεικτικά μέτρα διασποράς:
 - Εύρος.
 - Εκατοστημόρια.
 - Ενδοτερταμηνιακό εύρος.
 - Μέση απόκλιση.
 - Διακύμανση.
 - Τυπική απόκλιση.
 - Συντελεστής μεταβλητότητας.

Οι δύο κατανομές έχουν τα ίδια μέτρα κεντρικής τάσης αλλά διαφορετικά μέτρα διασποράς



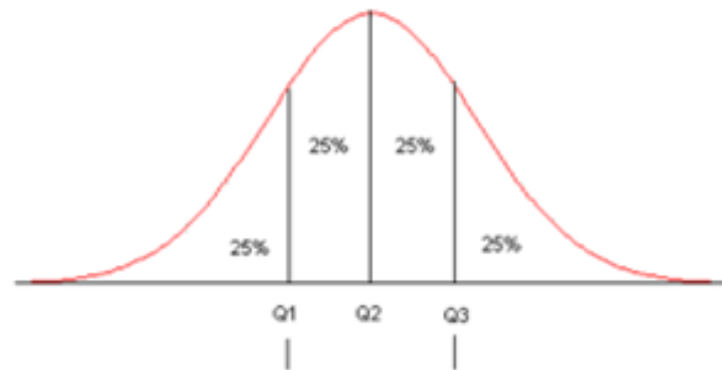
Στατιστικά Μέτρα Διασποράς

- **Εύρος (range):** είναι το απλούστερο από όλα τα μέτρα διασποράς και ορίζεται ως η διαφορά μεταξύ μέγιστης και ελάχιστης τιμής ενός συνόλου μετρήσεων.
 - Εύρος $r = \max - \min$
 - Εύκολο στον προσδιορισμό του αλλά η χρηστικότητά του είναι εξαιρετικά περιορισμένη διότι στον υπολογισμό του υπεισέρχονται δύο μόνο τιμές, οι πλέον ακραίες και είναι εξαιρετικά ευαίσθητο στην ύπαρξη τιμών που διαφοροποιούνται πολύ των υπολοίπων.
- Τα **εκατοστημόρια (percentiles)** αποτελούν γενίκευση της έννοιας της διαμέσου. Το k -εκατοστημόριο ενός συνόλου μετρήσεων (n) είναι εκείνη η τιμή, η οποία, όταν οι τιμές διαταχθούν κατ' αύξουσα σειρά, έχει από αριστερά της το $k\%$ του συνόλου των μετρήσεων και από δεξιά της το υπόλοιπο $(100-k)\%$
- Το k -εκατοστημόριο είναι ο αριθμός που στη διάταξη των τιμών καταλαμβάνει $(n + 1)/k$ τη θέση.



Στατιστικά Μέτρα Διασποράς

- Το 25^ο, το 50^ο και το 75^ο εκατοστημόριο μιας κατανομής μετρήσεων ονομάζονται πρώτο, δεύτερο και τρίτο τεταρτημόριο αντίστοιχα και συμβολίζονται Q1, Q2, Q3. Η διαφορά τρίτου και πρώτου τεταρτημορίου Q3 - Q1 ονομάζεται **ενδοτεταρτημοριακό εύρος (interquartile range)**.



- Μέση απόκλιση: η μέση τιμή των αποστάσεων των τιμών από τη μέση τιμή τους:

$$\frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$$

Στατιστικά Μέτρα Διασποράς

- Διακύμανση (variance):

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

- Τυπική απόκλιση (standard deviation):

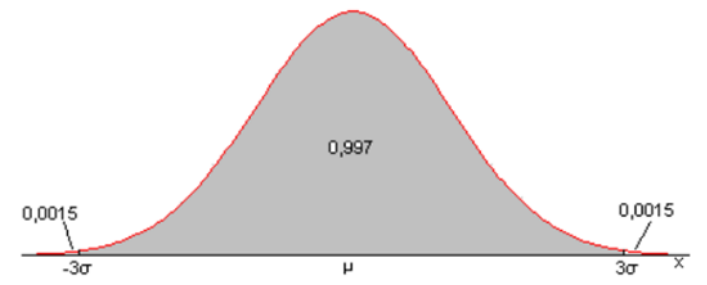
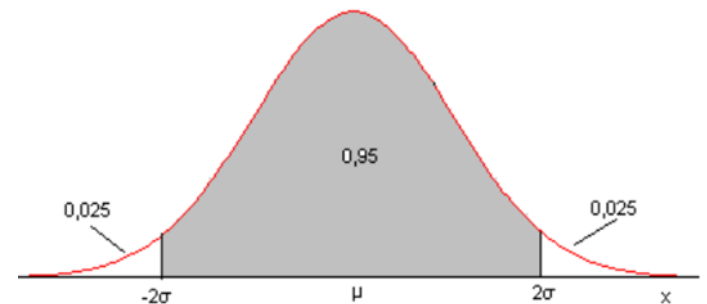
$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

- Συντελεστής μεταβλητότητας (coefficient of variation): είναι ένα σχετικό μέτρο διασποράς και εκφράζει την τυπική απόκλιση ενός συνόλου μετρήσεων ως ποσοστό (%) επί της μέσης τιμής τους.

