

# Γεωγραφική Ανάλυση

---

## 1 - Στατιστική Ανάλυση

Δρ. Δημήτρης Καβρουδάκης  
Αν. Καθηγητής Γεωγραφικής Ανάλυσης  
[dimitrisk@aegean.gr](mailto:dimitrisk@aegean.gr)

[www.dimitrisk.gr](http://www.dimitrisk.gr)



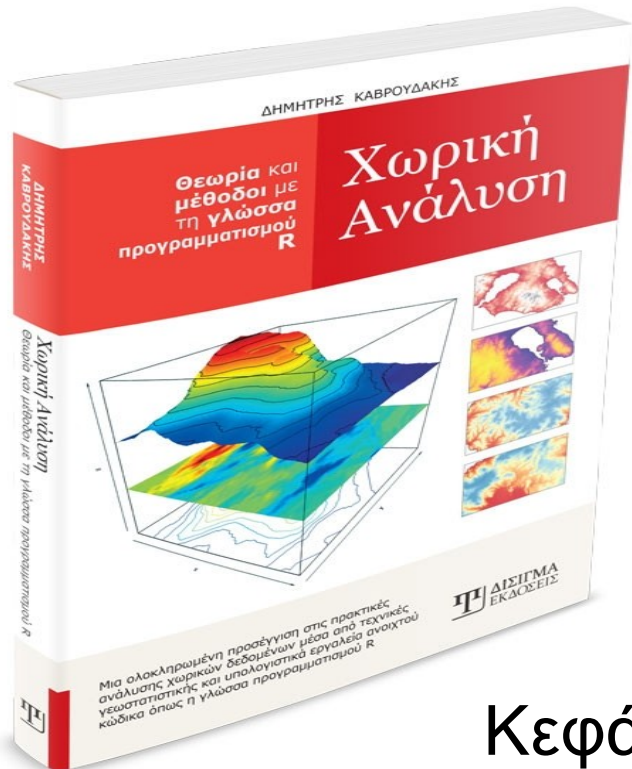
Πανεπιστήμιο Αιγαίου  
Τμήμα Γεωγραφίας

**SAG*i*SRS**  
ΧΑ-ΣΓΠ-ΤΗ

**Spatial Analysis, Geographical Information Systems and Remote Sensing**  
Ερευνητική ομάδα Χωρικής Ανάλυσης, Συστημάτων Γεωγραφικών Πληροφοριών και Τηλεπισκόπησης

# Δημήτρης Καβρουδάκης (2020) Χωρική Ανάλυση

Εκδόσεις ΔΙΣΙΓΜΑ, ISBN13: 978-618-202-021-0



Κεφάλαιο: 1  
Σελίδες: 1 - 26



# 1.1 Περιγραφική Στατιστική

- Βασικό εισαγωγικό κομμάτι της επιστήμης της Στατιστικής.
- Μεθόδους και εργαλεία για την παρουσίαση αριθμητικών δεδομένων μέσω:
  - **σύνοψης, ομαδοποίησης, απεικόνισής**
- πίνακες και τα γραφήματα: ραβδόγραμμα, θηκόγραμμα, διασποροδιάγραμμα
- Στατιστικά Μέτρα τα οποία είναι αριθμητικοί υπολογισμοί που περιγράφουν την κεντρική τάση / μεταβλητότητα των δεδομένων.
- Πληθυσμός και Δείγμα:
  - **Πληθυσμός:** σύνολο όλων των δυνατών στοιχείων ή μετρήσεων που είναι δυνητικά εφικτό να παρατηρηθούν στη μελέτη μιας μεταβλητής: όλοι οι κάτοικοι της Ελλάδας
  - **Δείγμα:** ένα υποσύνολο στοιχείων ή μετρήσεων από ένα πληθυσμό: οι κάτοικοι μιας γειτονιάς
- Χρησιμοποιούμε δείγματα πληθυσμού ώστε στη συνέχεια να γενικεύσουμε τα συμπεράσματα που βγάλαμε από το δείγμα, σε επίπεδο πληθυσμού
- Οι Μεταβλητές πληθυσμού είναι μεταβλητές η χαρακτηριστικά που περιγράφουν ένα πληθυσμό όπως π.χ., η ηλικία ή το βάρος
- Οι Παράμετροι είναι ποσοτικά μέτρα τα οποία περιγράφουν μια μεταβλητή του πληθυσμού όπως π.χ., μέσος όρος ηλικίας ή ο μέσος όρος του βάρους στην Ελλάδα.
- Στατιστικά Μέτρα είναι ποσοτικά μέτρα που περιγράφουν μια μεταβλητή του δείγματος π.χ., μέσος όρος ηλικίας των κατοίκων μιας γειτονιάς.

- Στατιστική Δειγματοληψία είναι η διαδικασία λήψης ενός αντιπροσωπευτικού δείγματος από έναν πληθυσμό
  - π.χ., εξαγωγή δείγματος φοιτητών από όλους τους φοιτητές στην Ελλάδα.
- **Τυχαίο Δείγμα** ονομάζεται το δείγμα στο οποίο κάθε στοιχείο ή μονάδα έχει την ίδια πιθανότητα να συμπεριληφθεί στο δείγμα χωρίς περιορισμούς και μεροληψία.
- **Στατιστική Εκτίμηση** είναι η βέλτιστη πρόταση σχετικά με την τιμή μιας παραμέτρου πληθυσμού από ένα δείγμα
- **Έλεγχος Υποθέσεων** και στατιστική συμπερασματολογία ονομάζεται η διαδικασία με την οποία γίνεται έλεγχος για το αν τα δεδομένα υποστηρίζουν μία υπόθεση σχετικά με την τιμή (ή εύρος τιμών) μιας παραμέτρου του πληθυσμού.



- Τυχαίο δείγμα ?:

- (a) **Επιλογή όλων των** αφιχθέντων ανδρών στο αεροδρόμιο Μυτιλήνης για 1 ημέρα
- (b) **Επιλογή 1** ατόμου κάθε 10 αφιχθέντων στο αεροδρόμιο Μυτιλήνης για 30 ημέρες
- (c) **Επιλογή 10** περαστικών ατόμων απο τη γειτονιά Χρυσομαλλούσα στη Μυτιλήνη
- (d) **Επιλογή 40** γυναικών και 60 ανδρών από τη προκυμαία Μυτιλήνης (πληθυσμός Μυτιλήνης 40%-60%)



## 1.2 Ιστόγραμμα Συχνοτήτων

- βασικό γράφημα της Περιγραφικής Στατιστικής
- συνοψίζει τα δεδομένα σε ομάδες και καταμετρά τη συχνότητα εμφάνισης των τιμών
- Απεικονίζει τη συχνότητα των τιμών ανά ομάδες, δίνοντας πληροφορίες σχετικά με την κατανομή των τιμών σε ομάδες με συγκεκριμένο εύρος.
- Το ύψος κάθε ράβδου αντιστοιχεί στη συχνότητα των μετρήσεων σε κάθε διάστημα, λέγεται ιστόγραμμα συχνοτήτων.

$n_k$

$k$



- Το παρακάτω παράδειγμα, παράγει 20 τυχαίους αριθμούς και τους απεικονίζει σε ένα Ιστόγραμμα Συχνοτήτων με 5 μπάρες.
- Η ομαδοποίηση των τιμών έγινε σε 5 ομάδες (διαστήματα)
- Το πλήθος παρατηρήσεων ανά ομάδα  $n_k$  (συχνότητα) παρουσιάζεται στον παρακάτω Πίνακα 1.1.

```
x=sample(1:10, size=20, replace=TRUE)
x
1 7 7 1 3 9 4 10 9 8 4 8 8 8 6 3 9 10 8 9
hist(x, breaks=5, col="cyan", main="Ιστόγραμμα Συχνοτήτων")
```

| Διάστημα | (0-2] | (2-4] | (4-6] | (6-8] | (8-10] |
|----------|-------|-------|-------|-------|--------|
| $n_k$    | 2     | 4     | 1     | 7     | 6      |



- Το προηγούμενο ιστόγραμμα καταμετρά και απεικονίζει τις τιμές ως απόλυτες συχνότητες.
- Υπάρχει η δυνατότητα να υπολογιστούν σχετικές συχνότητες ως προς το σύνολο του δείγματος ( $n=20$ )
- Δηλαδή τα αποτελέσματα να μην είναι ακέραιοι αριθμοί πλήθους, αλλά να είναι ποσοστά εμφάνισης ως προς το σύνολο των παρατηρήσεων
- Ο υπολογισμός της Σχετικής Συχνότητας για κάθε ομάδα , είναι το πηλίκο του πλήθους παρατηρήσεων σε κάποιο διάστημα προς το συνολικός πλήθος παρατηρήσεων  $k$  , όπως φαίνεται και στην παρακάτω εξίσωση 1.1

$$f_k = \frac{n_k}{N}$$

- Το παρακάτω παράδειγμα χρησιμοποιεί τις προηγούμενες αρχικές τιμές για να παράξει ένα ιστόγραμμα σχετικών συχνοτήτων
- Η Σχετική Συχνότητα των παρατηρήσεων (σχετική συχνότητα) ανά ομάδα, παρουσιάζεται στον παρακάτω Πίνακα 1.2.

```
hist(x, breaks=5, col="cyan",  
     main="Ιστόγραμμα Σχετικών Συχνοτήτων",  
     xlab="Τιμές", ylab="Σχετική Συχνότητα",  
     freq = FALSE)
```

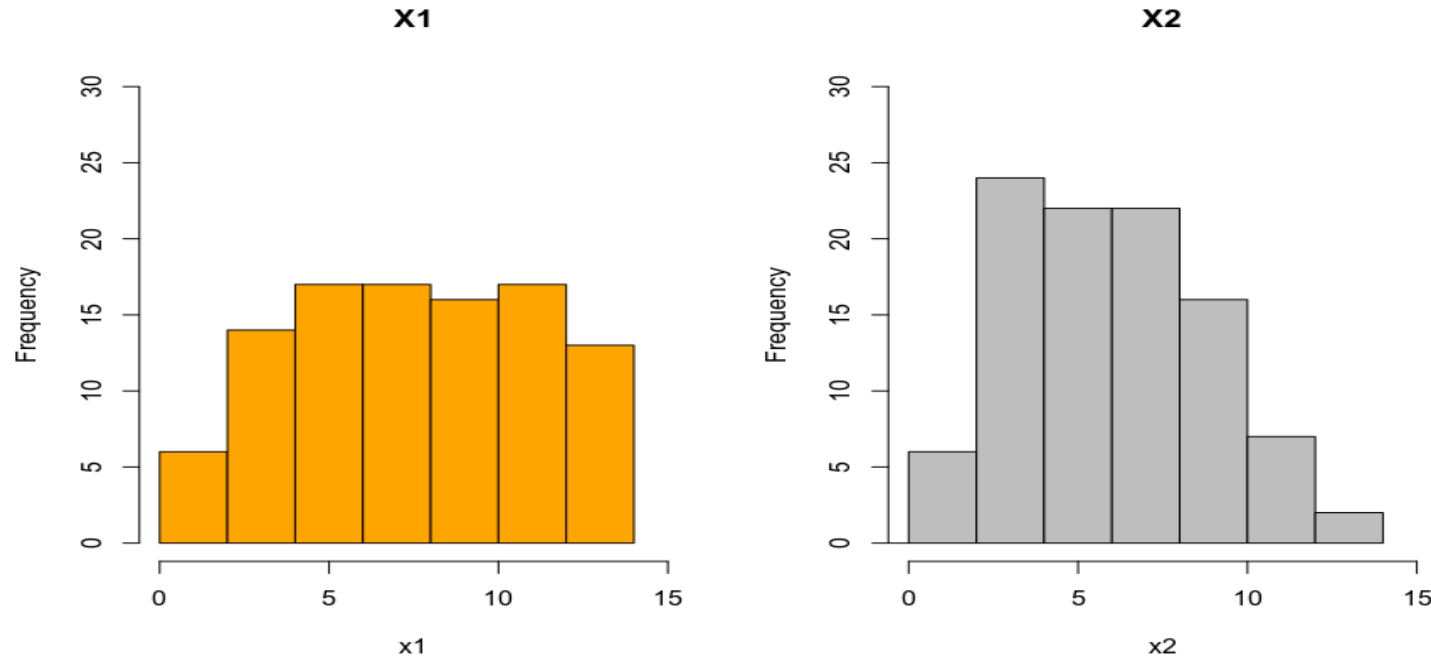
| Διάστημα | (0-2] | (2-4] | (4-6] | (6-8] | (8-10] |
|----------|-------|-------|-------|-------|--------|
| $f_k$    | 0.050 | 0.100 | 0.250 | 0.175 | 0.150  |



