



# Ανάλυση Δεδομένων στη Λογιστική και Χρηματοοικονομική

## Ενότητα 1<sup>η</sup> Εισαγωγή

# Προτεινόμενα Βιβλία

- Χρήστου, Γ., 2011. Εισαγωγή στην Οικονομετρία, 2011, Εκδόσεις: Gutenberg.
- Τζαβαλής Η., 2008. Οικονομετρία, 1<sup>η</sup> Έκδοση, 2008, Εκδόσεις: ΟΠΑ.
- Griffiths, Hill and Lim, 2018. “Principles of Econometrics”, 5<sup>th</sup> Edition, Wiley.
- Brooks, C., 2019, Introductory Econometrics for Finance, 4th edition, Cambridge University Press.
- Brooks, C., 2022, Εισαγωγή στη Χρηματοοικονομική Οικονομετρία, Gutenberg

# Περιγραφή Ενότητας

- Εισαγωγή στην οικονομετρία
- Διαφορές μεταξύ πληθυσμού και δείγματος
- Διαφορές περιγραφικής και επαγωγικής στατιστικής
- Απλές και λογαριθμικές αποδόσεις (διαφορές)
- Σύνολα δεδομένων και μεταβλητές
- Κατηγορικά δεδομένα (categorical data), διατεταγμένα δεδομένα (ordinal data), ποσοτικά δεδομένα (quantitative data).
- Βήματα ανάπτυξης ενός οικονομετρικού μοντέλου

# Εισαγωγή

# Η φύση και ο σκοπός της Οικονομετρίας

□ Τι είναι η Οικονομετρία;

- Είναι κυριολεκτικά η «μέτρηση των οικονομικών μεταβλητών» και των σχέσεων τους.

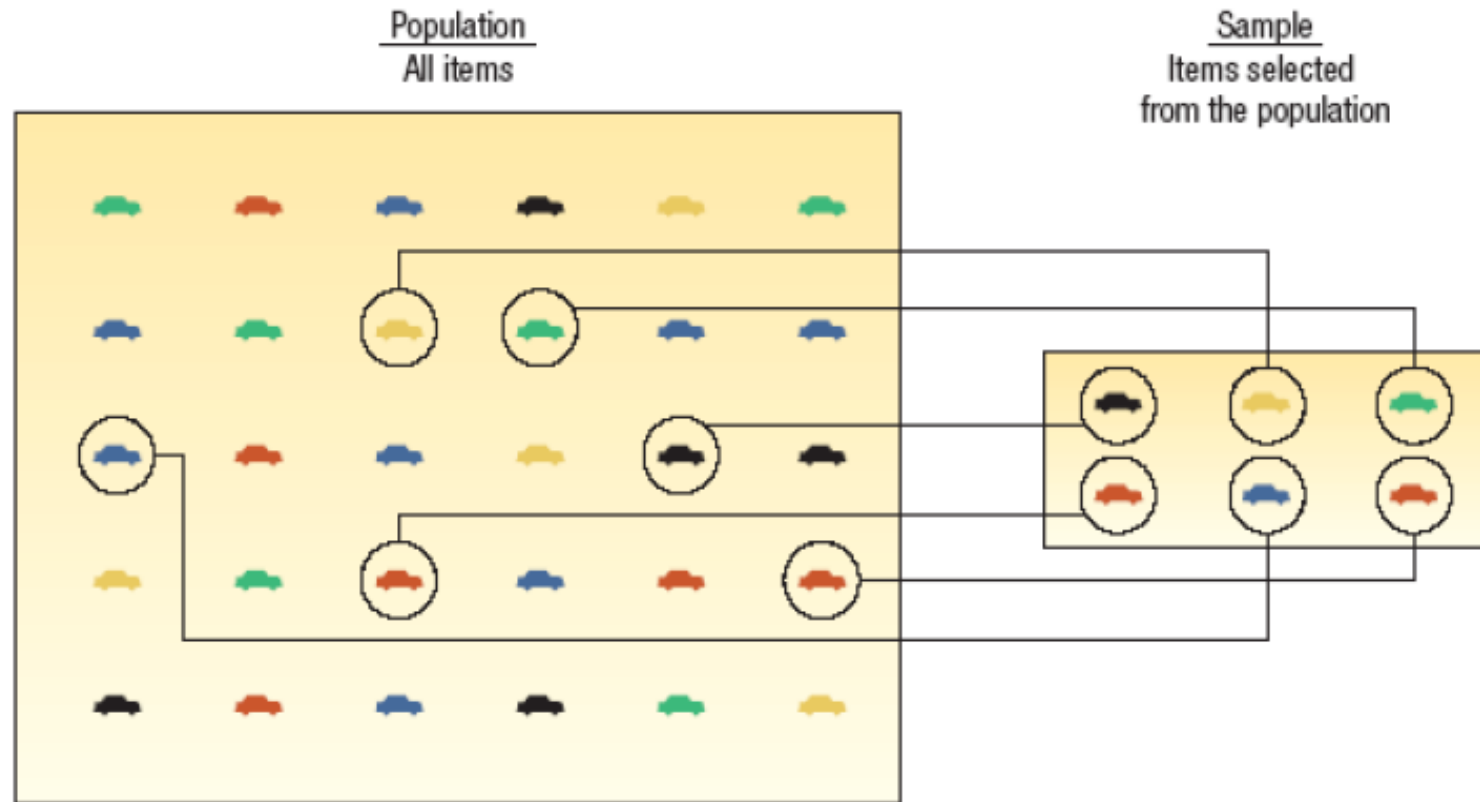
□ Τι είναι η χρηματοοικονομική οικονομμετρία;

- Η εφαρμογή στατιστικών και μαθηματικών μεθόδων σε προβλήματα των χρηματοοικονομικών.

# Δείγμα και πληθυσμός

- Ο **πληθυσμός** είναι το **σύνολο όλων των εξεταζόμενων ατόμων**, αντικειμένων ή μετρήσεων σε μια έρευνα.
- Παραδείγματα:
  - το σύνολο των ξενοδοχείων της χώρας.
  - το σύνολο των ποδοσφαιρικών ομάδων της Ελλάδας.
  - το σύνολο των ημερομισθίων μίας επιχείρησης.
- **Δείγμα (Sample):** Είναι κάθε **υποσύνολο του πληθυσμού**.
  - Παραδείγματα:
    - το σύνολο των ξενοδοχείων του νομού Καβάλας.
    - το σύνολο των ομάδων ποδοσφαίρου της *Super League*.

- Γιατί να μη μελετήσουμε **ολόκληρο τον πληθυσμό**;
  - Μπορεί να είναι μια **δαπανηρή** διαδικασία (π.χ. κάθε πολίτης που κατοικεί στην Κίνα)
  - Μπορεί να είναι **ανέφικτο** (π.χ. να περάσουμε από crash test κάθε αυτοκίνητο που φτιάχνουμε)
- Ένα δείγμα θα πρέπει να είναι **τυχαίο** (να μην αποκλείσουμε κανένα άτομο του πληθυσμού, ώστε να είναι ισο-πίθανο να συμπεριληφθεί οποιοσδήποτε από τον πληθυσμό) και **αντιπροσωπευτικό** (π.χ. να συμπεριλάβουμε άτομα από όλες τις ηλικιακές ομάδες του πληθυσμού).



# Επαγωγική και Περιγραφική Στατιστική

- Το **μειονέκτημα** των στατιστικών ενός δείγματος είναι ότι μπορούμε να βγάλουμε **συμπεράσματα μόνο για το δείγμα** στο οποίο αναφέρονται και όχι για τον πληθυσμό από τον οποίο προέρχεται το δείγμα αυτό.
- Η **Επαγωγική ή Αναλυτική Στατιστική** ασχολείται με τις τεχνικές οι οποίες με βάση τα στοιχεία ενός δείγματος **οδηγούν στην εξαγωγή συμπερασμάτων για τον πληθυσμό με σχετική ασφάλεια** (π.χ. 90%, 95%, 99%) από τον οποίο συλλέχθηκε το δείγμα.
- Διαμέσου της **θεωρίας πιθανοτήτων** μπορούμε να χρησιμοποιήσουμε τις ιδιότητες ενός πληθυσμού για να κάνουμε προβλέψεις για ένα δείγμα.
  - Χρειαζόμαστε τη θεωρία πιθανοτήτων για την επαγωγική στατιστική
  - Παραδείγματα:
    - Πως θα συμπεριφερόταν το δείγμα αν προερχόταν από τον πληθυσμό  $X$ ;
    - Αν προερχόταν από τον πληθυσμό  $\Psi$ ;
    - Σε ποιον από τους δύο μοιάζει;



# Παράμετρος και στατιστική

- **Παράμετρος** είναι μια αριθμητική μέτρηση που αναφέρεται σε ένα χαρακτηριστικό του πληθυσμού.
- **Στατιστική** είναι μια αριθμητική μέτρηση που αναφέρεται σε ένα χαρακτηριστικό του δείγματος.