$$CV = \frac{s}{\bar{x}} \cdot 100$$

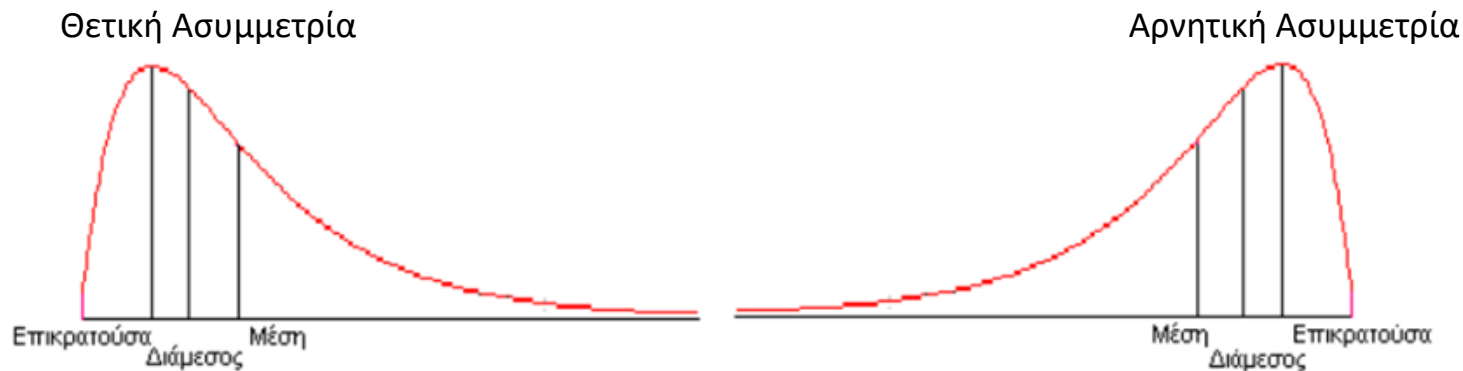
Στατιστικά Μέτρα Διασποράς

- Σε κωδωνοειδείς συμμετρικές κατανομές ορίζονται τα παρακάτω διαστήματα τιμών στα οποία ανήκει κάθε φορά ένα συγκεκριμένο ποσοστό του συνόλου των τιμών της (κανονικής) κατανομής:
 - Μέσος όρος (μ) $\pm \sigma$: 68%
 - Μέσος όρος (μ) $\pm 2\sigma$: 95%
 - Μέσος όρος (μ) $\pm 3\sigma$: 99%



Μέτρα Ασυμμετρίας

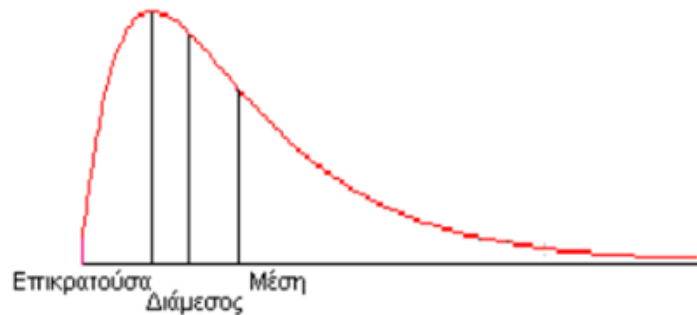
- Η **ασυμμετρία (skewness)** μιας κατανομής έχει να κάνει με την εκτροπή της από την κανονικότητα.
- Σε μια κανονική κατανομή η διάταξη των τιμών της γύρω από το μέσο είναι συμμετρική, και όπως ήδη έχουμε αναφέρει, οι τρεις τιμές θέσεως της κατανομής (μέση, διάμεσος και επικρατούσα) συμπίπτουν.
- Αντίθετα σε περιπτώσεις μη συμμετρικών κατανομών, η διάταξη των τιμών περί το μέσο είναι ασύμμετρη και οι τιμές θέσεως διαφοροποιούνται η μια της άλλης.
- Στις ασύμμετρες κατανομές η εκτροπή μπορεί να εμφανίζεται είτε από τη δεξιά πλευρά τους, να έχουμε δηλαδή παρατεταμένη ανάπτυξη του δεξιού κλάδου της κατανομής (θετική ασυμμετρία), είτε από την αριστερή πλευρά με εκτεταμένη ανάπτυξη του αριστερού κλάδου (αρνητική ασυμμετρία).



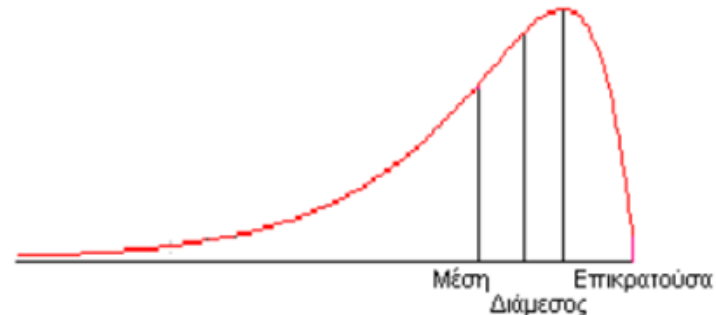
Μέτρα Ασυμμετρίας και Μέτρα Κεντρικής Τάσης

- Οι σχετικές θέσεις των τιμών κεντρικής τάσης μιας κατανομής μπορούν να χρησιμοποιηθούν για τον προσδιορισμό της ασυμμετρίας μιας κατανομής . Ειδικότερα ισχύουν τα ακόλουθα:
- Μέση = Διάμεσος = Επικρατούσα η κατανομή είναι συμμετρική.
- Μέση > Διάμεσος > Επικρατούσα η κατανομή είναι θετικά ασύμμετρη.
- Μέση < Διάμεσος < Επικρατούσα η κατανομή είναι αρνητικά ασύμμετρη.