- Το ιστόγραμμα, μπορεί να βοηθήσει στην κατανόηση των χαρακτηριστικών του δείγματος καθώς μπορεί να παρέχει χρήσιμες πληροφορίες σχετικά με την κατανομή των τιμών
- Το **σχήμα** του ιστογράμματος μπορεί να φανερώνει την ύπαρξη μιας επικρατούσας τιμής ή την απουσία της από το δείγμα όπως στην Εικόνα 1.1
  - Το σχήμα του ιστογράμματος επηρεάζεται σημαντικά από τον αριθμό των διαστημάτων και τα όρια. Επηρεάζεται από τη θέση και το πλάτος των διαστημάτων που διακρίνουν το εύρος τιμών μιας μεταβλητής.
  - Τις περισσότερες φορές χρησιμοποιούνται μη-αλληλεπικαλυπτόμενα διαστήματα με απλά ορίσματα. Ένας από τους πολλούς κανόνες για την επιλογή του πλήθους των διαστημάτων είναι ο ακόλουθος:

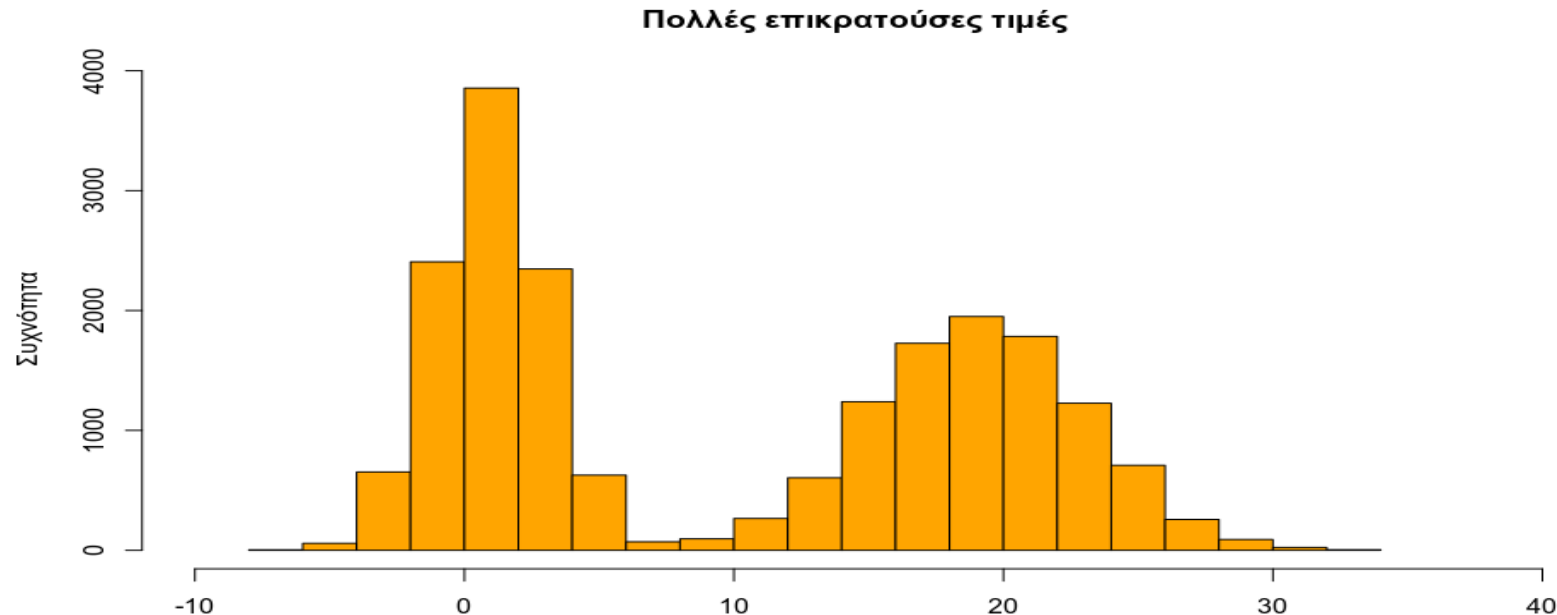
$$5 * \log_{10}(N)$$

- Η **επικρατούσα τιμή** είναι η τιμή του δείγματος η οποία έχει τη μεγαλύτερη συχνότητα. Η τιμή που εμφανίζεται περισσότερες φορές σε σχέση με τις συχνότητες όλων των άλλων τιμών στο δείγμα.



Εικόνα 1.1: Ιστογράμματα δυο κατανομών: **X1** χωρίς επικρατούσα τιμή και **X2** με επικρατούσα τιμή

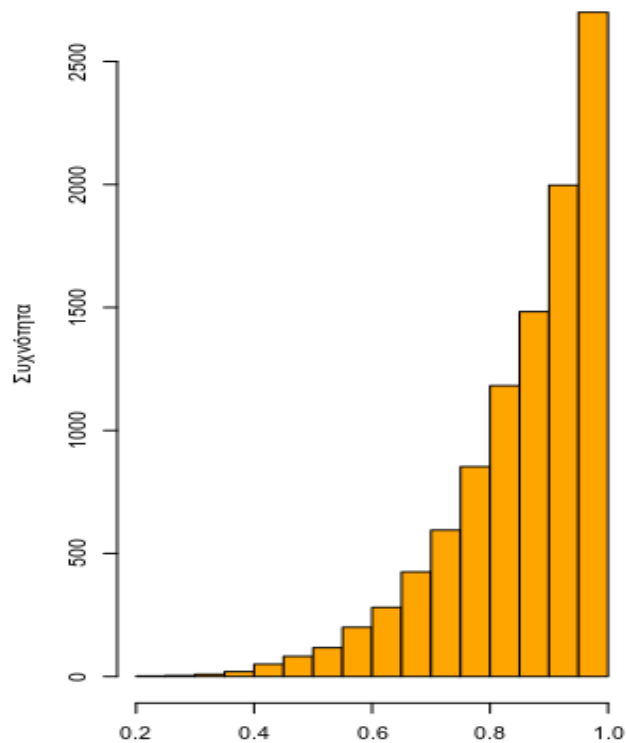
- Κάποιες φορές μπορεί να υπάρχουν περισσότερες από μια **επικρατούσες τιμές** όπως στην Εικόνα 1.2.
- Μπορεί να υποδηλώνει την ύπαρξη περισσότερων από ένα **πληθυσμό** ή δειγματοληψία με διαφορετικά **όργανα** μέτρησης



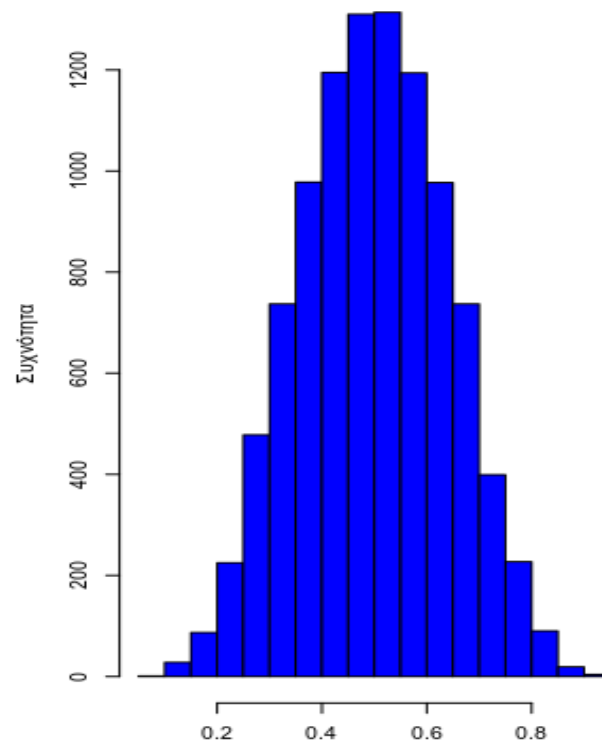
Εικόνα 1.2: Ιστόγραμμα με περισσότερες επικρατούσες τιμές

- Ένα άλλο χαρακτηριστικό των κατανομών είναι η **ασυμμετρία**.
  - Κατανομή των τιμών εκατέρωθεν του μέσου όρου και της επικρατούσας τιμής, με πολύ διαφορετικό τρόπο σε κάθε πλευρά.
- Εικόνα 1.3 παρουσιάζει ένα πολύ χαρακτηριστικό παράδειγμα 3 κατανομών με διαφορετικές ασυμμετρίες.
  - Αριστερά εμφανίζεται μια κατανομή με **αρνητική ασυμμετρία** έχοντας τα κοίλα προς τα αριστερά. Έχει δηλαδή λίγες μικρές τιμές και περισσότερες μεγάλες τιμές
  - Στη μέση εμφανίζεται μια **συμμετρική κατανομή** όπου οι τιμές κατανέμονται σχεδόν ομοιόμορφα εκατέρωθεν του μέσου όρου, έχοντας περισσότερες “μεσαίες τιμές” και λιγότερες μικρές/μεγάλες τιμές
  - Δεξιά εμφανίζεται μια κατανομή με **θετική ασυμμετρία** έχοντας τα κοίλα προς τα δεξιά με πολλές μικρές τιμές και λιγότερες μεγάλες τιμές