# Μεταβλητή

- **Μεταβλητή (Variable):** Ο όρος **μεταβλητή** ο οποίος χρησιμοποιείται σε πλήθος ερευνών αναφέρεται σε ένα χαρακτηριστικό ή μια ιδιότητα ενός προσώπου, αντικειμένου ή κατάστασης, που ερευνάται και μπορεί να μεταβάλλεται ή να παίρνει διαφορετικές τιμές σε διαφορετικές καταστάσεις.
- Είναι πολύ σημαντικό σε κάθε έρευνα να βρεθούν οι μεταβλητές που μας ενδιαφέρουν, να ελεγχθεί αν είναι διαθέσιμες και να μετρηθούν με αξιόπιστο και συγκεκριμένο τρόπο.
- Παραδείγματα:
  - Το **εισόδημα** μπορεί να χαρακτηριστεί ως μεταβλητή καθώς μπορεί να πάρει διαφορετικές τιμές για διαφορετικά άτομα ή για το ίδιο άτομο σε διαφορετικές χρονικές περιόδους.
  - Το **θρήσκευμα** μπορεί επίσης να χαρακτηριστεί ως μεταβλητή καθώς στο θρήσκευμα κάθε ατόμου μπορεί να αποδοθεί κάποια τιμή.

Δεδομένα

# Τύποι Δεδομένων



## ❑ Χρονολογικές σειρές (time-series data)

- ΑΕΠ (GDP) (ετήσια, τριμηνιαία)
- Ανεργία (ετήσια, τριμηνιαία, μηνιαία)
- Δημοσιονομικό έλλειμμα (ετήσια)
- Η προσφορά χρήματος (μηνιαία)
- Η τιμή κλεισίματος ενός δείκτη μετοχών (ημερήσια, εβδομαδιαία, μηνιαία, ...)

## ❑ Διαστρωματικά δεδομένα (cross-sectional data)

- Οι αποδόσεις των μετοχών του ΧΑΑ για μία συγκεκριμένη ημέρα διαπραγμάτευσης.
- Τα αποτελέσματα των εξετάσεων των φοιτητών σε μία συγκεκριμένη εξεταστική περίοδο για ένα συγκεκριμένο μάθημα.

## ❑ Χρονολογικές Σειρές και Διαστρωματικά στοιχεία ταυτόχρονα (Panel Data)

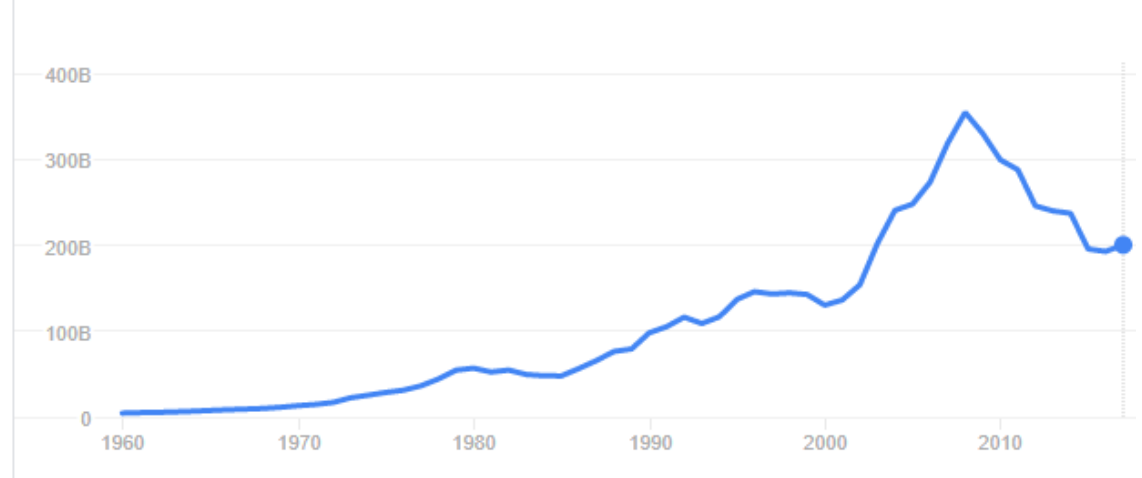
- Οι ημερήσιες τιμές κλεισίματος όλων των μετοχών στον κλάδο της ενέργειας για ένα έτος.

# Παραδείγματα (1)

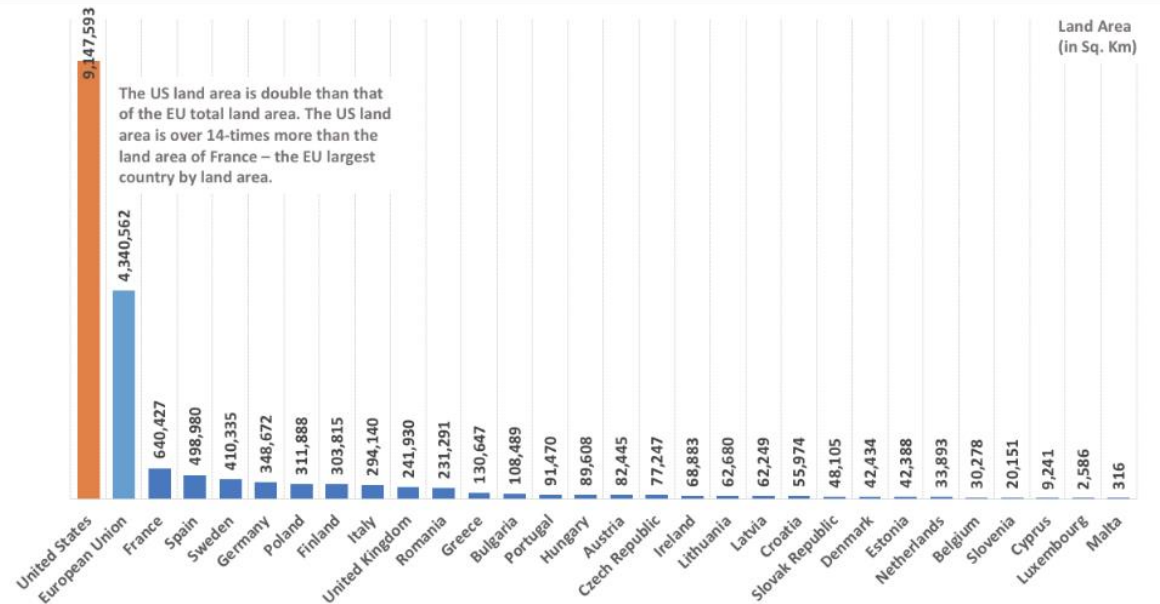
- Μία χρονολογική σειρά (time series) είναι η συλλογή παρατηρήσεων για μία μεταβλητή σε διαφορετικές χρονικές στιγμές, π.χ. η διαχρονική εξέλιξη του Ελληνικού ΑΕΠ.
- Τα διαστρωματικά δεδομένα (cross-sectional data) αναφέρονται σε παρατηρήσεις μίας μεταβλητής σε μία και μόνη συγκεκριμένη στιγμή στο χρόνο για διαφορετικά άτομα, αντικείμενα ή μετρήσεις, π.χ. η έκταση διαφορετικών χωρών σε τετραγωνικά χιλιόμετρα

Greece / Gross domestic product

200.3 billion USD (2017)



US VS EU: LAND AREA COMPARISON



## Παραδείγματα (2)

- Όταν μία μεταβλητή (ή πολλές) έχει (έχουν) συλλεχθεί και στη διάσταση του χρόνου (χρονολογική σειρά - time series) και στη διάσταση των διαφορετικών ατόμων, αντικειμένων, περιουσιακών στοιχείων (διαστρωματικά στοιχεία – cross sectional data) τότε έχουμε τα λεγόμενα **panel data**.
- **π.χ. 1** Ένα σύνολο παρατηρήσεων που περιλαμβάνει τις ιστορικές τιμές όλων των μετοχών του Χρηματιστηρίου Αξιών Αθηνών για το διάστημα 2010-2019. Διατεταγμένες η μία κάτω από την άλλη.
- **π.χ. 2** Το εισόδημα, η ηλικία και το φύλο για διαφορετικά άτομα σε διαφορετικά έτη

person	year	income	age	sex
1	2001	1300	27	1
1	2002	1600	28	1
1	2003	2000	29	1
2	2001	2000	38	2
2	2002	2300	39	2
2	2003	2400	40	2

# Balanced vs. Unbalanced Panel Data



<i>FIRM</i>	<i>YEAR</i>	<i>PROD</i>	<i>AREA</i>	<i>LABOR</i>	<i>FERT</i>
1	1990	7.87	2.50	160	207.5
1	1991	7.18	2.50	138	295.5
1	1992	8.92	2.50	140	362.5
1	1993	7.31	2.50	127	338.0
1	1994	7.54	2.50	145	337.5
1	1995	4.51	2.50	123	207.2
1	1996	4.37	2.25	123	345.0
1	1997	7.27	2.15	87	222.8
2	1990	10.35	3.80	184	303.5
2	1991	10.21	3.80	151	206.0
2	1992	13.29	3.80	185	374.5
2	1993	18.58	3.80	262	421.0
2	1994	17.07	3.80	174	595.7
2	1995	16.61	4.25	244	234.8
2	1996	12.28	4.25	159	479.0
2	1997	14.20	3.75	133	170.0