Θετική Ασυμμετρία

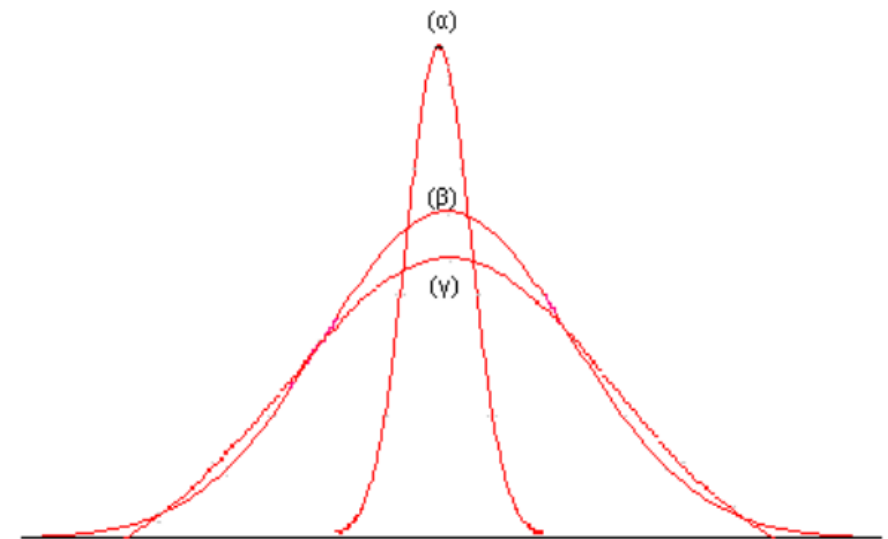


Αρνητική Ασυμμετρία



Μέτρα Ασυμμετρίας

- Η **κύρτωση (kurtosis)** μιας κατανομής έχει να κάνει με το βαθμό συγκέντρωσης των τιμών της κατανομής περί το μέσο της. Με το ποσοστό δηλαδή των τιμών της που βρίσκονται στο κεντρικό διάστημα του εύρους διακύμανσής της.
- Ο προσδιορισμός της κύρτωσης γίνεται, όπως και στην ασυμμετρία, σε σχέση με την κανονική κατανομή. (περίπτωση β)
- Αν δηλαδή το ποσοστό των παρατηρήσεων της κατανομής που βρίσκονται στο κέντρο της, είναι μεγαλύτερο του αντίστοιχου της κανονικής κατανομής, η κύρτωση της κατανομής είναι θετική και η κατανομή χαρακτηρίζεται ως **λεπτόκυρτη**. (περίπτωση α)
- Σε αντίθετη περίπτωση, η κύρτωση της κατανομής είναι αρνητική και η κατανομή χαρακτηρίζεται ως **πλατύκυρτη**. (περίπτωση γ)



Θηκογράμματα (box plots)

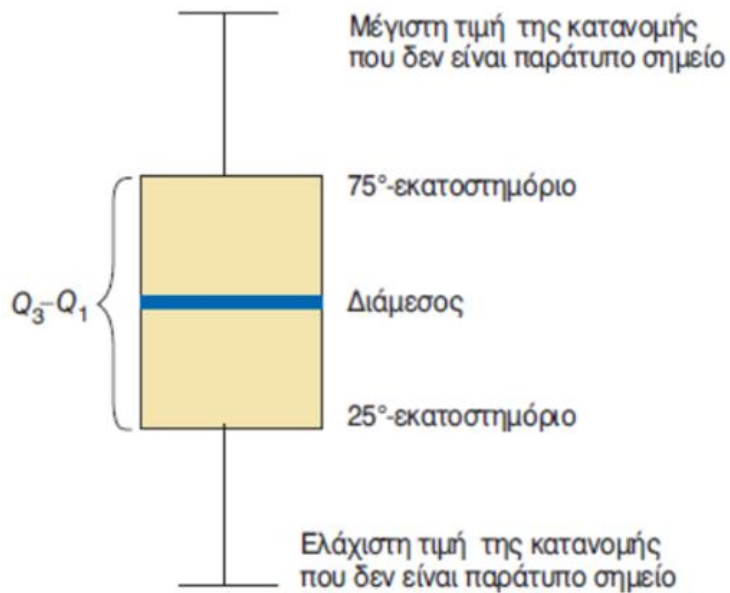
- Τα **Θηκογράμματα (box plots)** είναι διαγραμματικές απεικονίσεις οι οποίες συνοψίζουν υπό μορφή γραφήματος, βασικά περιγραφικά μέτρα μιας κατανομής, όπως η διάμεσος, τα τεταρτημόρια το ενδοτεταρτημοριακό εύρος και οι ακραίες τιμές.
- Ένα θηκόγραμμα είναι πιο συμπαγές από ότι ένα ιστόγραμμα σε σχέση με την πληροφορία που εμπειρικλείει, αλλά είναι λιγότερο λεπτομερές. Επιπλέον ένα θηκόγραμμα, μπορεί να χρησιμοποιηθεί για την ταυτόχρονη απεικόνιση και σύγκριση δύο ή περισσότερων κατανομών.
- Η μορφή του διαγράμματος είναι ένα ορθογώνιο παραλληλόγραμμο το ύψος του οποίου αντιστοιχεί στο ενδοτεταρτημοριακό εύρος της κατανομής. Η κάτω οριζόντια πλευρά του παραλληλογράμμου αντιστοιχεί στο 25^ο εκατοστημόριο της κατανομής ενώ η επάνω οριζόντια πλευρά στο 75^ο εκατοστημόριο. Στο εσωτερικό του υπάρχει μια οριζόντια γραμμή η οποία αντιστοιχεί στη διάμεσο της κατανομής.