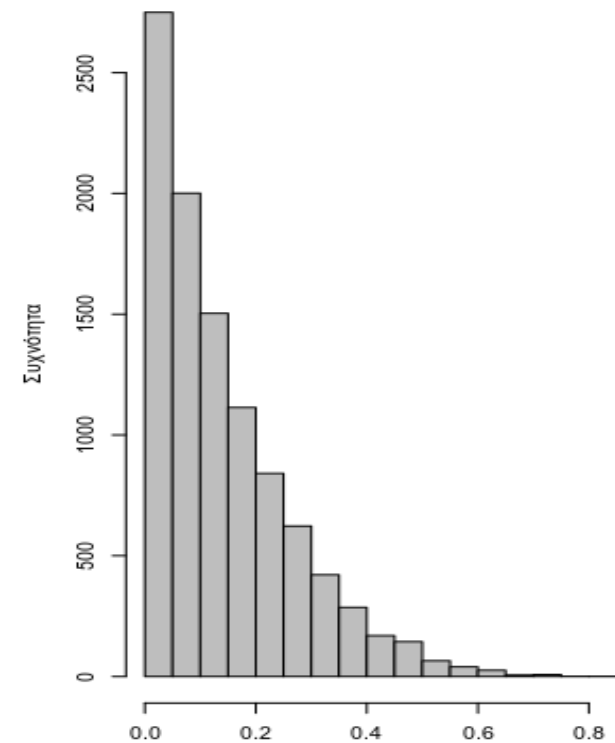
**Αρνητική ασυμμετρία**



**Συμμετρική κατανομή**

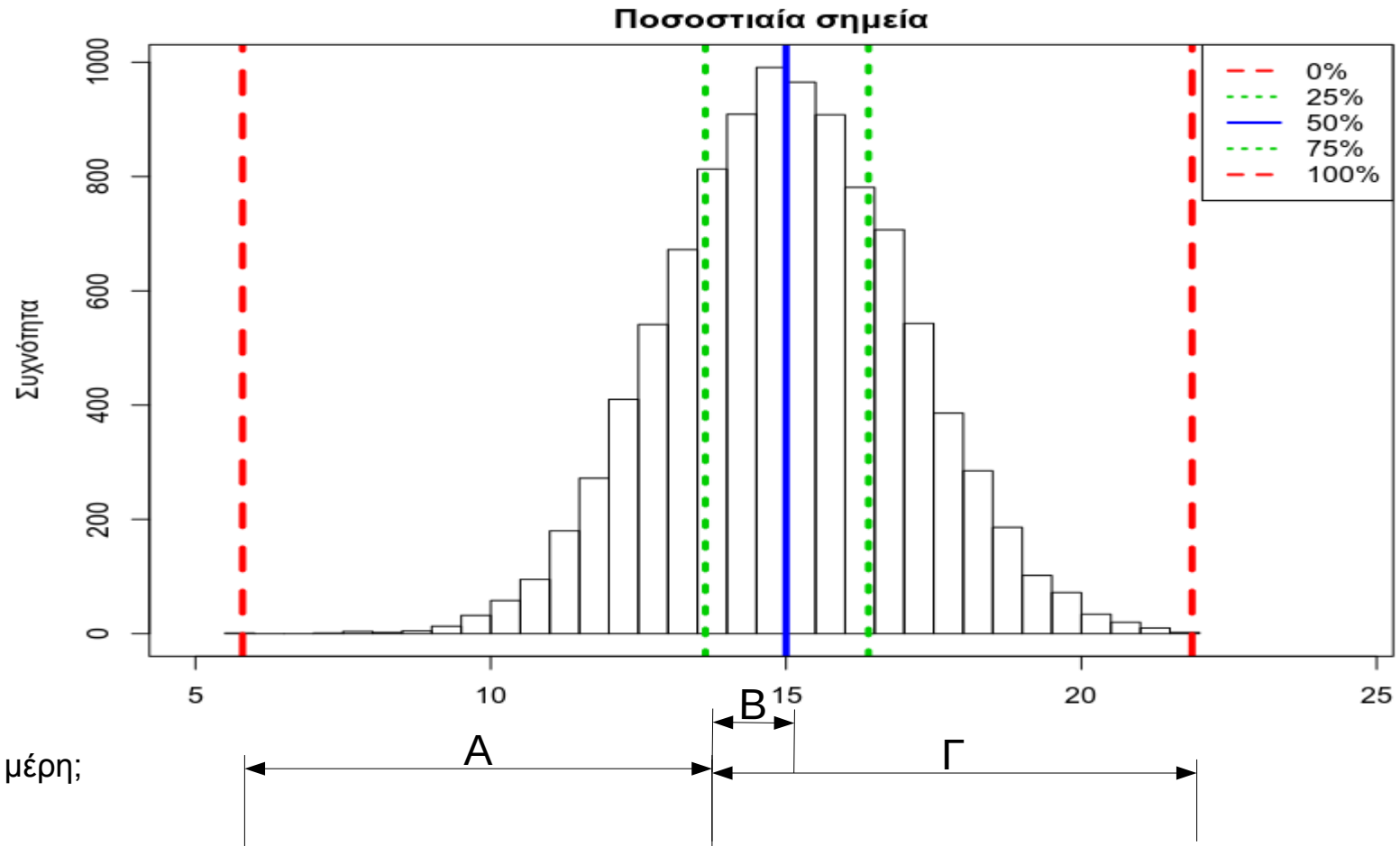


**Θετική ασυμμετρία**



Εικόνα 1.3: Ιστογράμματα με διαφορετικές ασυμμετρίες (αρνητική, συμμετρική, θετική)

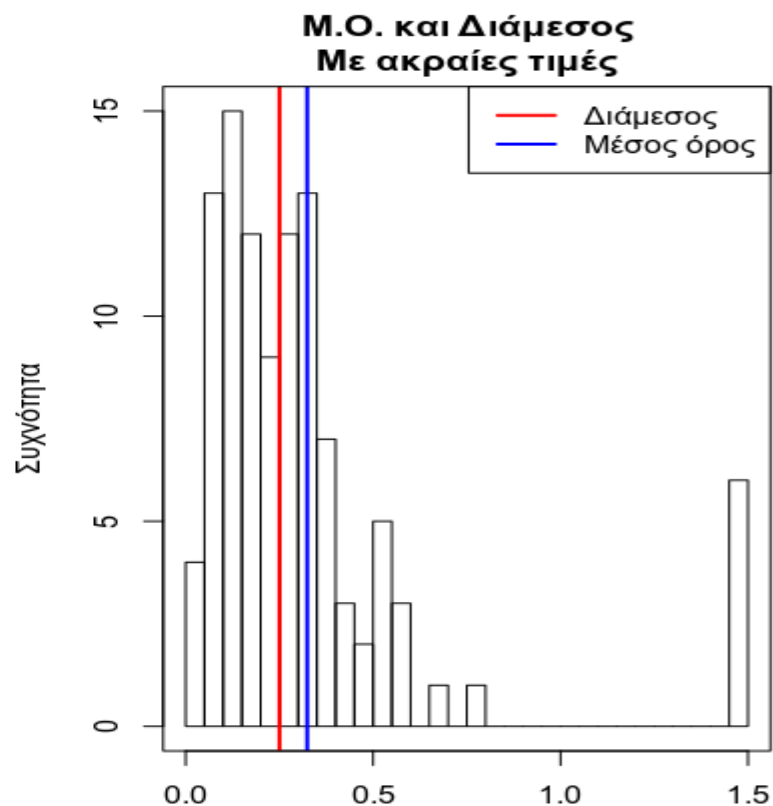
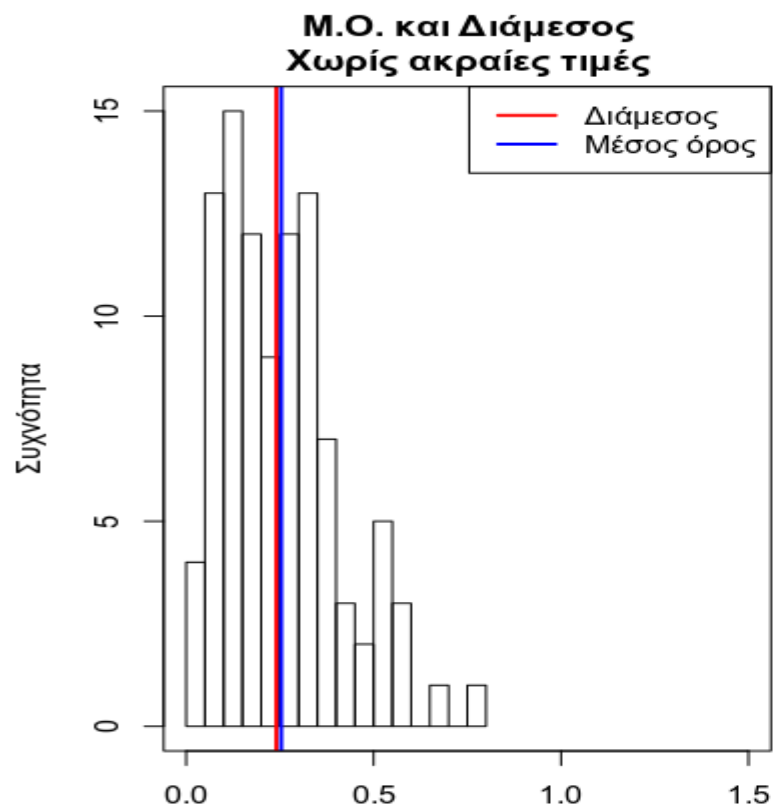
- Ως  $p$ -ποσοστημόριο αι η τιμή μίας μει  $x_p$  λητής που αντιστοιχε  $X$  ε ένα αριθμητικό ποσοστό μετρήσεων του δείγματος οι οποίες είναι  $p$  κρότερες ή ίσες από την τιμή
- Το ποσοστιαία σημεία (ή αλλιώς **ποσοστημόρια**) διαχωρίζουν την κατανομ  $x_p$  ον τιμών σε κομμάτια με βάση το πλήθος των παρατηρήσεων.
  - Το κατώτερο τεταρτημόριο 0%-25% περιλαμβάνει το 25% των παρατηρήσεων.
  - Το ποσοστιαίο σημείο 50% φανεώνει τον διάμεσο της κατανομής τιμών.
- Παραγωγή ποσοστημορίων εναλλακτικά σε:
  - Εκατό εκατοστημόρια (100-μόρια) με το 1% των τιμών
  - Δέκα δεκατημόρια (10-μόρια) με 10% των τιμών
  - Κ.Ο.Κ.



Τι % παρατηρήσεων  
περιλαμβάνουν τα ακόλουθα μέρη;

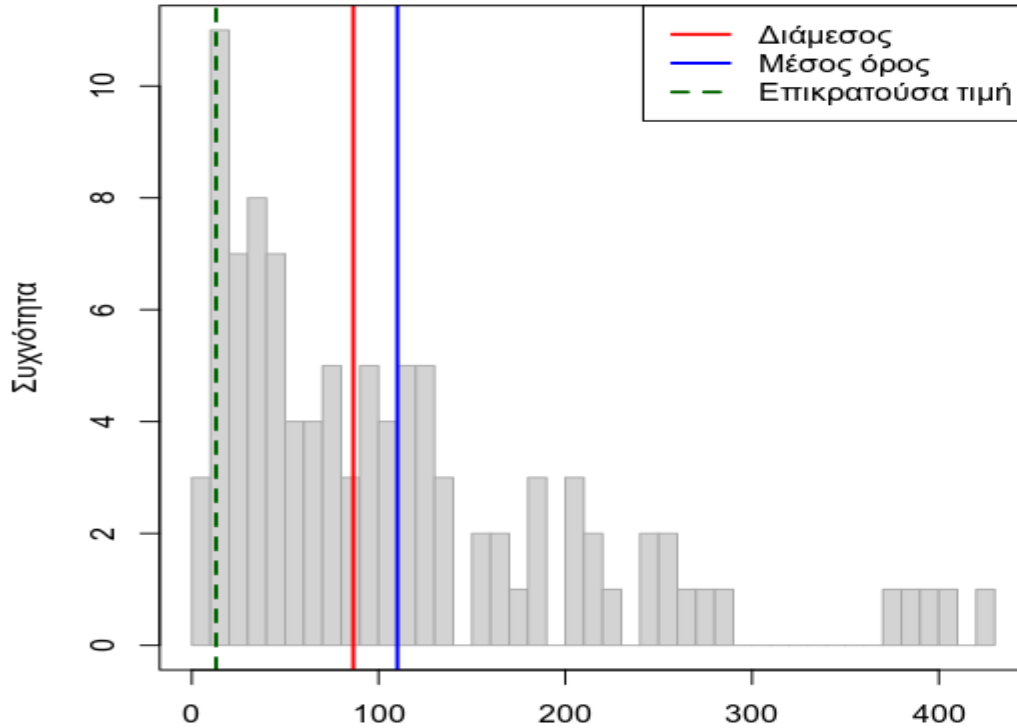
- Στατιστικά Μέτρα τα οποία χαρακτηρίζουν τις κατανομές,
  - για την κεντρικότητα που παρουσιάζουν
- **Διάμεσος** είναι η τιμή , της  $\tilde{x}$  αβλητής που χωρίζει ένα διατεταγμένο σε αύξουσα σειρά δείγμα σε δύο ίσα μέρη.
  - Εναλλακτικά το 50% ποσοστημόριο
- **Επικρατούσα Τιμή** η τιμή της μεταβλητής που εμφανίζεται πιο συχνά στο δείγμα
- **Μέσος Όρος** είναι το ισοσταθμισμένο άθροισμα των τιμών του δείγματος με βάρος
  - Ο μέσος όρος, ενδέχεται να μην αντιστοιχεί σε κάποια από τις τιμές του δείγματος, αλλά μπορεί να είναι κάποια τιμή η οποία μπορεί να περιλαμβάνει δεκαδικά ψηφία

$$m_x = \frac{1}{N} \sum_{i=1}^{i=N} x_i \quad \text{ή} \quad m_x = \sum_{i=1}^{i=N} \frac{1}{N} x_i$$