<i>FIRM</i>	<i>YEAR</i>	<i>PROD</i>	<i>AREA</i>	<i>LABOR</i>	<i>FERT</i>
1	1990	7.87	2.50	160	207.5
1	1991	7.18	2.50	138	295.5
1	1992	8.92	2.50	140	362.5
1	1993	7.31	2.50	127	338.0
1	1994	7.54	2.50	145	337.5
1	1995	4.51	2.50	123	207.2
1	1996	4.37	2.25		345.0
1	1997				222.8
2	1990	10.35	3.80	184	303.5
2	1991	10.21	3.80	151	206.0
2	1992	13.29	3.80	185	374.5
2	1993	18.58	3.80	262	421.0
2	1994	17.07	3.80	174	595.7
2	1995	16.61	4.25	244	234.8
2	1996	12.28	4.25	159	479.0
2	1997	14.20	3.75	133	170.0

## Υπολογισμός Αποδόσεων

### □ Απλές Αποδόσεις (simple returns)

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \times 100\%$$

όπου,  $R_t$  δηλώνει την απόδοση τη χρονική στιγμή  $t$   
 $P_t$  δηλώνει την τιμή τη χρονική στιγμή  $t$



### □ Λογαριθμικές Αποδόσεις (log returns ή continuously compounded returns)

$$R_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \times 100\%$$

όπου,  $R_t$  δηλώνει την απόδοση τη χρονική στιγμή  $t$   
 $P_t$  δηλώνει την τιμή τη χρονική στιγμή  $t$   
 $\ln$  δηλώνει τον φυσικό λογάριθμο

□ Προσοχή! Οι λογαριθμικές διαφορές (αποδόσεις) δεν ισχύουν για τα χαρτοφυλάκια μετοχών. Αυτό συμβαίνει γιατί η απόδοση ενός χαρτοφυλακίου είναι ο σταθμισμένος μέσος όρος των απλών αποδόσεων κάθε περιουσιακού στοιχείου που περιλαμβάνεται στο χαρτοφυλάκιο.



# Λογαριθμικές Αποδόσεις

1. Οι λογαριθμικές αποδόσεις έχουν την επιθυμητή κάποιες φορές ιδιότητα ότι μπορούν να προστεθούν και να παραμείνουν ακριβείς.

2. Για παράδειγμα αν θέλουμε να βρούμε την εβδομαδιαία απόδοση μίας μετοχής, μπορούμε απλά να προσθέσουμε τις λογαριθμικές αποδόσεις αυτής της μετοχής για τις πέντε ημέρες διαπραγμάτευσης που περιέχει μία ημερολογιακή εβδομάδα (trading days: 5, calendar days: 7)

$$r_1 = \ln p_1/p_0 = \ln p_1 - \ln p_0$$

$$r_2 = \ln p_2/p_1 = \ln p_2 - \ln p_1$$

$$r_3 = \ln p_3/p_2 = \ln p_3 - \ln p_2$$

$$r_4 = \ln p_4/p_3 = \ln p_4 - \ln p_3$$

$$r_5 = \ln p_5/p_4 = \ln p_5 - \ln p_4$$

---

$$\text{Απόδοση εβδομάδας: } \ln p_5 - \ln p_0 = \ln p_5/p_0$$

## Δεδομένα και Παρατηρήσεις

- **Δεδομένα (data):** Είναι τα στοιχεία τα οποία συλλέγονται, αναλύονται και συνοψίζονται για παρουσίαση και ερμηνεία. Δεδομένα αποκτώνται με τη συλλογή των μετρήσεων κάθε μεταβλητής και μπορεί να είναι αριθμοί, λέξεις ή και σύμβολα.
- **Παρατήρηση (observation):** Η κάθε μέτρηση, κάθε μεταβλητής ονομάζεται παρατήρηση.
- Οι πηγές δεδομένων είναι δύο ειδών:
  - **(i) Πρωτογενή δεδομένα:** Αυτά δεν υπάρχουν σε καμία μορφή και έτσι θα πρέπει να συλλεχθούν για πρώτη φορά από τον πληθυσμό ή το δείγμα που έχει επιλεγεί, π.χ. οι απαντήσεις σε ερωτηματολόγια για τη χρήση ηλεκτρονικών πληρωμών.
  - **(ii) Δευτερογενή δεδομένα:** Αυτά υπάρχουν ήδη σε κάποια μορφή δημοσιευμένα ή αδημοσίευτα, π.χ. τα δεδομένα του Ελληνικού ΑΕΠ που συγκεντρώνει η Ελληνική Στατιστική Αρχή.
- **Σύνολο δεδομένων (dataset):** είναι μια καταγραφή των τιμών που αντιστοιχούν σε κάθε μεταβλητή και παρατήρηση
- Συχνά είναι σε μια **μορφή πίνακα** όπου οι παρατηρήσεις βρίσκονται στις γραμμές και οι μεταβλητές σε στήλες και μπορεί να εισαχθεί σε ένα **αρχείο excel**.

# Κατηγορικά δεδομένα (nominal or categorical data)

- **Κατηγορικά δεδομένα (categorical data):** Κατηγορικά δεδομένα ονομάζονται τα δεδομένα που προέρχονται από μεταβλητές οι τιμές των οποίων εκφράζουν τάξεις ή κατηγορίες.
- Τα κατηγορικά δεδομένα **δεν μπορούν να διαταχθούν κατά σειρά μεγέθους** αλλά ούτε και κατά μία φυσική διάταξη. **Εξ'ορισμού, λοιπόν, δεν μπορεί να είναι αριθμοί.**
- Πολλές φορές, όμως, **αντιστοιχίζουμε τα δεδομένα αυτά με αριθμούς** ούτως ώστε να μπορέσουμε να τα επεξεργαστούμε με τη βοήθεια Η/Υ. Η αντιστοιχία αυτή είναι απλός συμβολισμός και επομένως δεν νοείται κανενός είδους αριθμητικής πράξης.
- **Αντιπροσωπευτικά παραδείγματα** κατηγορικών δεδομένων αποτελούν το φύλο, με τιμές αρσενικό, θηλυκό, το θρήσκευμα με τιμές π.χ. Ορθόδοξος, Καθολικός, Προτεστάντης, Μουσουλμάνος , κ.ά., η ομάδα αίματος με τιμές Α, Β, ΑΒ, Ο.

# Διατάξιμα δεδομένα (ordinal level)

- **Διατάξιμα δεδομένα (ordinal data):** Διατάξιμα ονομάζονται τα δεδομένα που προέρχονται από μεταβλητές που εκφράζουν ένα είδος ποιοτικού χαρακτηριστικού της στατιστικής μονάδας που υπόκειται στη μέτρηση.
- Όπως και τα κατηγορικά, τα διατάξιμα δεδομένα κατηγοριοποιούνται σε τάξεις με τη διαφορά, όμως, ότι οι εν λόγω τάξεις βρίσκονται τώρα σε μια φυσική διάταξη μεταξύ τους, π.χ. από το “μικρότερο στο μεγαλύτερο” ή από το “θετικό στο αρνητικό”.
- Δεν γνωρίζουμε, όμως, την ακριβή απόσταση μεταξύ αυτών των τάξεων ή την απόσταση τους από κάποιο ανεξάρτητο σημείο αναφοράς. Τα διατάξιμα δεδομένα τα αντιστοιχίζουμε πολλές φορές με αριθμούς η επιλογή των οποίων ουσιαστικά είναι αυθαίρετη αλλά γίνεται έτσι ώστε να διατηρείται η διάταξη που υπάρχει μεταξύ των τάξεων.
- **Αντιπροσωπευτικά παραδείγματα** διατάξιμων δεδομένων αποτελούν το επίπεδο εκπαίδευσης ενός ατόμου όταν εκφράζεται από τις κατηγορίες: πρωτοβάθμια, δευτεροβάθμια και τριτοβάθμια εκπαίδευση, το μέγεθος μιας οικογένειας σε αριθμό μελών όταν αποδοθεί για παράδειγμα από τις τιμές: μικρό (μέγεθος), μέτριο και μεγάλο όπως επίσης και η στάση ενός ατόμου σε σχέση, για παράδειγμα, με την οικονομική πολιτική της κυβέρνησης όταν αξιολογηθεί ως: πολύ θετική, θετική, αδιάφορη, αρνητική, πολύ αρνητική.

# Ποσοτικά δεδομένα

- **Ποσοτικά δεδομένα με αυθαίρετο μηδέν:** Ποσοτικά με αυθαίρετο μηδέν ονομάζονται τα δεδομένα που προέρχονται από ποσοτικές μεταβλητές. Συνεπώς, τώρα τα δεδομένα είναι αριθμοί και ισχύουν τα παρακάτω.
  - κατηγοριοποιούνται σε τάξεις
  - μπορούν να ταξινομηθούν κατά σειρά μεγέθους
  - μπορούμε να προσδιορίσουμε την μεταξύ τους απόσταση
- Το μηδέν που ορίζουμε εμείς ως αρχή της κλίμακας μέτρησης δεν εκφράζει εννοιολογικά και την αρχή των εννοιών των αντίστοιχων μεταβλητών.
- Τα ποσοτικά δεδομένα με αυθαίρετο μηδέν προέρχονται ως επί το πλείστον από **κοινωνικές μεταβλητές**.
- **Αντιπροσωπευτικά παραδείγματα** ποσοτικών δεδομένων με αυθαίρετο μηδέν αποτελούν από το χώρο των κοινωνικών επιστημών η νοημοσύνη, η προσωπικότητα και η δημοτικότητα ενός ατόμου (δεν υπάρχει μία σταθερή μονάδα μέτρησης αποδεκτή από όλους) και
- Από το χώρο των **φυσικών επιστημών**, η θερμοκρασία όταν αυτή εκφράζεται σε βαθμούς Κελσίου ή βαθμούς Φαρενάιτ, ο χρόνος όπου ως μονάδες μέτρησης του μπορούν να οριστούν το δευτερόλεπτο, το λεπτό, η ώρα κ.τ.λ.