Θηκογράμματα (box plots)

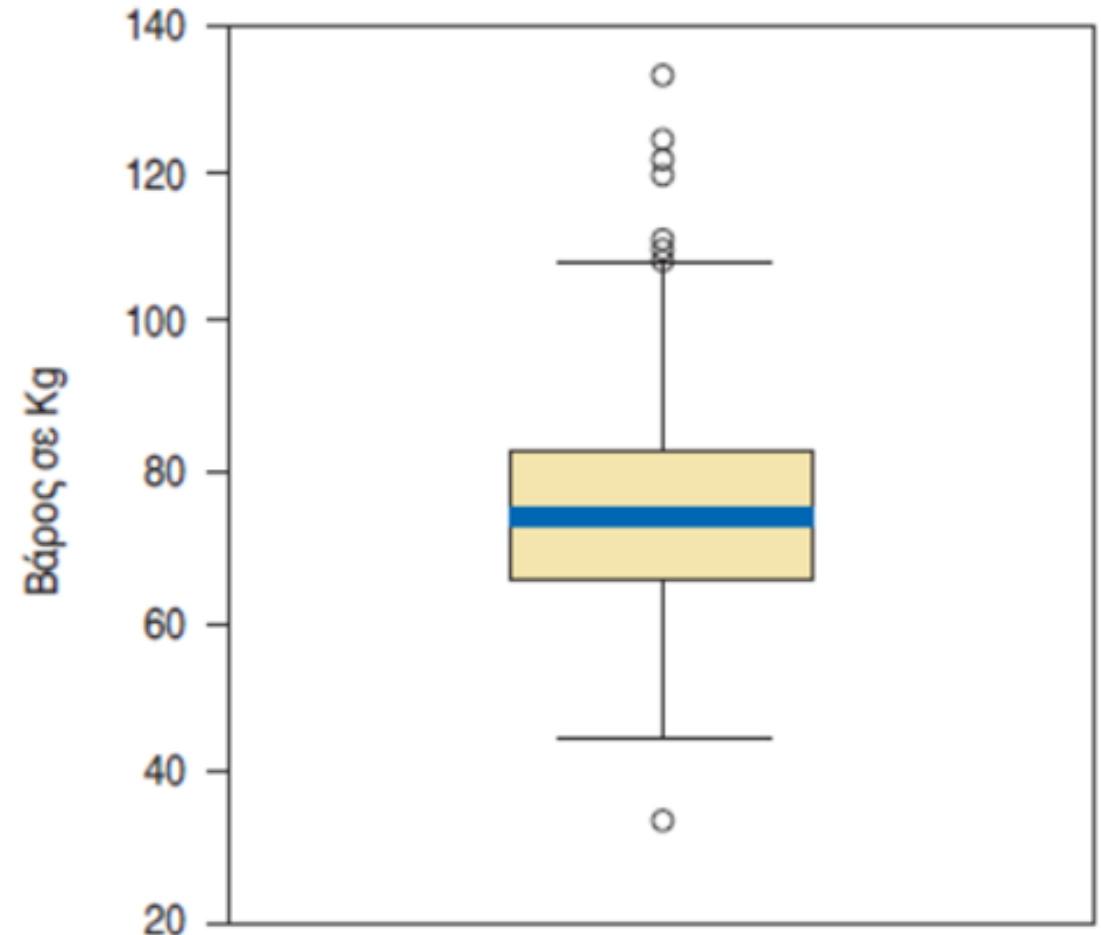
- Οριζόντιες γραμμές οι οποίες ονομάζονται **φράκτες (whiskers)** φέρονται πέραν των δύο οριζοντίων πλευρών του παραλληλογράμμου σε αποστάσεις ίσες το πολύ με 1,5 φορά το ενδοτεταρτημοριακό εύρος της κατανομής, $1,5 \times (Q3 - Q1)$. Αν η μικρότερη ή η μεγαλύτερη τιμή της κατανομής βρίσκονται εντός των περιοχών αυτών των αποστάσεων, τότε οι φράκτες φέρονται ακριβώς στο ύψος αυτών των τιμών.
- Τιμές της κατανομής που βρίσκονται εκτός των περιοχών που ορίζονται από τους φράκτες ονομάζονται **παράτυπα σημεία της κατανομής (outliers)**.
- Αν τα παράτυπα σημεία βρίσκονται σε απόσταση μικρότερη από 3 φορές το ενδοτεταρτημοριακό εύρος, δηλαδή μεταξύ $1,5 \times (Q3 - Q1)$ και $3 \times (Q3 - Q1)$, συμβολίζονται επί του θηκογράμματος με ένα μικρό κύκλο (o), διαφορετικά αν βρίσκονται σε απόσταση μεγαλύτερη από $3 \times (Q3 - Q1)$, συμβολίζονται με ένα αστερίσκο (*).