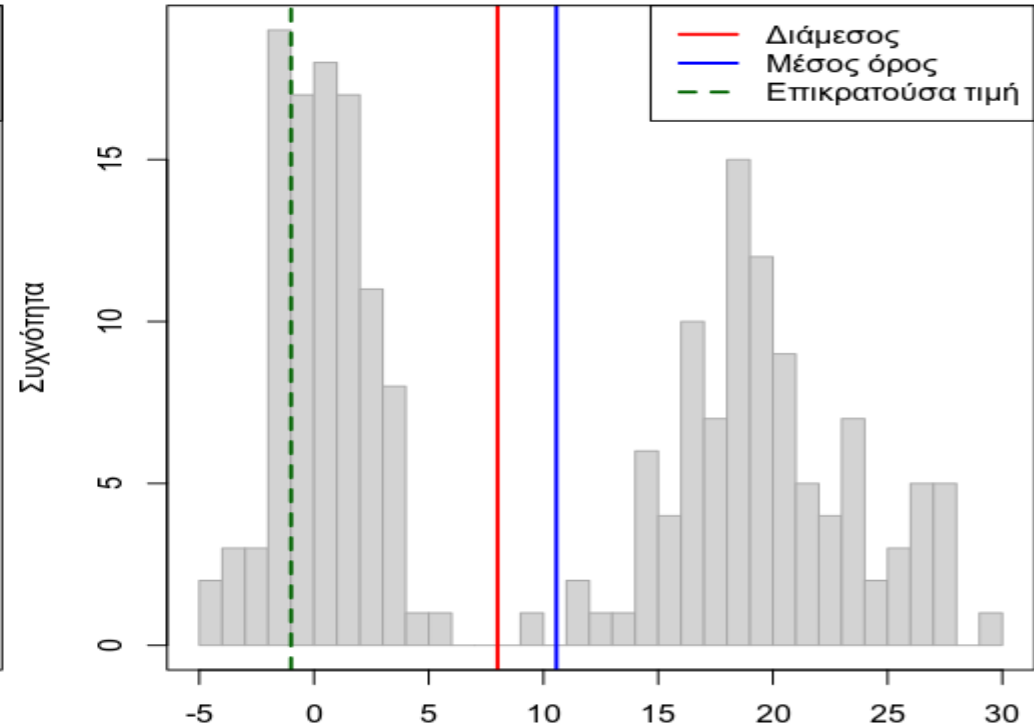


Εικόνα 1.5: Ευαισθησία μέσου όρου και διαμέσου σε ακραίες τιμές της κατανομής ενός δείγματος

### Μ.Ο. και Διάμεσος ασύμετρη κατανομή



### Μ.Ο. Επικρατούσα Τιμή Διάμεσος



Εικόνα 1.6: Μέσος όρος, Διάμεσος και Επικρατούσα τιμή σε διαφορετικές κατανομές

- πληροφορούν σχετικά με τη διασπορά των τιμών του δείγματος.
- **Εύρος**: αποτυπώνει τη διαφορά μεταξύ μέγιστης και ελάχιστης τιμής στο δείγμα
- **Ενδοτεταρτημοριακό Πλ**  $x_{max} - x_{min}$  ιερώνει σχετικά με τη διακύμανση των τιμών γύρω από τη διάμεσο του δείγματος και υπολογίζεται με τη διαφορά μεταξύ ανώτερου και κατώτερου τεταρτημορίου του δείγματος
- **Διακύμανση**: η μέση τετραγωνική απόκλιση των τιμών του δείγματος από το μέσο όρο του δείγματος και η μονάδα μέτρησης της  $x_{0.75} - x_{0.25}$  'άδα μέτρησης της μεταβλητής στο τετράγωνο
- **Τυπική Απόκλιση**: η τετραγωνο 
$$S_x^2 = \frac{1}{N-1} \sum_{i=1}^{i=N} (x_i - m_x)^2$$
 S .

$$S_x = \sqrt{S_x^2}$$

- **Μέση Απόλυτη Απόκλιση** (Mean Absolute Deviation: MAD): ποσοτικοποιεί την απόκλιση των τιμών από το μέσο όρο σε απόλυτες τιμές

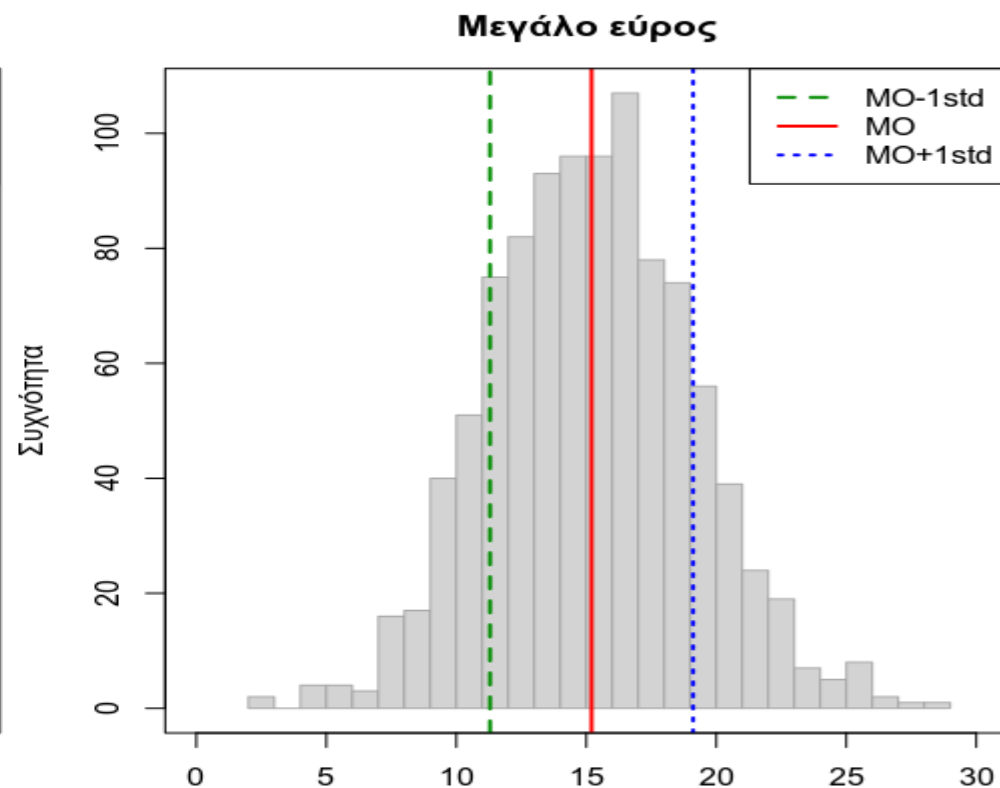
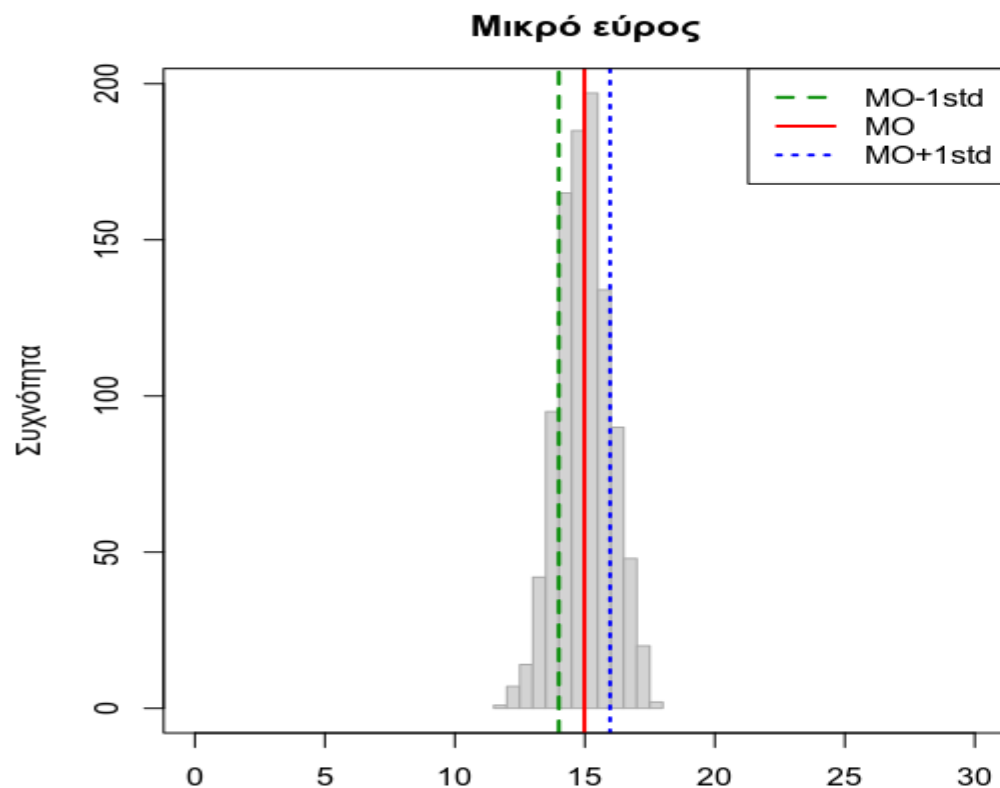
$$MAD = \frac{1}{N-1} \sum_{i=1}^{i=N} |x_i - m_x|$$

- **Συντελεστής Μεταβλητότητας** απόκλισης προς το μέσο όρο του δείγματος

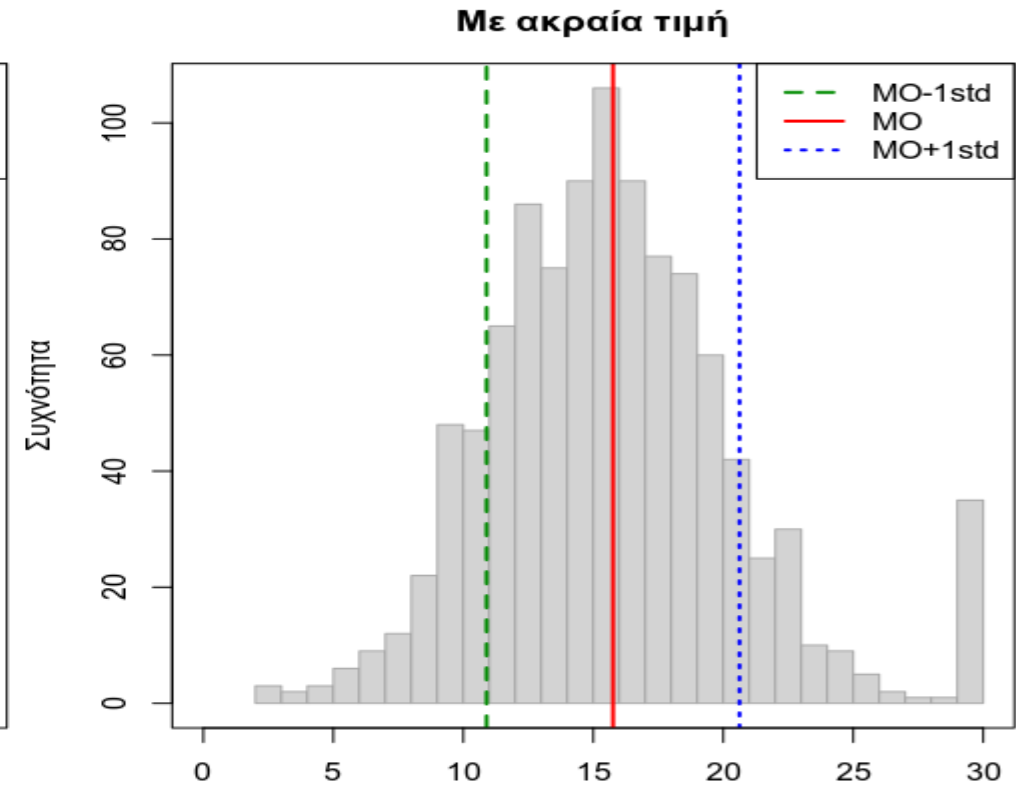
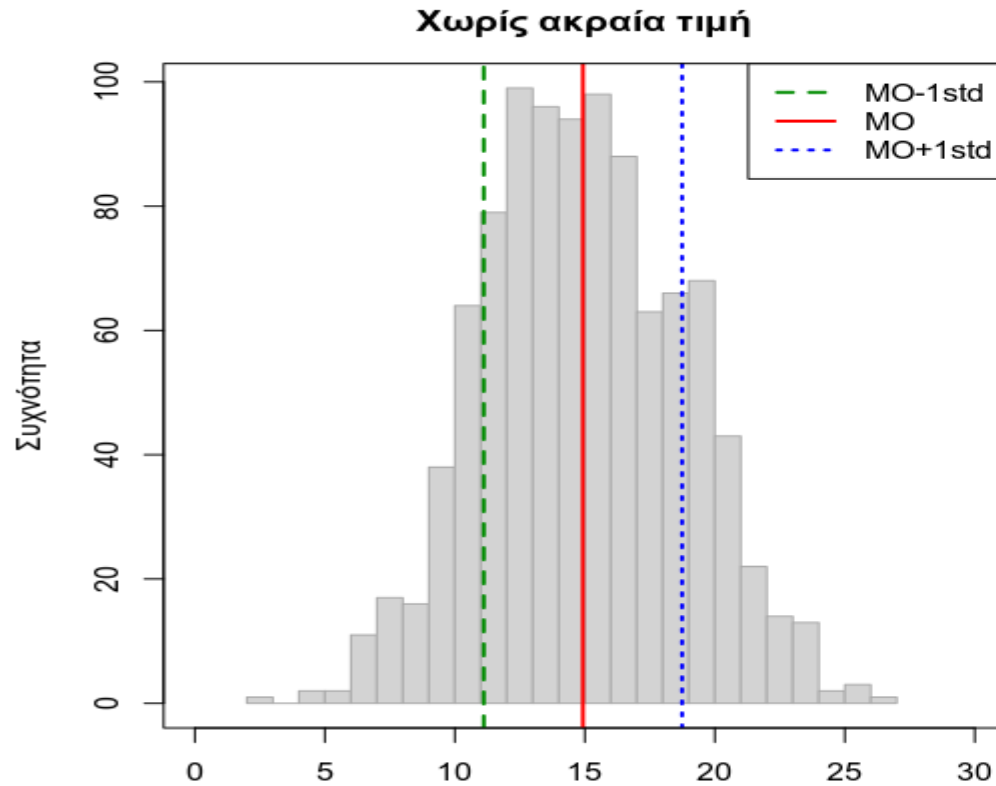
$$CV = \frac{S_x}{m_x} * 100$$