# Ποσοτικά δεδομένα με απόλυτο μηδέν (ratio level)

- Τα ποσοτικά δεδομένα με απόλυτο μηδέν μπορούν να θεωρηθούν ως μία ειδική περίπτωση ποσοτικών δεδομένων με αυθαίρετο μηδέν.
- Η διαφορά έγκειται στο ότι τώρα η κλίμακα μέτρησης που χρησιμοποιούμε έχει ως μηδέν ένα πραγματικό σημείο αναφοράς, το οποίο εκφράζει την πραγματική (φυσική) κατάσταση της μεταβλητής.
- Τα δεδομένα αυτά προκύπτουν συχνότερα όταν έχουμε να μελετήσουμε μεταβλητές από τις φυσικές επιστήμες.
- **Αντιπροσωπευτικά παραδείγματα** ποσοτικών δεδομένων με απόλυτο μηδέν βρίσκουμε κυρίως στις φυσικές επιστήμες, όπως για παράδειγμα η **ταχύτητα ενός αυτοκινήτου**.

# Δεδομένα και Παρατηρήσεις

**Nominal level** - data that is classified into categories and cannot be arranged in any particular order.

EXAMPLES: eye color, gender, religious affiliation.

**Interval level** - similar to the ordinal level, with the additional property that meaningful amounts of differences between data values can be determined. There is no natural zero point.

EXAMPLES: Temperature on the Fahrenheit scale.

**Ordinal level** – data arranged in some order, but the differences between data values cannot be determined or are meaningless.

EXAMPLE: During a taste test of 4 soft drinks, Mellow Yellow was ranked number 1, Sprite number 2, Seven-up number 3, and Orange Crush number 4.

**Ratio level** - the interval level with an inherent zero starting point. Differences and ratios are meaningful for this level of measurement.

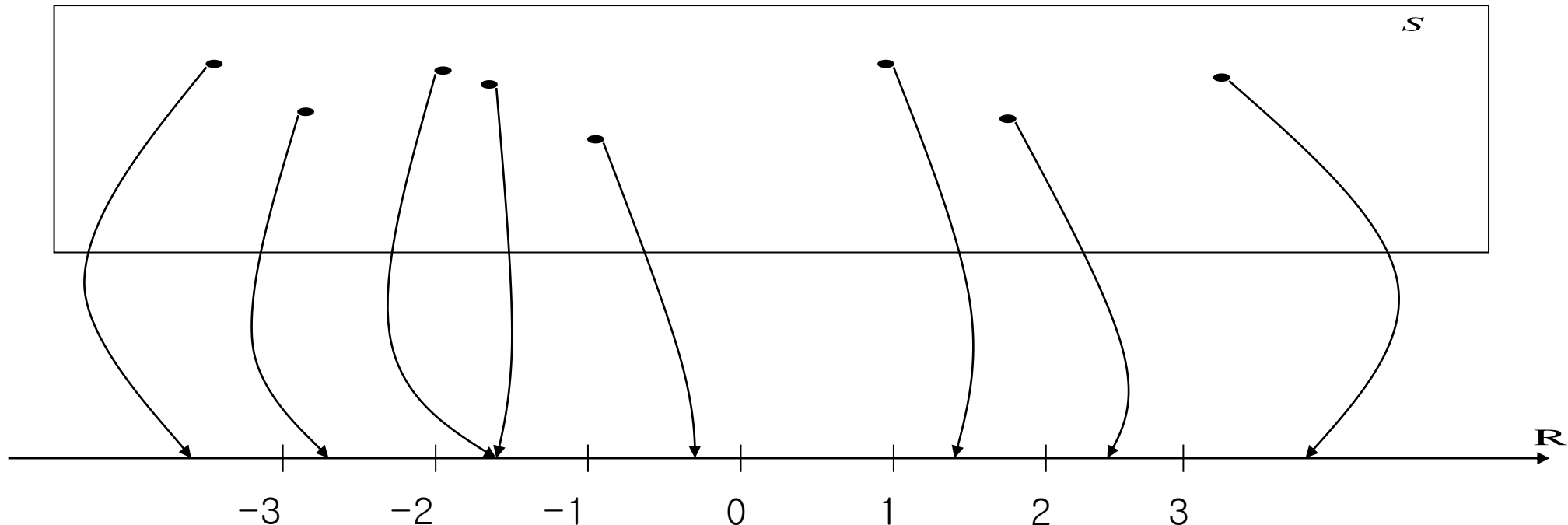
EXAMPLES: Monthly income of surgeons, or distance traveled by manufacturer's representatives per month.

# Τυχαία Μεταβλητή & Κεντρικό Οριακό Θεώρημα



# Τυχαία μεταβλητή- Ορισμός

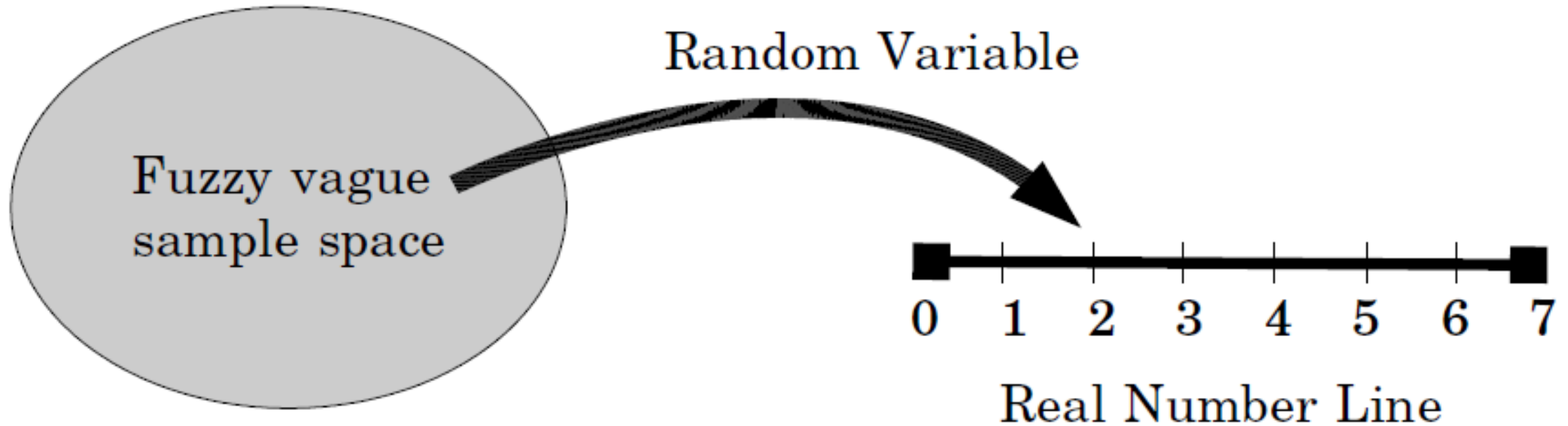
- Μια τυχαία μεταβλητή είναι μια μεταβλητή που λαμβάνει αριθμητικές τιμές που σχετίζονται με το τυχαίο αποτέλεσμα ενός **πειράματος**, όπου μια (και μόνο μία) αριθμητική τιμή εκχωρείται σε κάθε σημείο δείγματος.



# Τυχαία μεταβλητή- Ορισμός

- Μια τυχαία μεταβλητή λαμβάνει ένα καθορισμένο σύνολο τιμών με διαφορετικές πιθανότητες.
  - Για παράδειγμα, εάν ρίξετε ένα ζάρι, το αποτέλεσμα είναι τυχαίο (όχι σταθερό) και υπάρχουν 6 πιθανά αποτελέσματα, καθένα από τα οποία συμβαίνει με πιθανότητα το ένα έκτο.
  - Για παράδειγμα, εάν κάνετε δημοσκόπηση σχετικά με τις προτιμήσεις ψήφου, το ποσοστό του δείγματος που απαντά "Ναι στην πρόταση 100" είναι επίσης μια τυχαία μεταβλητή (το ποσοστό θα είναι ελαφρώς διαφορετικό κάθε φορά που κάνετε δημοσκόπηση).
- Σε γενικές γραμμές, η πιθανότητα είναι πόσο συχνά αναμένουμε να προκύψουν διαφορετικά αποτελέσματα εάν επαναλαμβάνουμε **το πείραμα ξανά και ξανά**.