Θηκογράμματα (box plots)

- * Τιμές μεγαλύτερες κατά $3(Q_3 - Q_1)$ τουλάχιστον από το 75°-εκατοστημόριο
- Τιμές μεγαλύτερες κατά $1,5(Q_3 - Q_1)$ τουλάχιστον από το 75°-εκατοστημόριο



- Τιμές μικρότερες κατά $1,5(Q_3 - Q_1)$ τουλάχιστον από το 25°-εκατοστημόριο
- * Τιμές μικρότερες κατά $3(Q_3 - Q_1)$ τουλάχιστον από το 25°-εκατοστημόριο



Στατιστικός Έλεγχος

- Μία συνήθης περίπτωση στατιστικού ελέγχου αφορά την μέση τιμή (μ) μία μεταβλητής και ειδικότερα αν αυτή είναι ίση, μεγαλύτερη ή μικρότερη από μία συγκεκριμένη τιμή (μ_0). Παράδειγμα: η μέση αξία του ενεργητικού των ελληνικών μη εισηγμένων εταιρειών τη χρήση 2023 ήταν 5.000.000 ευρώ.
- Αν μας ενδιαφέρει να εξετάσουμε ότι η μέση τιμή (μ) μία μεταβλητής είναι ίση ή (εναλλακτικά) διάφορη μία συγκεκριμένης τιμής (αμφίπλευρος έλεγχος – two tailed) τότε διατυπώνεται ως εξής:

$$H_0: \mu = \mu_0 \text{ έναντι } \mu \neq \mu_0$$

- Για τον στατιστικό έλεγχο χρησιμοποιείται η κατανομή t-student και ακολουθούνται τα ακόλουθα βήματα:
- **Βήμα 1^ο:** Καθορίζεται το ανεκτό επίπεδο πιθανότητας (α) να διαπράξουμε Σφάλμα Τύπου I (στατιστική σημαντικότητα).
- **Βήμα 2^ο:** Για το ανεκτό επίπεδο πιθανότητας (α) να διαπράξουμε Σφάλμα Τύπου I καθορίζονται οι κριτικές τιμές της κατανομής t-student.

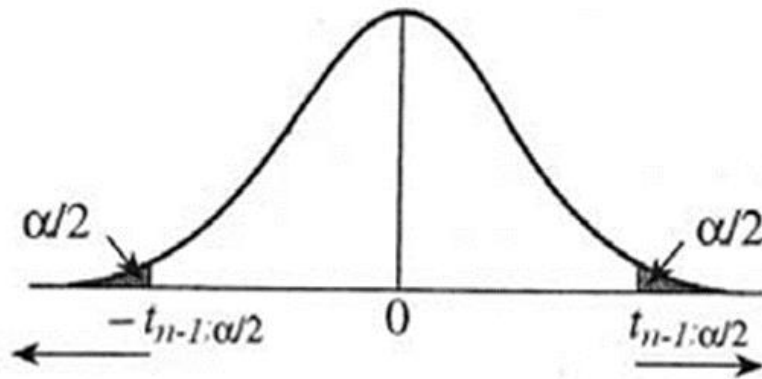
Στατιστικός Έλεγχος

- **Βήμα 3^ο:** Με βάση το δείγμα των δεδομένων υπολογίζεται ο μέσος όρος του δείγματος (\bar{x}).
- **Βήμα 4^ο:** Δεν αποδεχόμαστε την H_0 εάν ο μέσος όρος του δείγματος (\bar{x}) ικανοποιεί τα ακόλουθα κριτήρια:

$$\bar{x} \leq \mu_0 - \frac{s}{\sqrt{n}} t_{n-1; \alpha/2} \quad \text{ή} \quad \bar{x} \geq \mu_0 + \frac{s}{\sqrt{n}} t_{n-1; \alpha/2}$$

ή, ισοδύναμα, όταν,

$$|t| = \frac{|\bar{x} - \mu_0| \sqrt{n}}{s} \geq t_{n-1; \alpha/2}$$



Δεν αποδεχόμαστε την H_0 εάν ο μέσος όρος του δείγματος (\bar{x}) ευρίσκεται στις σκιασμένες περιοχές.

Τα πακέτα λογισμικού στατιστικής ανάλυσης παρέχουν και την τιμή p-value.

Δεν αποδεχόμαστε την H_0 εάν η τιμή p-value είναι μικρότερη από το ανεκτό επίπεδο πιθανότητας (α) να διαπράξουμε Σφάλμα Τύπου I (στατιστική σημαντικότητα).

Στατιστικός Έλεγχος

- Αν μας ενδιαφέρει να εξετάσουμε ότι η μέση τιμή (μ) μία μεταβλητής είναι ίση ή (εναλλακτικά) μεγαλύτερη μία συγκεκριμένης τιμής (μονόπλευρος δεξιόπλευρος έλεγχος – one tailed) τότε διατυπώνεται ως εξής:

$$H_0: \mu \leq \mu_0 \text{ έναντι } \mu > \mu_0$$

- Για τον στατιστικό έλεγχο χρησιμοποιείται η κατανομή t-student και ακολουθούνται τα ακόλουθα βήματα:
- **Βήμα 1^ο:** Καθορίζεται το ανεκτό επίπεδο πιθανότητας (α) να διαπράξουμε Σφάλμα Τύπου I (στατιστική σημαντικότητα).
- **Βήμα 2^ο:** Για το ανεκτό επίπεδο πιθανότητας (α) να διαπράξουμε Σφάλμα Τύπου I καθορίζονται η κριτική (δεξιόπλευρη) τιμή της κατανομής t-student.
- **Βήμα 3^ο:** Με βάση το δείγμα των δεδομένων υπολογίζεται ο μέσος όρος του δείγματος (\bar{x}).