- Τα μέτρα μεταβλητότητας συμβάλλουν σημαντικά στην κατανόηση των **χαρακτηριστικών** των τιμών του δείγματος

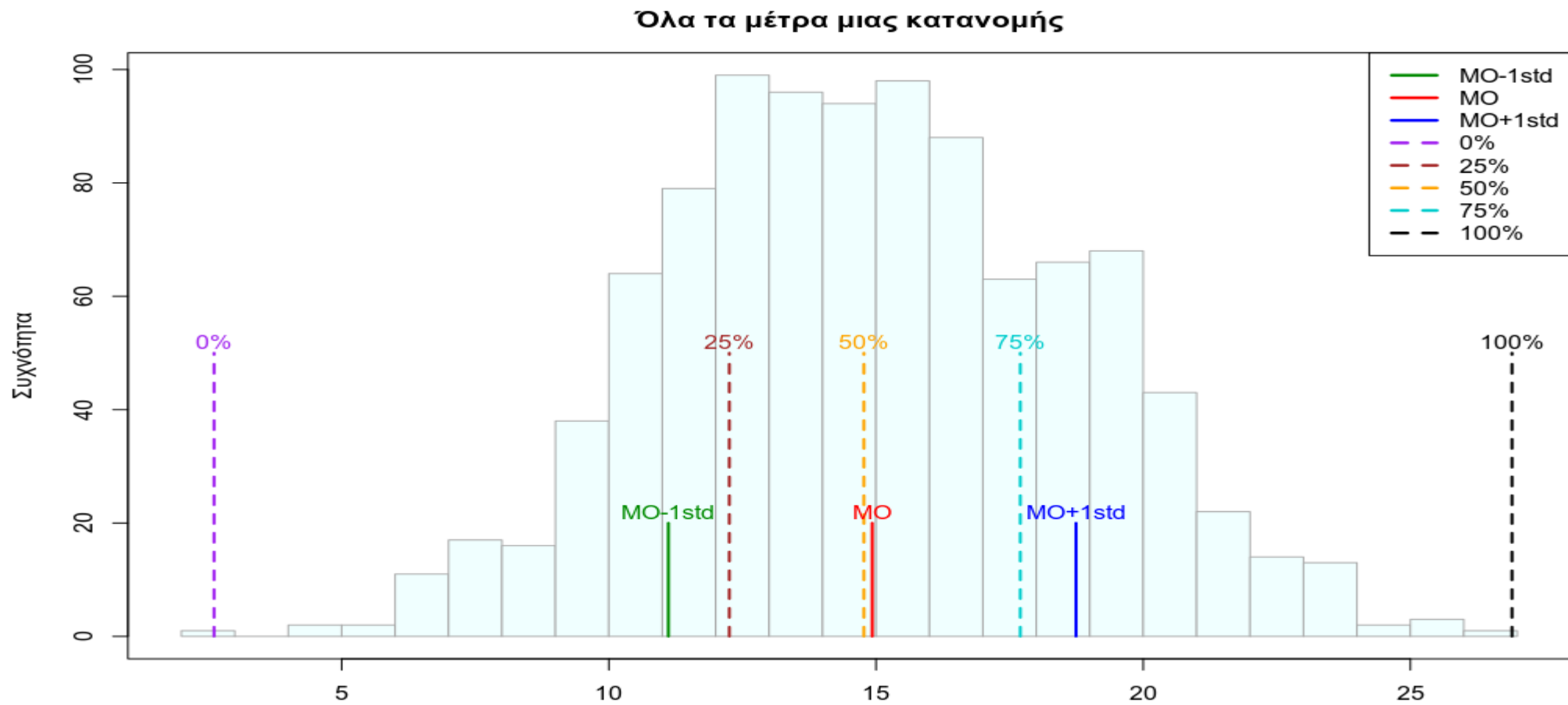
- Τα μέτρα μεταβλητότητας που βασίζονται σε **ποσοστημόρια** καθώς και αυτά που δεν περιλαμβάνουν τετραγωνικές αποκλίσεις, είναι γενικά πιο εύρωστα.



Εικόνα 1.7: Σχέση 1 τυπικής απόκλισης εκατέρωθεν του μέσου όρου για κατανομές με (αριστερά) μικρό εύρος τιμών (δεξιά) μεγάλο εύρος τιμών

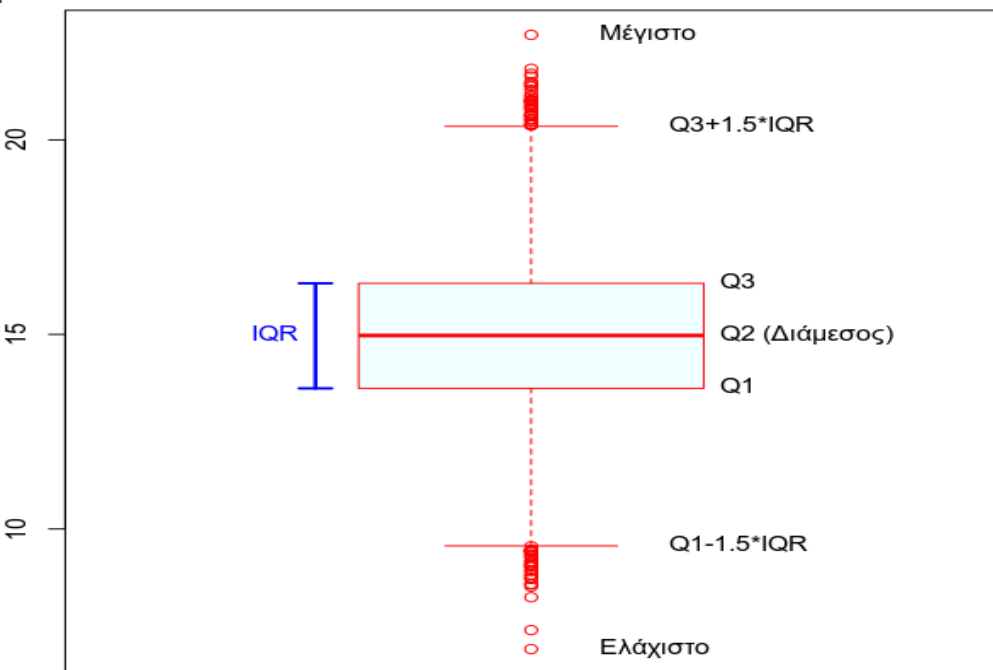


Εικόνα 1.8: Σχέση 1 τυπικής απόκλισης εκατέρωθεν του μέσου όρου για κατανομές χωρίς ακραίες τιμές (αριστερά), με ακραίες τιμές (δεξιά)



Εικόνα 1.9: Όλα τα μέτρα μιας κατανομής σε ένα γράφημα

- Είδος γραφήματος το οποίο βοηθάει να κατανοήσουμε τα στοιχεία της κατανομής ενός δείγματος



- απεικονίζει τα τεταρτημόρια και τις ακραίες τιμές σύμφωνα με τον κλασσικό ορισμό χαρακτηρισμού ακραίων τιμών:
- Ακραίες τιμές ως σημ  $1.5 * IQR$
- Οριζόντιες γραμμές τα ποσοστημόρια:
  - 25% (Q1), 50% (Q2), 75% (Q3)
- Κάποιες τιμές χαρακτηρίστηκαν ως ακραίες καθώς βρίσκονται πέραν του κλασσικού ορίου:
  - 1.5 φορές το IQR

## 1.7 Αθροιστική Σχετική Συχνότητα

- Βασικό στοιχείο της περιγραφικής Στατιστικής είναι η έννοια της αθροιστικής συχνότητας
- Τη χρησιμοποιούμε σε περιπτώσεις όπου μας ενδιαφέρει η ποσοτικοποίηση του πλήθους παρατηρήσεων που είναι μέχρι κάποια τιμή της μεταβλητής.
- η Αθροιστική Σχετική Συχνότητα που αντιστοιχεί σε μία τιμή είναι το ποσοστό των μετρήσεων στο δείγμα, οι οποίες έχουν τιμή μικρότερη ή ίση από
- Κάθε αθροιστική συχνότητα ανήκει στο  $F(x)$  ημα και αν ισχύει τότε θα ισχύει και

$$x < \acute{x}$$

$$F(x)$$

$$[0,1]$$

$$F(x) < F(\acute{x})$$

$$F(x) = \frac{1}{N} \sum_{i=1}^{i=N} c(x_i)$$

$$c(x_i) = 1 \quad \text{αν} \quad x_i \leq x$$

$$c(x_i) = 0 \quad \text{αν} \quad x_i > x$$

```

x = c(1,4,7,5,9,4,4)
x
1 4 7 5 9 4 4
x2 = x[order(x)] # αύξουσα διάταξη
x2
1 4 4 4 5 7 9
sum(x2) # σύνολο
34
cumsum(x2) # αθροιστική συχνότητα
1 5 9 13 18 25 34
df = data.frame(x2, cumsum=cumsum(x2))
df$asxs = 1:length(x2) / length(x2) # αθροιστική σχετική συχνότητα
df

```

| x2 | cumsum | asxs      |
|----|--------|-----------|
| 1  | 1      | 0.1428571 |
| 4  | 5      | 0.2857143 |
| 4  | 9      | 0.4285714 |
| 4  | 13     | 0.5714286 |
| 5  | 18     | 0.7142857 |
| 7  | 25     | 0.8571429 |
| 9  | 34     | 1.0000000 |

## 1.8 Συνάρτηση αθροιστικής κατανομής

- Η αντιστοίχιση των αθροιστικών σχετικών συχνοτήτων

$$\{F(x_i), i = 1 : N\}$$

που ορίζονται για μετ  $N$  ρεις μίας μεταβλητής

με τις  $\{x_i, i = 1 : N\}$

διατεταγμένες σε αύξουσα σειρά, λέγεται **Συνάρτηση Αθροιστικής Κατανομής**

- Cumulative Distribution Function ή CDF

```
P = ecdf(x2) # Συνάρτηση εμπειρικής αθροιστικής Κατανομής του x2
```

```
P(0)
```

```
0
```