# Παράδειγμα μιας τυχαίας μεταβλητής



- Σκεφτείτε το πείραμα της ρίψης ενός νομίσματος 5 φορές.
- Μερικές τυχαίες μεταβλητές που μπορούν να προκύψουν από το πείραμα:
  - Πόσες φορές το αποτέλεσμα ήταν κορώνα;
  - Πόσες φορές το αποτέλεσμα ήταν γράμματα;
  - Παίρνει την τιμή '1' αν είχε το ίδιο αποτέλεσμα κάθε φορά, '0' αν όχι.

# Κατανομή πιθανότητας

- Μια κατανομή πιθανότητας είναι μια λίστα με όλες τις πιθανές τιμές μιας τυχαίας μεταβλητής και μια πιθανότητα που σχετίζεται με το κάθε αποτέλεσμα.
  - Παράδειγμα: Τα ταξί επιτρέπεται να μεταφέρουν μέχρι 4 επιβάτες. Η πιθανότητα να μεταφέρουν ένα συγκεκριμένο αριθμό επιβατών είναι:

<b>Number</b>	<b>Probability</b>
0	.30
1	.35
2	.15
3	.10
4	<i>guess!</i>

# Διακριτές και συνεχείς μεταβλητές

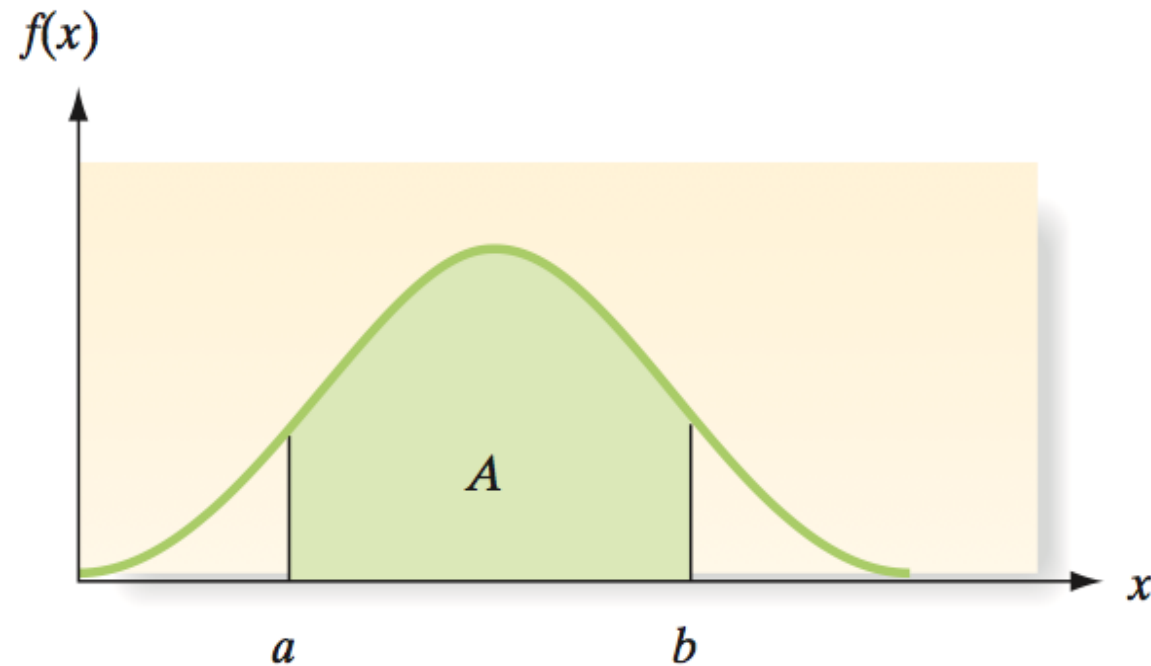
- Οι διακριτές τυχαίες μεταβλητές συνήθως παίρνουν νούμερα από ένα πεπερασμένο αριθμό αποτελεσμάτων έτσι ώστε τα αποτελέσματα που σχετίζονται με τις πιθανότητες μπορούν να καταγραφούν σε έναν πίνακα.
- Οι συνεχείς τυχαίες μεταβλητές μπορούν να πάρουν τιμές από ένα εύρος αποτελεσμάτων, ένα άπειρο αριθμό πιθανών αποτελεσμάτων.
- Διακριτές τυχαίες μεταβλητές
  - Αριθμός πωλήσεων
  - Αριθμός κλήσεων
  - Αριθμός μετοχών
  - Άνθρωποι στην ουρά
  - Λάθη ανά σελίδα
- Συνεχείς τυχαίες μεταβλητές
  - Μήκος
  - Βάθος
  - Χρόνος
  - Βάρος

# Συνάρτηση Πιθανότητας

- Μια συνάρτηση πιθανότητας χαρτογραφεί τις πιθανές τιμές μιας τυχαίας μεταβλητής  $x$  έναντι των αντίστοιχων πιθανοτήτων εμφάνισής τους,  $p(x)$ .
- Το  $p(x)$  είναι ένας αριθμός από 0 έως 1.0.
- Η περιοχή κάτω από μια συνάρτηση πιθανότητας είναι πάντα 1.

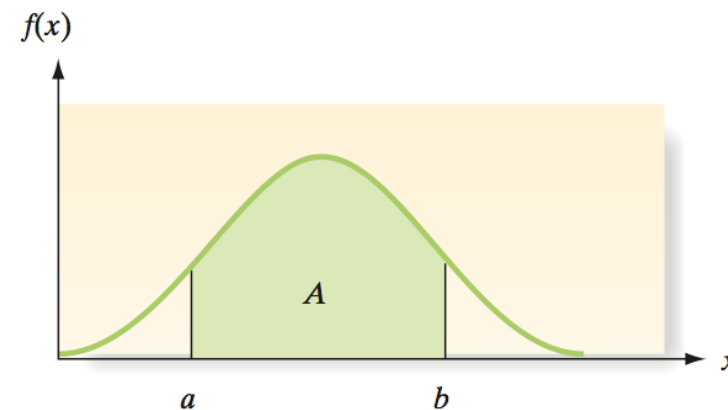
## Συνάρτηση Πυκνότητας Πιθανότητας

- Η γραφική μορφή της κατανομής πιθανότητας για μια συνεχή τυχαία μεταβλητή  $x$  είναι μια ομαλή καμπύλη



□ Αυτή η καμπύλη, συνάρτηση του  $x$ , συμβολίζεται με το σύμβολο  $f(x)$  και ονομάζεται συνάρτηση πυκνότητας πιθανότητας (pdf), συνάρτηση συχνότητας ή κατανομή πιθανότητας.

□ Οι περιοχές με κατανομή πιθανότητας αντιστοιχούν σε πιθανότητες για  $x$ . Το εμβαδόν  $A$  κάτω από την καμπύλη μεταξύ δύο σημείων  $a$  και  $b$  είναι η πιθανότητα ότι το  $x$  λαμβάνει μια τιμή μεταξύ  $a$  and  $b$ .





# Τυχαία Μεταβλητή

## Πρώτη Ροπή

- Οι περισσότερες σειρές δεδομένων στη λογιστική και χρηματοοικονομική θεωρούνται καλύτερα τυχαίες μεταβλητές.
- Έστω η τυχαία μεταβλητή  $X$
- **Μέσος (Expectation)  $E(X)$**

### Ιδιότητες

$$E(c) = c, c \text{ σταθερά}$$

$$E(cX) = cE(X), c \text{ σταθερά}$$

$$E(cX + d) = cE(X) + d, c, d \text{ σταθερές}$$

Όταν  $X, Y$  είναι ανεξαρτητες μεταβλητες  $E(XY) = E(X)E(Y)$

## Τυχαία Μεταβλητή Δεύτερη Ροπή

- Διακύμανση Variance  $\text{Var}(X)$

### Ιδιότητες

$$\text{Var}(X) = E[X - E(X)]^2, c \text{ σταθερά}$$

$$\text{Var}(cX + d) = c^2 \text{Var}(X), c, d \text{ σταθερές}$$

$$\text{Var}(c) = 0, c \text{ σταθερά}$$

$$\text{Var}(cX + dY) = c^2 \text{Var}(X) + d^2 \text{Var}(Y) + 2cd \text{Cov}(X, Y)$$

$c, d$  σταθερές,

$\text{Cov}(X, Y)$  η συνδυακύμανση των  $X$  και  $Y$

# Τυχαία Μεταβλητή

## Τρίτη και Τέταρτη Ροπή

- **Ασυμμετρία**

$$\mu_3 = E[X - E(X)]^3$$

❖ *Μετρά το βαθμό (την έκταση) στον οποίο(στην οποία) η κατανομή δεν είναι συμμετρική*

- Εάν  $\mu_3 > 0$  η δεξιά ουρά είναι η μακρύτερη
- Εάν  $\mu_3 < 0$  η αριστερή ουρά είναι η μακρύτερη
- Εάν  $\mu_3 = 0$  κανονική κατανομή