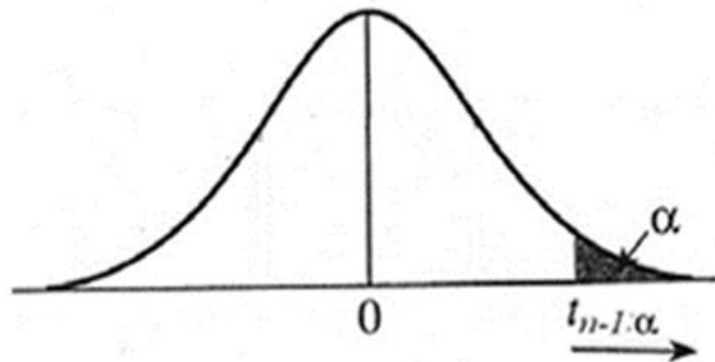
Στατιστικός Έλεγχος

- **Βήμα 3^ο:** Με βάση το δείγμα των δεδομένων υπολογίζεται ο μέσος όρος του δείγματος (\bar{x}).
- **Βήμα 4^ο:** Δεν αποδεχόμαστε την H_0 εάν ο μέσος όρος του δείγματος (\bar{x}) ικανοποιεί τα ακόλουθα κριτήρια:

$$\bar{x} \geq \mu_0 + \frac{s}{\sqrt{n}} t_{n-1;\alpha}$$

ή, ισοδύναμα, όταν,

$$t = \frac{(\bar{x} - \mu_0)\sqrt{n}}{s} \geq t_{n-1;\alpha}$$



Δεν αποδεχόμαστε την H_0 εάν ο μέσος όρος του δείγματος (\bar{x}) ευρίσκεται στις σκιασμένες περιοχές.

Τα πακέτα λογισμικού στατιστικής ανάλυσης παρέχουν και την τιμή p-value.

Δεν αποδεχόμαστε την H_0 εάν η τιμή p-value είναι μικρότερη από το ανεκτό επίπεδο πιθανότητας (α) να διαπράξουμε Σφάλμα Τύπου I (στατιστική σημαντικότητα).

Στατιστικός Έλεγχος

- Αν μας ενδιαφέρει να εξετάσουμε ότι η μέση τιμή (μ) μία μεταβλητής είναι ίση ή (εναλλακτικά) μεγαλύτερη μία συγκεκριμένης τιμής (μονόπλευρος αριστερόπλευρος έλεγχος – one tailed) τότε διατυπώνεται ως εξής:

$$H_0: \mu \geq \mu_0 \text{ έναντι } \mu < \mu_0$$

- Για τον στατιστικό έλεγχο χρησιμοποιείται η κατανομή t-student και ακολουθούνται τα ακόλουθα βήματα:
- **Βήμα 1^ο:** Καθορίζεται το ανεκτό επίπεδο πιθανότητας (α) να διαπράξουμε Σφάλμα Τύπου I (στατιστική σημαντικότητα).
- **Βήμα 2^ο:** Για το ανεκτό επίπεδο πιθανότητας (α) να διαπράξουμε Σφάλμα Τύπου I καθορίζονται η κριτική (αριστερόπλευρη) τιμή της κατανομής t-student.
- **Βήμα 3^ο:** Με βάση το δείγμα των δεδομένων υπολογίζεται ο μέσος όρος του δείγματος (\bar{x}).

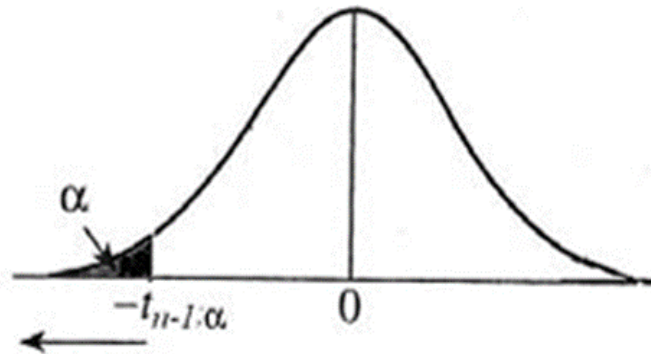
Στατιστικός Έλεγχος

- **Βήμα 3^ο:** Με βάση το δείγμα των δεδομένων υπολογίζεται ο μέσος όρος του δείγματος (\bar{x}).
- **Βήμα 4^ο:** Δεν αποδεχόμαστε την H_0 εάν ο ο μέσος όρος του δείγματος (\bar{x}) ικανοποιεί τα ακόλουθα κριτήρια:

$$\bar{x} \leq \mu_0 - \frac{s}{\sqrt{n}} t_{n-1; \alpha}$$

ή, ισοδύναμα, όταν,

$$t = \frac{(\bar{x} - \mu_0)\sqrt{n}}{s} \leq -t_{n-1; \alpha}$$



Δεν αποδεχόμαστε την H_0 εάν ο ο μέσος όρος του δείγματος (\bar{x}) ευρίσκεται στις σκιασμένες περιοχές.

Τα πακέτα λογισμικού στατιστικής ανάλυσης παρέχουν και την τιμή p-value.

Δεν αποδεχόμαστε την H_0 εάν η τιμή p-value είναι μικρότερη από το ανεκτό επίπεδο πιθανότητας (α) να διαπράξουμε Σφάλμα Τύπου Ι (στατιστική σημαντικότητα).