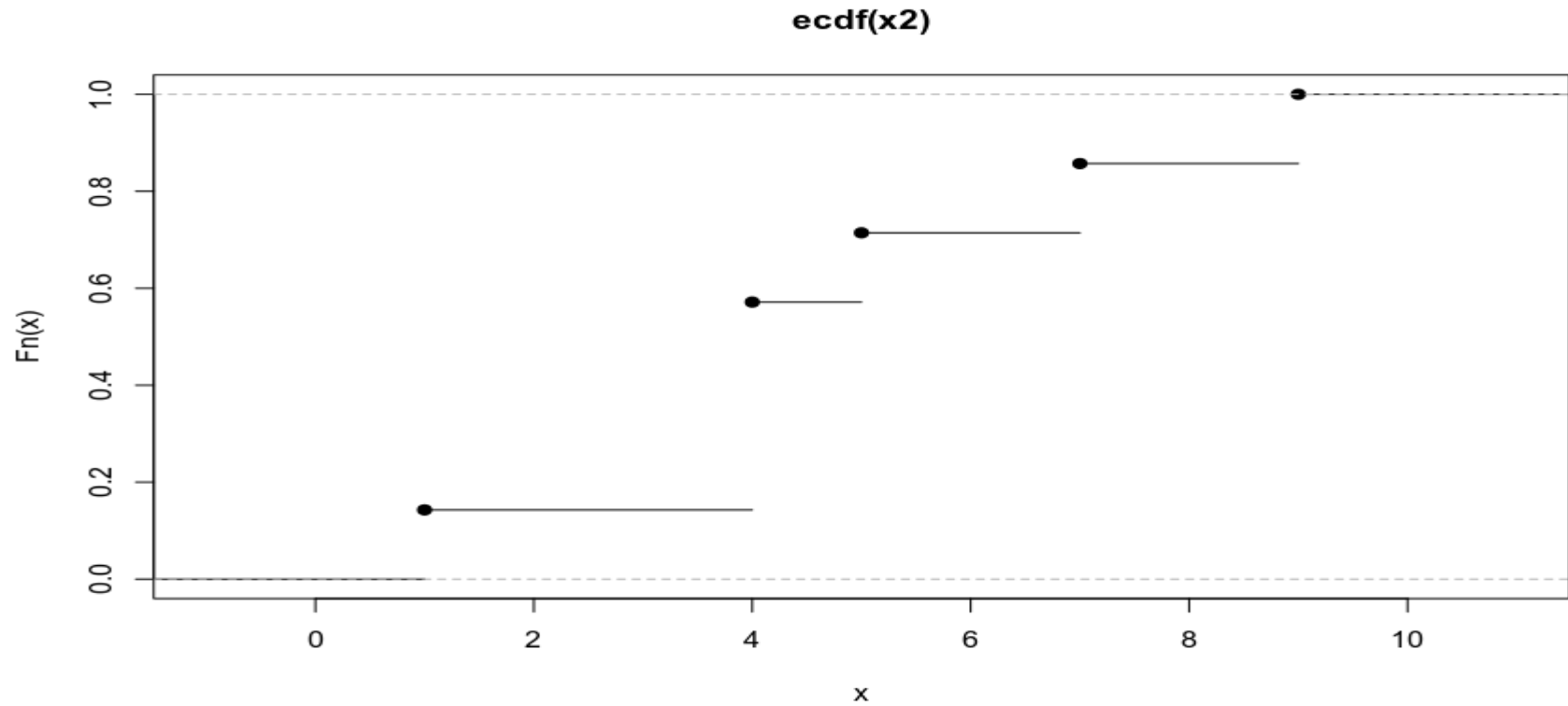
```
P(3)
```

```
0.1428571
```

```
P(6)
```

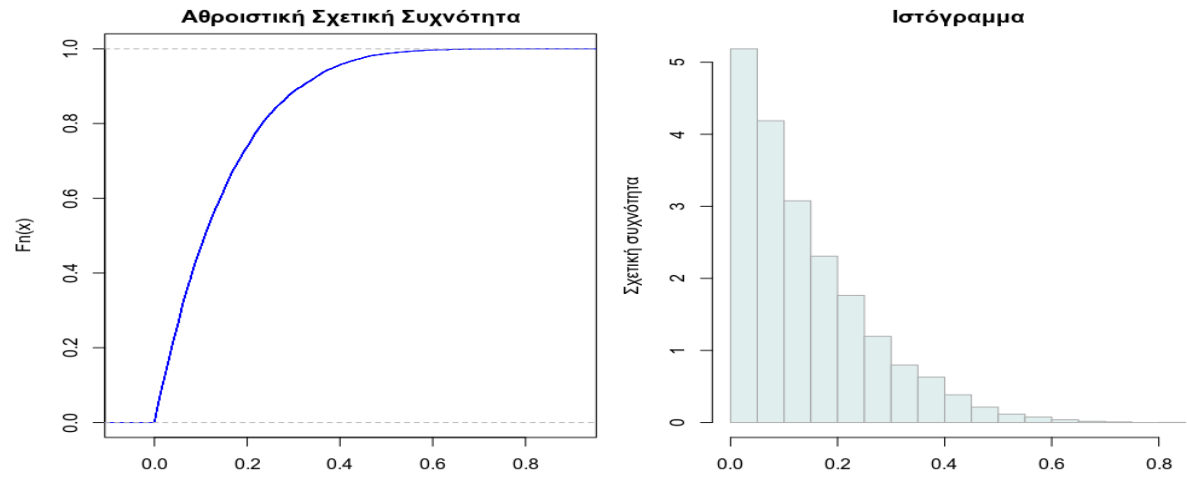
```
0.7142857
```

```
plot(P)
```

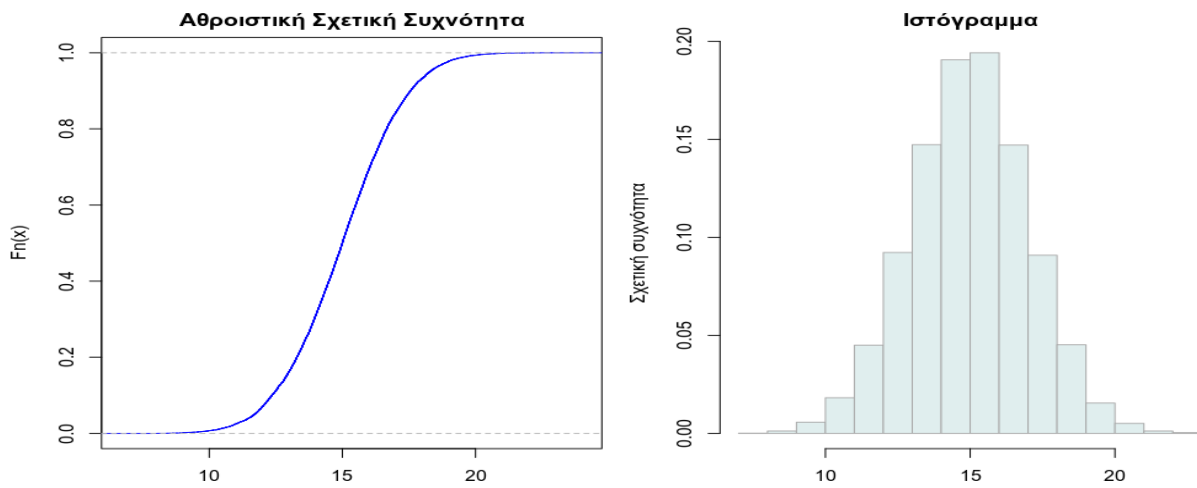


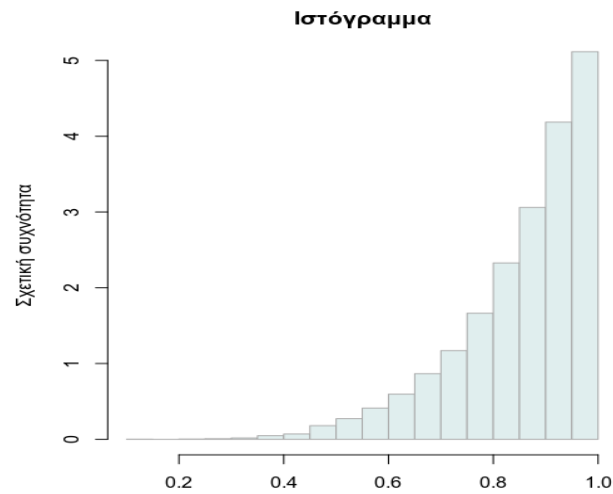
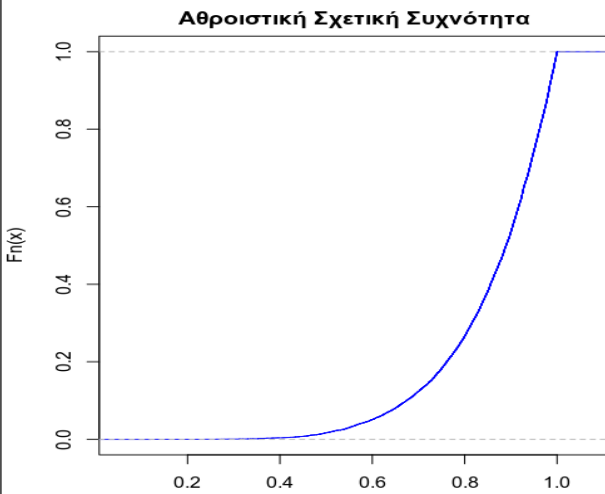
- Χρησιμοποιώντας τη καμπύλη αθροιστικής κατανομής μιας μεταβλητής, μπορούμε να μελετήσουμε την κατανομή των τιμών
- Μπορούμε δηλαδή να χρησιμοποιήσουμε την εξίσωση της αθροιστικής κατανομής και να “ρωτήσουμε” σχετικά με την πιθανότητα εμφάνισης μιας τιμής μικρότερη ή ίση με μια τιμή.
- Ένα οποιοδήποτε σημείο επάνω στην καμπύλη της συνάρτησης αθροιστικής κατανομής έχει δυο συντεταγμένες.
  - Στον οριζόντιο άξονα (τετμημένη) την τιμή της μεταβλητής, και στο κάθετο άξονα (τεταγμένη) το ποσοστό των μετρήσεων στο δείγμα που έχουν τιμή μικρότερη ή ίση του .
  - Το ποσοστό αντιστοιχεί στο αθροιστικό εμβαδό εκείνων των ράβδων του ιστογράμματος πυκνότητας που βρίσκονται στα αριστερά της τιμής .
- Παρακάτω παρατίθενται μια σειρά από ζευγάρια γραφημάτων διαφορετικών κατανομών τιμών
- Αριστερά η καμπύλη της συνάρτησης αθροιστικής κατανομής και στη δεξιά το ιστόγραμμα σχετικών συχνοτήτων
- Όπως φαίνεται και στις εικόνες 1.11 έως 1.15, αυξημένη σχετική συχνότητα στο ιστόγραμμα εμφανίζεται ως μεγαλύτερη κλίση της καμπύλης στη συνάρτηση αθροιστικής κατανομής.