- **Κύρτωση**

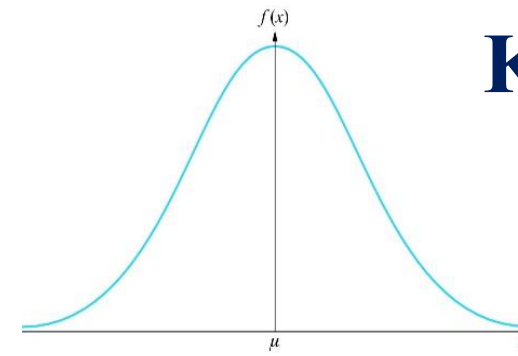
$$\mu_4 = E[X - E(X)]^4$$

❖ *Μετράει το “πάχος της ουράς”*

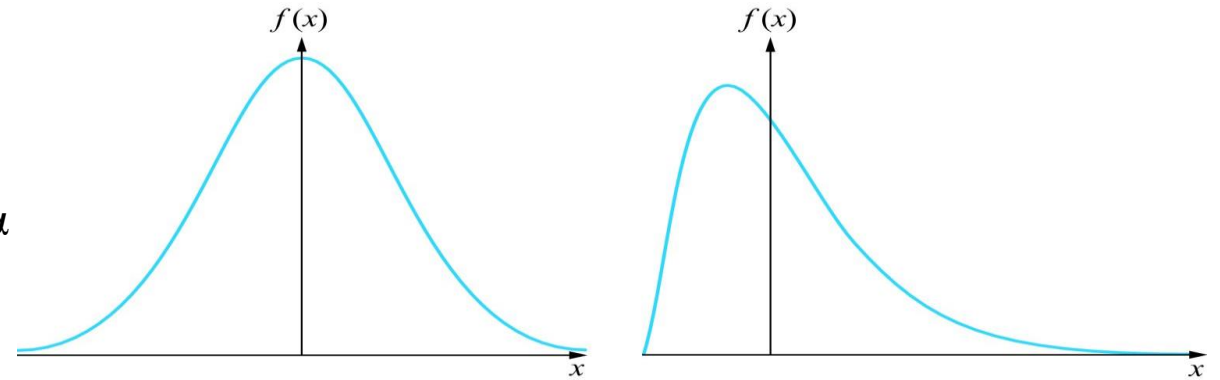
- Εάν  $\mu_4 > 3$  η κατανομή έχει μακριές ή παχιές ουρές (λεπτόκυρτη)
- Εάν  $\mu_4 < 3$  η κατανομή έχει κοντές ή λεπτές ουρές (πλατύκυρτη)
- Εάν  $\mu_4 = 3$  κανονική κατανομή

# Κατανομή

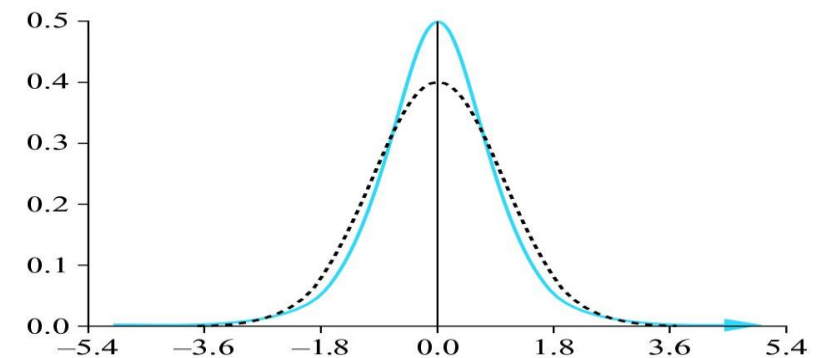
(a) Κανονική Κατανομή



(b) Κανονική Κατανομή έναντι Κατανομής με Ασυμμετρία



(c) Κανονική Κατανομή έναντι Λεπτόκυρτης Κατανομής



# Κανονική Κατανομή (Normal Distribution)

## □ Σημασία

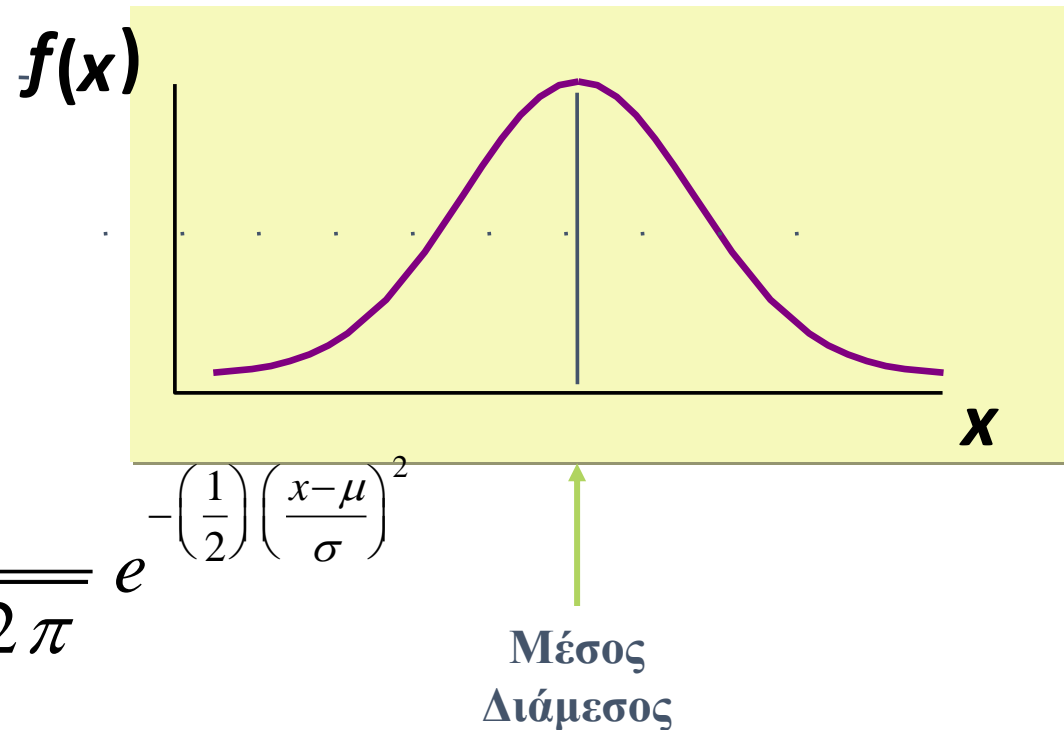
1. Περιγράφει πολλές τυχαίες διαδικασίες ή συνεχή φαινόμενα
2. Βάση για κλασικά στατιστικά συμπεράσματα

- Συμμετρική
- Μέσος και διάμεσος είναι ίσοι
- Συνάρτηση πυκνότητας πιθανότητας (Probability Density Function)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{1}{2}\right)\left(\frac{x-\mu}{\sigma}\right)^2}$$

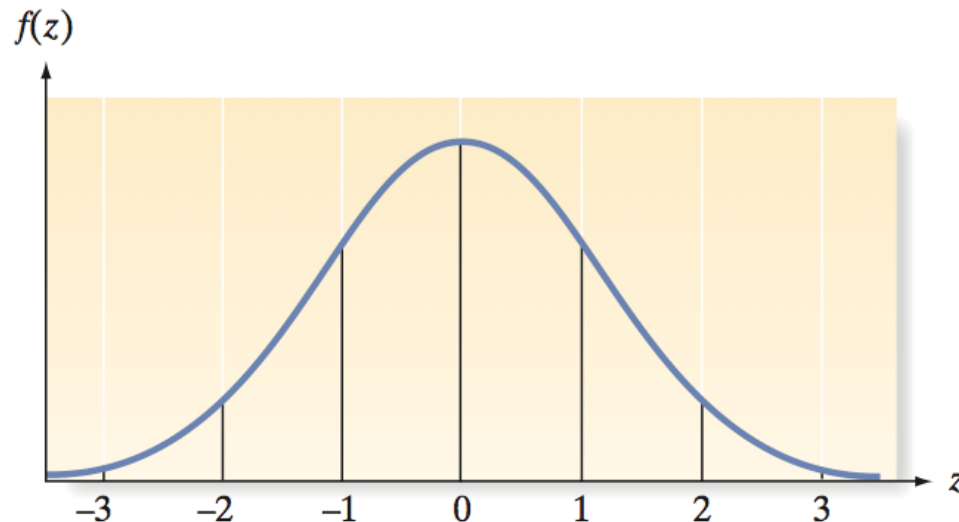
Όπου

$\mu$  = μέση τιμή της τυχαίας μεταβλητής  $X$   
 $\sigma$  = τυπική απόκλιση



# Τυποποιημένη Κανονική Κατανομή (Standard Normal Distribution)

- Η Τυποποιημένη Κανονική Κατανομή είναι μια κανονική κατανομή με  $\mu = 0$  και  $\sigma = 1$ . Μια τυχαία μεταβλητή με τυπική κανονική κατανομή, που συμβολίζεται με το σύμβολο  $z$ , ονομάζεται τυπική κανονική τυχαία μεταβλητή.