Περιγραφική Στατιστική και Στατιστικός Έλεγχος

Εφαρμογή 4^η

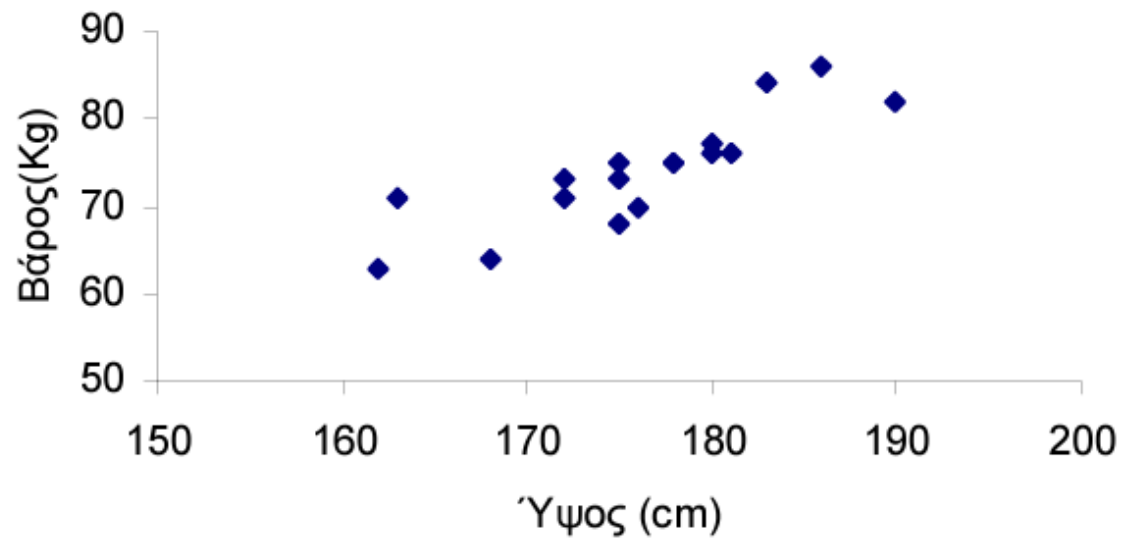


Συσχέτιση

- Η συσχέτιση εκφράζει πόσο ισχυρή (έντονη) είναι η σχέση ανάμεσα σε δύο μεταβλητές.
- Για παράδειγμα, μπορεί να υπάρχει ισχυρή συσχέτιση ανάμεσα στον οικονομικό κύκλο και το ύψος των πωλήσεων των εταιρειών.
- Αν οι δύο μεταβλητές κινούνται προς την ίδια κατεύθυνση, η συσχέτιση είναι θετική. Αν κινούνται προς αντίθετες κατευθύνσεις, τότε η συσχέτιση χαρακτηρίζεται αρνητική. Αν το αποτέλεσμα ισούται με μηδέν, τότε δεν υπάρχει σχέση μεταξύ τους.
- Η ανάλυση συσχέτισης μεταξύ δύο μεταβλητών αποτελεί το πρώτο βήμα για τη μελέτη της σχέσης τους.

Συσχέτιση

- Ένας απλός τρόπος για να αποκτήσουμε μια πρώτη ιδέα για το αν και πώς δυο μεταβλητές συµμεταβάλλονται-συσχετίζονται, είναι να κατασκευάσουμε το **διάγραμμα διασποράς (Scatter Diagram)**. Να αναπαραστήσουμε δηλαδή τα ζεύγη των παρατηρήσεων σε ένα διάγραμμα.



Μέτρα Γραμμική Συνάρτησης

- Ο δειγματικός συντελεστής γραμμικής συσχέτισης του Pearson δύο μεταβλητών X και Y συμβολίζεται με r και ορίζεται ως ο λόγος της συνδυακόμενης δύο μεταβλητών (s_{xy}) ως προς το γινόμενο των επιμέρους τυπικών αποκλίσεων των δυο επιμέρους μεταβλητών (s_x και s_y):

$$r = \frac{s_{xy}}{s_x \cdot s_y}$$

Αν $r = \pm 1$ υπάρχει **τέλεια γραμμική** συσχέτιση.

Αν $-0,3 \leq r < 0,3$ **δεν υπάρχει γραμμική** συσχέτιση. Αυτό, όμως, δεν σημαίνει ότι δεν υπάρχει άλλου είδους συσχέτιση μεταξύ των δύο μεταβλητών.

Αν $-0,5 < r \leq -0,3$ ή $0,3 \leq r < 0,5$ υπάρχει **ασθενής γραμμική** συσχέτιση.

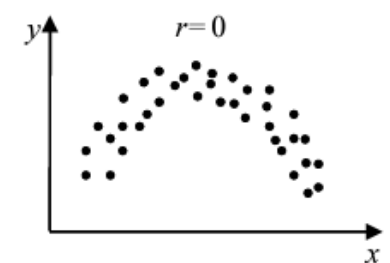
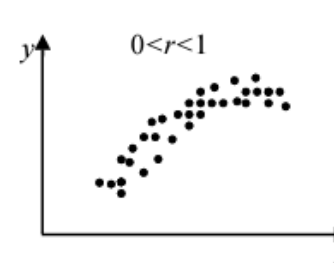
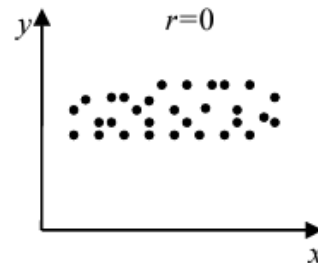
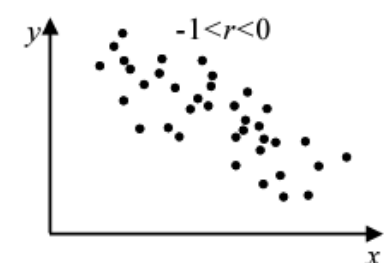
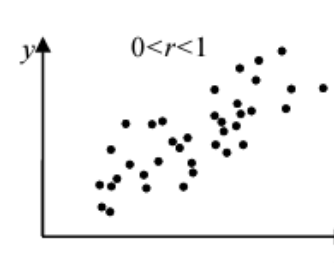
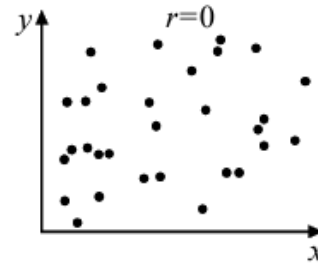
Αν $-0,7 < r \leq -0,5$ ή $0,5 \leq r < 0,7$ υπάρχει **μέση γραμμική** συσχέτιση.

Αν $-0,8 < r \leq -0,7$ ή $0,7 \leq r < 0,8$ υπάρχει **ισχυρή γραμμική** συσχέτιση.

Αν $-1 < r \leq -0,8$ ή $0,8 \leq r < 1$ υπάρχει **πολύ ισχυρή γραμμική** συσχέτιση.

Μέτρα Γραμμική Συνάρτησης

- Στην πράξη, υπολογίζουμε το συντελεστή γραμμικής συσχέτισης στις περιπτώσεις μόνο που το διάγραμμα διασποράς (στικτό διάγραμμα) έχει σχήμα επιμήκους κεκλιμένης έλλειψης ή πλατυσμένου J.
- Αν, όμως, τον υπολογίσουμε και σε περιπτώσεις που το διάγραμμα διασποράς έχει άλλη μορφή, η τιμή του η οποία θα είναι μικρή, δεν συνεπάγεται μη συσχέτιση αλλά μη γραμμική συσχέτιση. Είναι, δηλαδή, δυνατόν να υπάρχει μεγάλη μη γραμμική συσχέτιση.



Μέτρα Γραμμική Συνάρτησης

- **Συντελεστής συσχέτισης του Spearman (rho):** δίνει το μέγεθος της γραμμικής συσχέτισης ποιοτικών μεταβλητών διάταξης.
- Παίρνει τιμές στο κλειστό διάστημα $[-1, 1]$. Αν συμφωνούν πλήρως οι δύο κατατάξεις είναι $\rho = 1$, ενώ όταν η μια διάταξη είναι ριζικά διαφορετική από την άλλη (για παράδειγμα, αν το μέγεθος δείγματος είναι 8 τότε το X είναι 1 όταν το Y είναι 8, το X είναι 2 όταν το Y είναι 7, κ.ο.κ) είναι $\rho = -1$. Η τιμή 0 δείχνει το μικρότερο βαθμό συσχέτισης.

Στατιστικός Έλεγχος

- Η γενική ιδέα της διαδικασίας στατιστικού ελέγχου υποθέσεων είναι η εξής: θέτουμε ως μηδενική υπόθεση (H_0) αυτή για την οποία αμφιβάλουμε, αυτή που αμφισβητείται, και εξετάζουμε αν ένα τυχαίο δείγμα που παίρνουμε από τον πληθυσμό συνηγορεί υπέρ της απόρριψής της, έναντι της εναλλακτικής (H_1).
- Είμαστε σίγουροι ότι συμβαίνει σε ένα τυχαίο δείγμα αφορά τη συμπεριφορά του πληθυσμού; Όχι απαραίτητα. **ΑΛΛΑ** αν παρατηρηθεί κάτι ιδιαίτερα ακραίο ή σπάνιο (με άλλα λόγια έχει ιδιαίτερα μικρή πιθανότητα να συμβεί) τότε μπορούμε να απορρίψουμε την H_0 .
- **ΕΠΙΣΗΣ** , αν αυτό που παρατηρείται στο δείγμα δεν είναι ακραίο/σπάνιο τότε το δείγμα που πήραμε δε μας δίνει αρκετές ενδείξεις για την απόρριψη της H_0 και «αποτυγχάνουμε να την απορρίψουμε».

Στατιστικός Έλεγχος

- Ακόμη και έτσι, η παραπάνω συλλογιστική δεν μας προφυλάσσει απόλυτα από την εξαγωγή ενός εσφαλμένου συμπεράσματος. Κατά συνέπεια είναι δυνατόν να συμβούν οι ακόλουθοι δύο τύποι σφαλμάτων:
 - **Σφάλμα Τύπου I (type I error):** Απορρίφθηκε η H_0 ενώ ήταν αληθής. Με άλλα λόγια στο δείγμα που επιλέχθηκε παρατηρήθηκε κάτι ακραίο ή το πείραμα σχεδιάστηκε και εκτελέστηκε με τέτοιο τρόπο ώστε να μας οδηγήσει στην απόρριψη της H_0 ενώ ήταν αληθής.
 - **Σφάλμα Τύπου II (type II error):** Δεν απορρίφθηκε η H_0 ενώ ήταν ψευδής. Με άλλα λόγια στο δείγμα που επιλέχθηκε δεν παρατηρήθηκε κάτι ακραίο ή το πείραμα σχεδιάστηκε και εκτελέστηκε με τέτοιο τρόπο ώστε να μας οδηγήσει στη μη απόρριψη της H_0 ενώ ήταν ψευδής.

Ανάλυση Συσχέτισης

Εφαρμογή 5^η