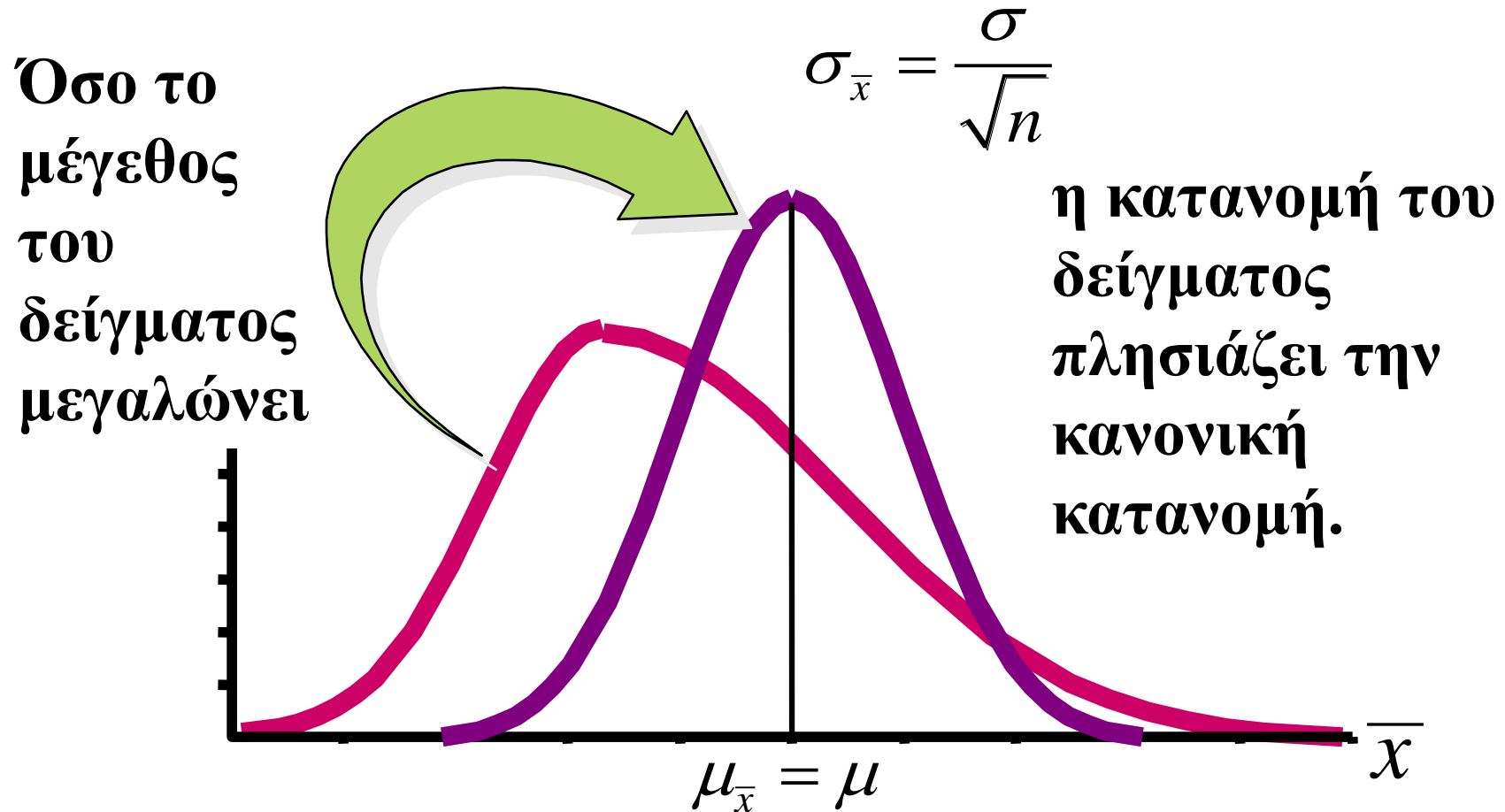


# Κεντρικό Οριακό Θεώρημα

Αν από έναν πληθυσμό που ακολουθεί οποιαδήποτε κατανομή με μέση τιμή  $\mu$  και τυπική απόκλιση  $\sigma$ , επιλέξουμε τυχαία δείγματα μεγέθους  $n$  και υπολογίσουμε τους μέσους τους, τότε, για μεγάλα  $n$  (θεωρητικά  $n \rightarrow \infty$ ) η κατανομή αυτών των μέσων (των δειγματικών) είναι κατά προσέγγιση κανονική κατανομή με μέση τιμή επίσης  $\mu_{\bar{x}} = \mu$  και τυπική απόκλιση

$$\sigma_{\bar{x}} = \sigma / \sqrt{n}.$$

# Κεντρικό Οριακό Θεώρημα

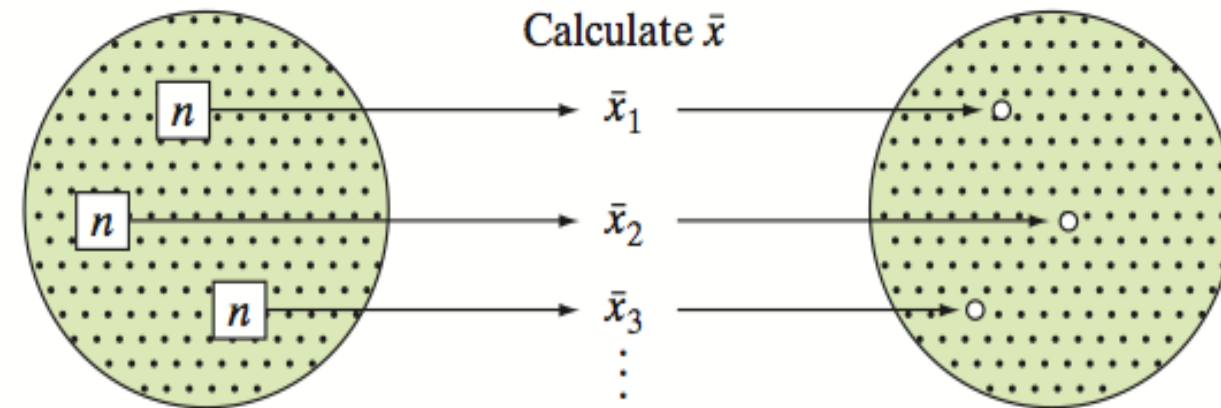




# ΣΥΝΟΠΤΙΚΑ

## Generating the Sampling Distribution of $\bar{x}$

Select sample size  $n$  (large)  
from target population



Population:

Mean =  $\mu$   
Std. Dev. =  $\sigma$   
Unknown shape

Repeat this  
process an  
infinite  
number  
of times

Sampling distribution of  $\bar{x}$   
(i.e., theoretical population of  $\bar{x}$ 's)

Mean =  $\mu_{\bar{x}} = \mu$   
Std. Dev. =  $\sigma_{\bar{x}} = \sigma/\sqrt{n}$   
Normal distribution (Central Limit Theorem)

# Κεντρικό Οριακό Θεώρημα- Παράδειγμα

Μας είναι γνωστό, ότι τα μήλα μιας δενδροκαλλιέργειας έχουν μέσο βάρος  $\mu = 50\text{gr}$  με τυπική απόκλιση  $\sigma = 10\text{gr}$ . Η παραγωγή της καλλιέργειας συσκευάζεται σε κιβώτια και προωθείται στην αγορά. Σε κάθε κιβώτιο τοποθετούνται 400 μήλα (τυχαία επιλεγμένα). Μπορούμε να υπολογίσουμε ποιο ποσοστό (κατά προσέγγιση) των κιβωτίων περιέχει μήλα με μέσο βάρος: α) μεγαλύτερο των 51.25gr και β) μικρότερο των 49gr;

## Απάντηση

Τα μήλα κάθε κιβωτίου είναι ένα τυχαίο δείγμα μεγέθους  $n = 400$  από τον πληθυσμό των μήλων της παραγωγής. Η μορφή της κατανομής των βαρών των μήλων δεν μας είναι γνωστή. Μπορεί να είναι οποιαδήποτε. Για την άγνωστη αυτή κατανομή γνωρίζουμε μόνο τη μέση τιμή της  $\mu = 50\text{ gr}$  και την τυπική απόκλισή της  $\sigma = 10\text{ gr}$ .

Μπορούμε με αυτά τα δεδομένα να απαντήσουμε στα ερωτήματα που θέσαμε; Η απάντηση είναι ναι και ας δούμε πώς. Τα ερωτήματά μας μπορούν να επαναδιατυπωθούν ως εξής: Ποιο ποσοστό (κατά προσέγγιση) των δειγματικών μέσων α) είναι μεγαλύτερο των 51.25gr και β) είναι μικρότερο των 49gr.

Είναι φανερό ότι για να μπορέσουμε να απαντήσουμε στα ερωτήματα που θέσαμε πρέπει να γνωρίζουμε την κατανομή των δειγματικών μέσων. Το Κ.Ο.Θ. μας βεβαιώνει ότι, παρότι δε γνωρίζουμε την κατανομή των βαρών των μήλων, εντούτοις, γνωρίζουμε την κατανομή των δειγματικών μέσων αφού τα δείγματά μας έχουν μέγεθος αρκετά μεγάλο (400). Δηλαδή, αρκεί μόνο, το ότι γνωρίζουμε τη μέση τιμή και την τυπική απόκλιση της κατανομής των βαρών των μήλων.

Έτσι, αν συμβολίσουμε με  $X$  την τυχαία μεταβλητή που εκφράζει το βάρος ενός τυχαία επιλεγμένου μήλου της καλλιέργειας, και με  $\bar{X}$  την τυχαία μεταβλητή που εκφράζει τους δειγματικούς μέσους, το Κ.Ο.Θ. μας βεβαιώνει ότι η  $\bar{X}$  ακολουθεί κατά προσέγγιση κανονική κατανομή με μέση τιμή  $\mu_{\bar{X}} = \mu = 50$  gr και τυπική απόκλιση  $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{400}} = 0.5$ . Αφού, πλέον, γνωρίζουμε ότι  $\bar{X} \sim N(50, 0.5^2)$ , η απάντηση στα ερωτήματά μας είναι η εξής. Ζητάμε τις πιθανότητες:  $P(\bar{X} > 51.25)$  και  $P(\bar{X} < 49)$ . Έτσι έχουμε:

α)  $P(\bar{X} > 51.25) = 1 - P(\bar{X} \leq 51.25) = 1 - P(Z \leq \frac{51.25 - 50}{0.50}) = 1 - \Phi(2.5) = 0.0062$ . Άρα, το ποσοστό των κιβωτίων που περιέχουν μήλα με μέσο βάρος μεγαλύτερο των 51.25gr είναι κατά προσέγγιση 0.62%.

β)  $P(\bar{X} < 49) = P(Z \leq \frac{49 - 50}{0.50}) = \Phi(-2) = 1 - \Phi(2) = 0.0228$ . Άρα, το ποσοστό των κιβωτίων που περιέχουν μήλα με μέσο βάρος μικρότερο των 49gr είναι κατά προσέγγιση 2.28%.

### Η συνάρτηση κατανομής της τυποποιημένης κανονικής κατανομής

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{t^2}{2}} dt$$

$$\Phi(-z) = 1 - \Phi(z)$$



Ο πίνακας τιμών της  $\Phi(z)$  στο διάστημα  $z \in [0, 3.49]$  με βήμα 0.01. Η τιμή της  $\Phi(z)$ , για παράδειγμα, στο  $z = 0.35 = 0.3 + 0.05$  είναι η τιμή του πίνακα που αντιστοιχεί στην 4η γραμμή (που βρίσκεται το 0.3) και 6η στήλη (που βρίσκεται το 0.05). Οπότε  $\Phi(0.35) = 0.6368$ ,  $\Phi(2.39) = \Phi(2.3 + 0.09) = 0.9916$  κ.ο.κ.

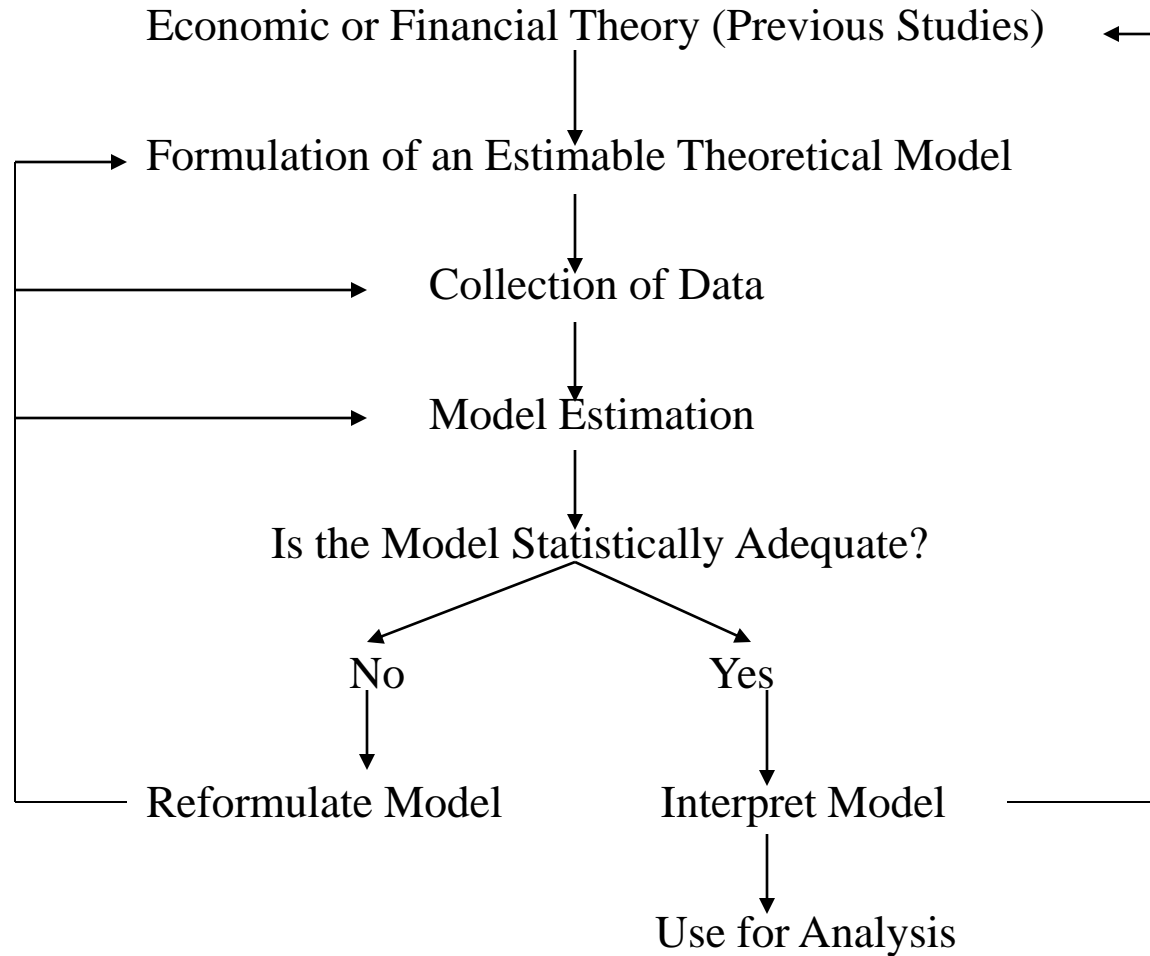
$z$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5398	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

# Γενικές Παρατηρήσεις

## Παραδείγματα Ερευνητικών Ερωτημάτων

- ❑ Πως διακυμαίνεται ο γενικός δείκτης του χρηματιστηρίου μίας χώρας σε σχέση με τα μακροοικονομικά μεγέθη της χώρας;
- ❑ Πως αντιδρά η τιμή της μετοχής μίας εισηγμένης εταιρείας μετά την ανακοίνωση για την καταβολή μερίσματος από την εταιρεία;
- ❑ Ποια είναι η επίδραση της συναλλαγματικής ισοτιμίας μίας χώρας εάν τα επιτόκια βάσης αυξηθούν;

# Βήματα κατά τη διάρκεια ανάπτυξης ενός οικονομετρικού μοντέλου (steps in formulating an econometric model)



# Κριτική σκέψη για τις ακαδημαϊκές μελέτες στη Λογιστική και Χρηματοοικονομική

- 1) Η μελέτη που διαβάσετε αναπτύσσει ένα θεωρητικό μοντέλο ή αποτελεί εφαρμογή (εμπειρική διερεύνηση) ;
- 2) Προέρχονται τα δεδομένα από κάποιον αξιόπιστο φορέα; Είναι τα δεδομένα συνεπή; Είναι το δείγμα επαρκές σε ότι αφορά την αντιπροσωπευτικότητα και το μέγεθος; Είναι οι μεταβλητές που ορίζουν οι συγγραφείς της μελέτης ξεκάθαρες και σωστά ορισμένες;
- 3) Διενεργούν οι συγγραφείς ελέγχους παραβίασης των υποθέσεων του μοντέλου που έχουν εκτιμήσει; (π.χ. στην κλασσική γραμμική παλινδρόμηση υπάρχουν αρκετοί τέτοιοι έλεγχοι όπως θα δούμε).
- 4) Πως ερμηνεύονται τα αποτελέσματα της μελέτης; Μήπως οι συγγραφείς υπερβάλουν σε ότι αφορά τη σημαντικότητα των αποτελεσμάτων; Μήπως οι συγγραφείς ερμηνεύουν τα αποτελέσματα με ανορθόδοξο τρόπο; Απαντούν τα αποτελέσματα το ερευνητικό ερώτημα της μελέτης (σε επαρκή βαθμό);



# Πηγές δεδομένων

- <http://www.bls.gov> - Bureau of Labor Statistics
- <http://www.federalreserve.gov> - Federal Reserve Board
- <http://research.stlouisfed.org/fred2> - Federal Reserve Bank of St. Louis
- [http://www.nationwide.co.uk/hpi/datadownload/data\\_download.htm](http://www.nationwide.co.uk/hpi/datadownload/data_download.htm) - Nationwide
- <http://www.oanda.com/convert/fxhistory> - Oanda
- <http://finance.yahoo.com> - Yahoo! Finance
- <http://www.dallasfed.org/> - Federal Reserve of Bank of Dallas
- <http://www.bankofengland.co.uk/Pages/home.aspx> - Bank of England
  
- Βάσεις Δεδομένων με Συνδρομή: Thomson Reuters Eikon, Bloomberg, etc.