Εικόνα 1.11:  
Κατανομή με **θετική ασυμμετρία** ή λοξότητα

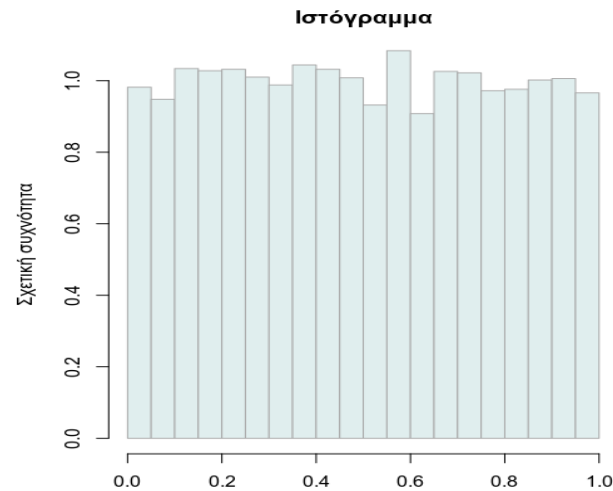
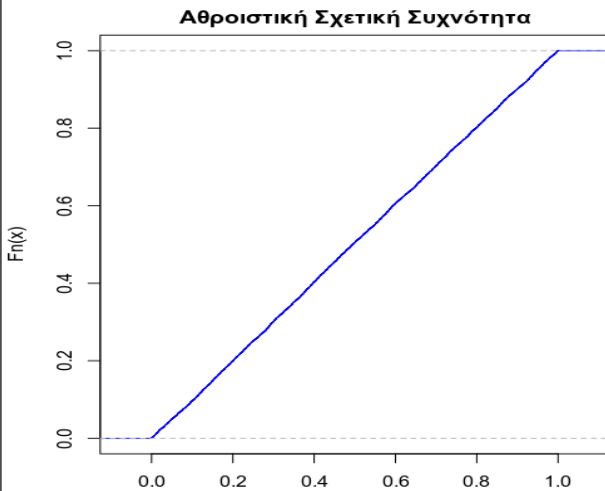


Εικόνα 1.12:  
Κατανομή με **συμμετρική** κατανομή και σαφή επικρατούσα τιμή

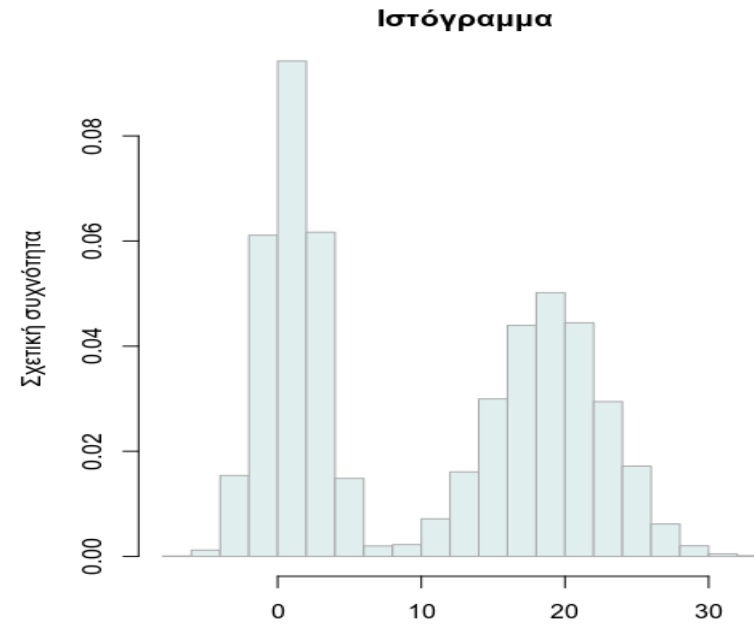
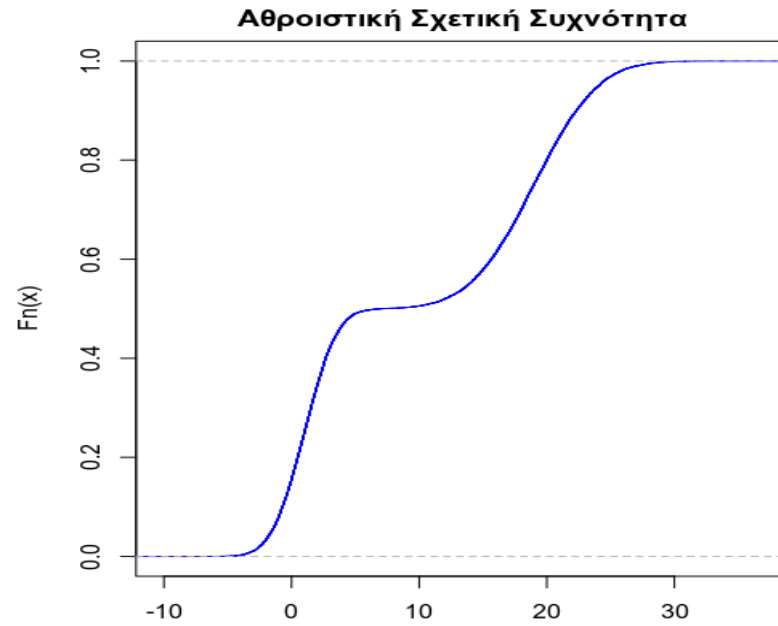




Εικόνα 1.13:  
Κατανομή με αρνητική ασυμμετρία



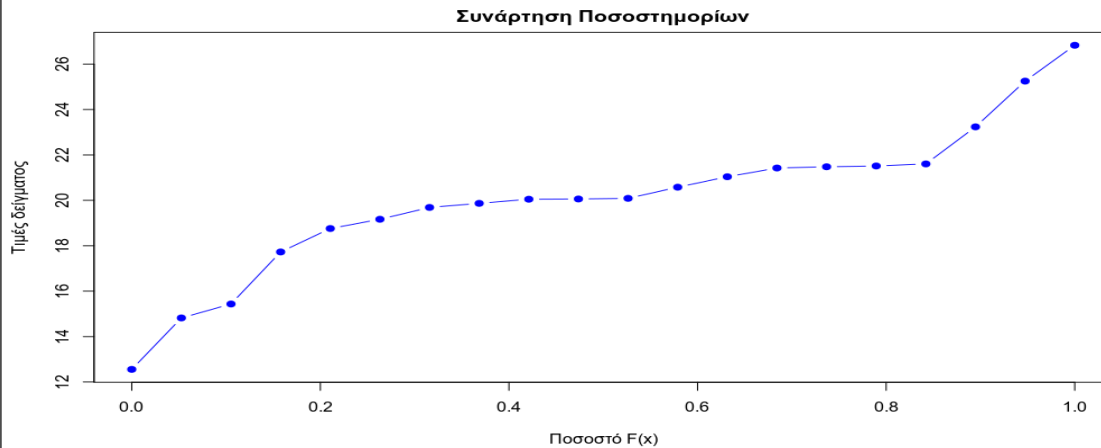
Εικόνα 1.14:  
Κατανομή χωρίς σαφή επικρατούσα τιμή (mode)



Εικόνα 1.15: Κατανομή με **δύο** επικρατούσες τιμές

## 1.9 Συνάρτηση Ποσοστημορίων

- Η έννοια των ποσοστημορίων που είδαμε (1.3) μπορεί να μετασχηματιστεί σε συνάρτηση.
- **Ποσοστημόριο** είναι η τιμή  $x_p$  μιας μεταβλητής που αντιστοιχεί σε ένα αθροιστικό ποσοστό  $p$ .
- Η αντιστοιχία αθροιστικών ποσοστών  $p$  με τα ποσοστημόρια τους  $x_p$  ονομάζεται **συνάρτηση ποσοστημορίων**:  
$$x_p = Q(p)$$
- Η αντιστοιχία αυτή ονομάζεται και αντίστροφη **συνάρτηση αθροιστικής κατανομής**:
- Το παρακάτω παράδειγμα οπτικοποιεί  $x_p = F^{-1}(p)$  α 1.16, τη συνάρτηση ποσοστημορίων για ένα δείγμα 20 τιμών με μέσο όρο 19 και τυπική απόκλιση 3.
- Χρησιμοποιώντας το γράφημα μπορούμε να “ρωτήσουμε” το ποσοστό των τιμών οι οποίες βρίσκονται μέχρι μια κάποια τιμή της μεταβλητής αλλά και να κατανοήσουμε την κατανομή των τιμών εντός του εύρους του δείγματος.



Εικόνα 1.16:  
Συνάρτηση ποσοστημορίων για ένα δείγμα 20 τιμών

```
n = 20
x1 = rnorm(n, mean=19, sd=3)
quantile(x1)
```

| 0%       | 25%      | 50%      | 75%      | 100%     |
|----------|----------|----------|----------|----------|
| 12.55424 | 19.06386 | 20.07325 | 21.48656 | 26.82857 |

```
horizontal = (1:n-1)/(n-1)
plot( horizontal, sort(x1), type="b", col="blue", pch=19,
      main="Συνάρτηση Ποσοστημορίων", xlab="Ποσοστό F(x)", ylab="Τιμές δείγματος")
```

## 1.10 Κανονικοποίηση μετρήσεων

- Διαδικασία κατά την οποία ένα δείγμα τιμών μιας μεταβλητής μετασχηματίζεται αναλόγως ώστε να υπακούει σε κάποιους ορισμένους κανόνες σχετικά με τα χαρακτηριστικά της κατανομής και να αποκτήσει κάποιες νέες στατιστικές ιδιότητες.
- Η διαδικασία της **κανονικοποίησης** είναι μια μετατροπή των αρχικών τιμών του δείγματος που ακολουθεί μια οποιαδήποτε κατανομή και έχει μέσο όρο και τυπική απόκλιση σε νέες  $x_i$  μέσ οι οποίες τώρα πλέον έχουν μέσο όρο και τυπική απόκλιση  $m_x$   $S_x$
- Η διαδικασία αυτή χρησιμοποιεί την παρακάτω Εξίσωση 1.  $m=0$  ον μετασχηματισμό κάθε παλιάς  $S=1$  σε μια νέα τιμή ως εξής:

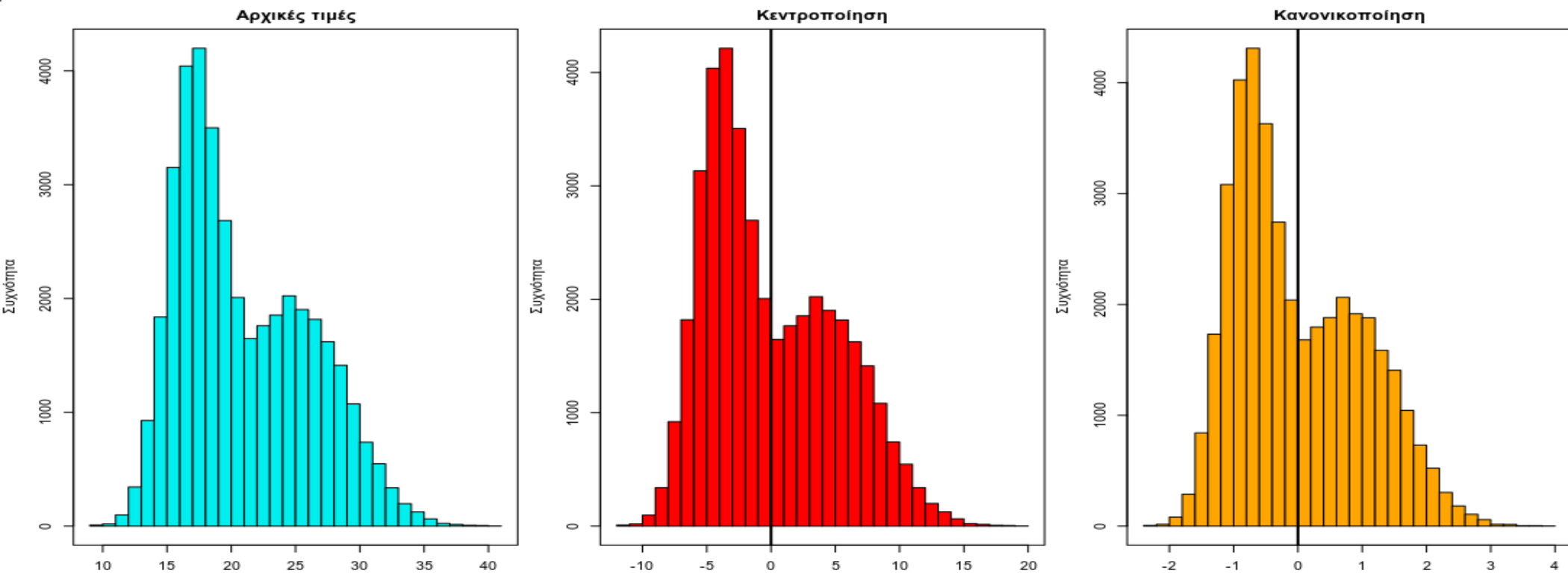
$x_i$

$z_i$

$$z_i = \frac{x_i - m_x}{S_x}$$

Εξίσωση 1.9: Υπολογισμός κανονικοποιημένων τιμών δείγματος

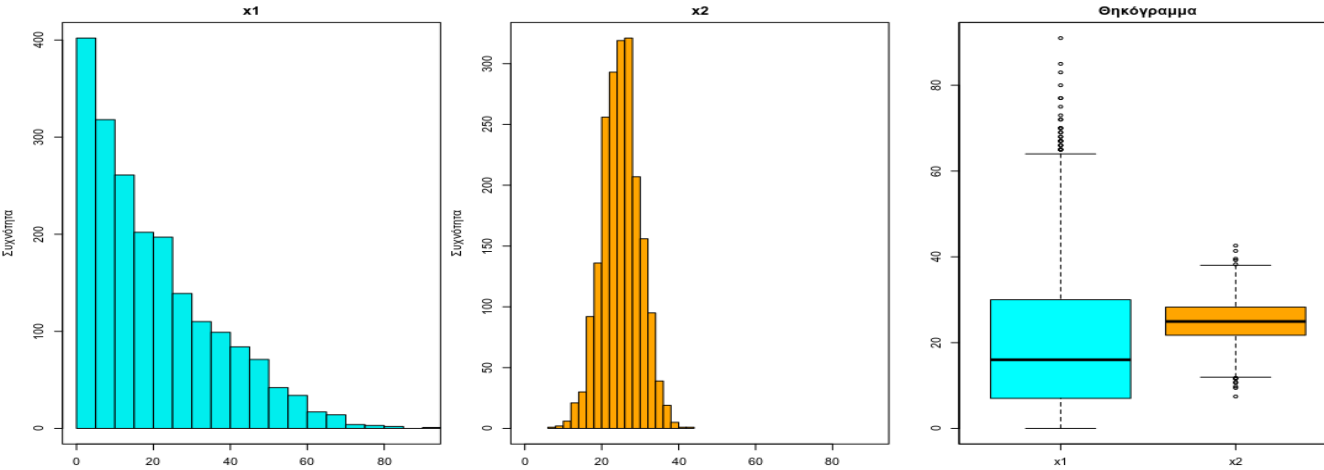
- Οι τιμές  $x_i - m_x$  / αριθμητή ονομάζονται **αποκλίσεις** και έχουν μέσο όρο και τυπική απόκλιση  $m = 0$  και η διαδικασία υπολογισμού  $S_x$  τους λέγεται **κεντροποίηση**.
- Όταν οι κεντροποιημένες τιμές, διαιρεθούν με την τυπική απόκλιση, τότε παράγουν τις **κανονικοποιημένες** τιμές
  - Η κανονικοποιημένη τιμή εκφράζει ουσιαστικά την απόκλιση της αρχικής τιμής από το μέσο όρο σε ποσοστιαίες τιμές επί της τυπικής απόκλισης
- Όπως φαίνεται και στην παράσταση Εικόνα 1.17, είναι ένας χρήσιμος μετασχηματισμός δεδομένων  $S_x$ 
  - για τη **σύγκριση** κατανομών μετρήσεων με διαφορετικές μονάδες.



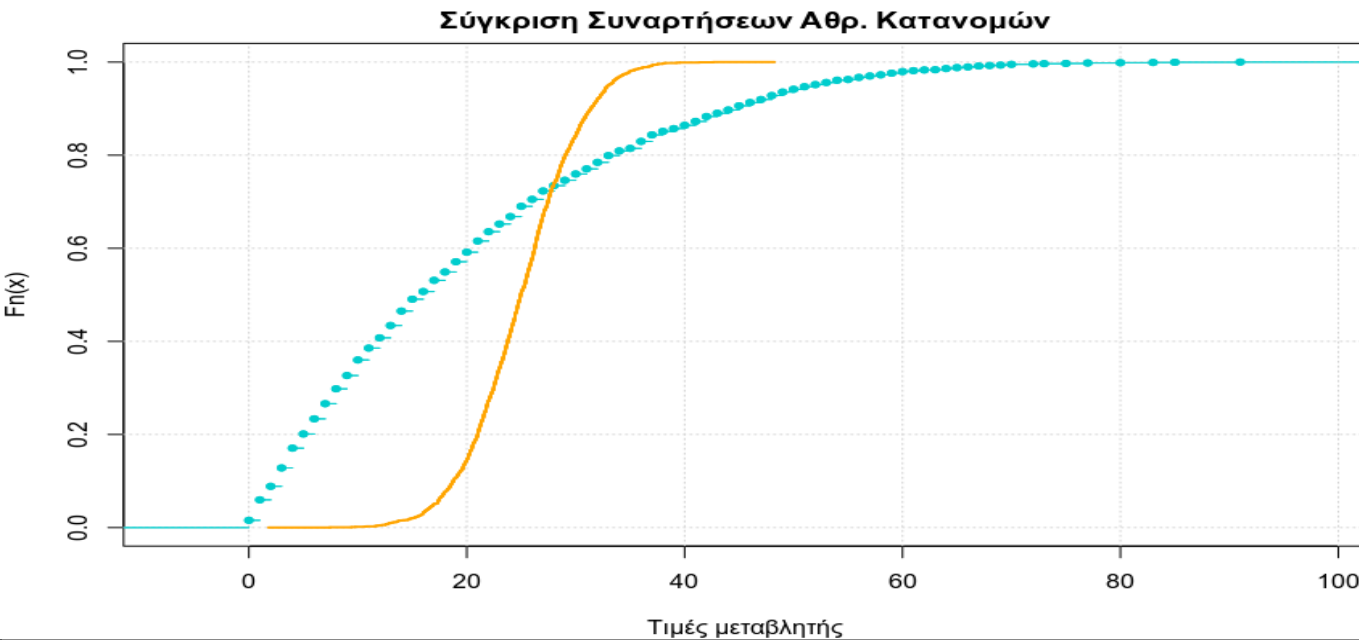
Εικόνα 1.17: Ιστόγραμμα  
 (α) αρχικών τιμών  
 (β) κεντροποιημένων τιμών  
 (γ) κανονικοποιημένων τιμών

- Σύγκριση δυο κατανομών τιμών οι οποίες μπορεί
  - να έχουν καταγράψει είτε σε διαφορετικές περιοχές
  - σε διαφορετικές χρονικές περιόδους
  - να περιλαμβάνουν και διαφορετικό πλήθος παρατηρήσεων.
- Οι παρατηρήσεις, δεν είναι απαραίτητο να έχουν καταγραφεί στις ίδιες θέσεις και έτσι δεν είναι απαραίτητο πως συγκρίνουμε ζευγάρια παρατηρήσεων σε ίδιες θέσεις.
- Η σύγκριση κατανομών μπορεί να γίνει με μια σειρά από εργαλεία όπως τα ιστογράμματα, θηκογράμματα ή ακόμα και με τη σύγκριση στατιστικών μέτρων όπως ο μέσος όρος, διάμεσος και διακύμανση.
- Με οπτική σύγκριση υπάρχει το πρόβλημα της πολυπλοκότητας
- Στην παρακάτω Εικόνα 1.18, παρουσιάζετε ένα απλό παράδειγμα σύγκρισης δυο κατανομών
  - με τα αρχικά τους ξεχωριστά ιστογράμματα
  - και ένα κοινό θηκόγραμμα

- Ένας ακόμα τρόπος σύγκρισης κατανομών είναι και η χρήση της **Συνάρτησης Αθροιστικής Κατανομής**.
- Η παράλληλη απεικόνιση δυο η περισσότερων εμπειρικών συναρτήσεων αθροιστικών σχετικών συχνοτήτων στο ίδιο διάγραμμα μπορεί να μας βοηθήσει να κατανοήσουμε την κατανομή των συχνοτήτων σε κάθε μια από τις κατανομές
  - όπως φαίνεται και στην παρακάτω Εικόνα 1.19.

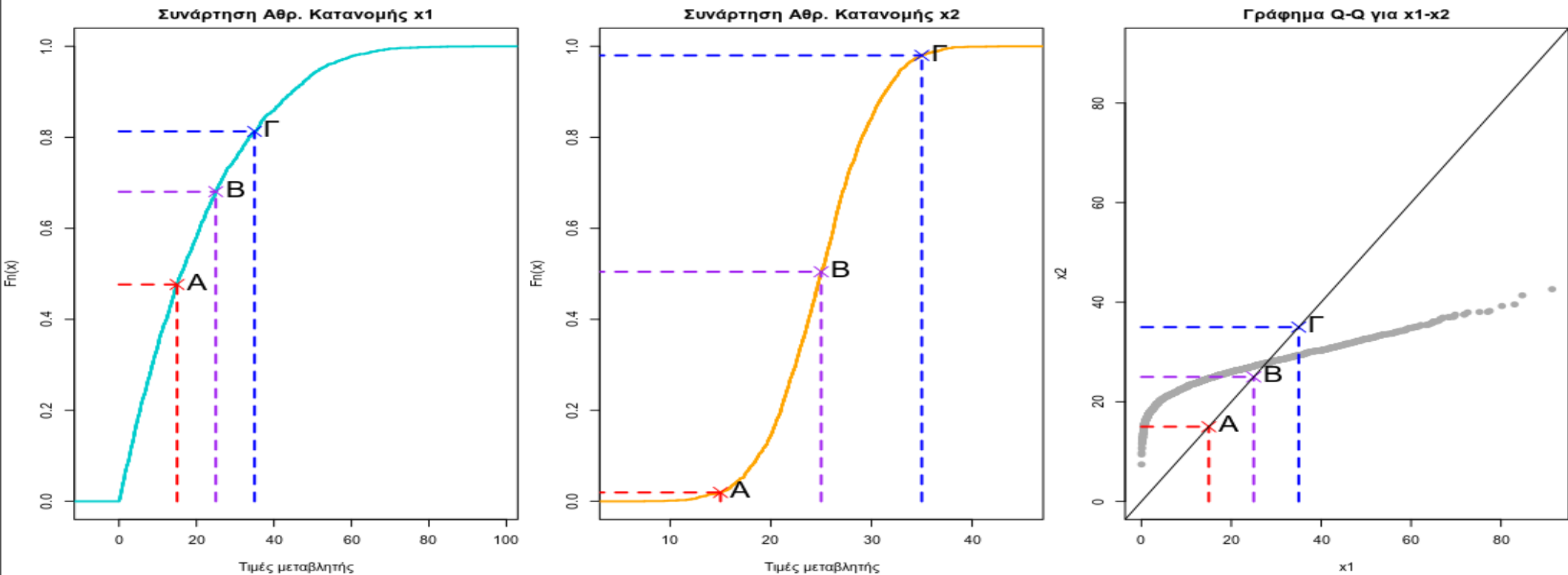


Εικόνα 1.18:  
Σύγκριση δυο κατανομών ( $x_1$ ,  $x_2$ ) με τα αρχικά  
ιστογράμματα τους και ένα κοινό θηκόγραμμα



Εικόνα 1.19:  
Σύγκριση δυο Συναρτήσεων Αθροιστικών  
Κατανομών ( $x_1$ ,  $x_2$ )

- Ένας άλλος τρόπος σύγκρισης δυο κατανομών είναι η χρήση ενός διασπορογράμματος ποσοστημορίων κάθε δείγματος τα οποία θα αντιστοιχούν στα ίδια αθροιστικά ποσοστά ώστε να μπορούμε να συγκρίνουμε.
- Ένα τέτοιο γράφημα ποσοστιαίων σημείων δυο κατανομών ονομάζεται **γράφημα Q-Q** και η διαδικασία κατασκευής του είναι η ακόλουθη.
  - Διακριτοποίηση του εύρους πιθανοτήτων σε ίσα διαστήματα π.χ. [0.5, 0.10, ... , 0.95, 1]
  - Υπολογισμός ποσοστιαίων σημείων κάθε δείγματος που αντιστοιχούν στις παραπάνω πιθανότητες
    - π.χ. για το σύνολο η πιθανότητα 0.15 οδηγεί στην τιμή δείγματος 23 ενώ η ίδια πιθανότητα για το άλλο σύνολο μπορεί να οδηγεί σε διαφορετική τιμή δείγματος 76.
- Το γράφημα Q-Q είναι ουσιαστικά η γραφική αντιστοίχιση των ποσοστιαίων σημείων που ορίζονται από το ίδιο ποσοστό
  - δηλαδή το γράφημα των ποσοστιαίων σημείων του ενός συνόλου δεδομένων με τα αντίστοιχα (του ίδιου ποσοστού) ποσοστιαία σημεία του άλλου συνόλου δεδομένων.
- Αν τα δυο σύνολο ήταν ακριβώς ίδια, τότε όλες οι τιμές στο γράφημα Q-Q θα εφάπτονταν στη **γραμμή της διαγωνίου των 45°**
  - Όσο αποκλίνουν από αυτή τη γραμμή, τόσο πιο διαφορετικές είναι οι κατανομές τιμών στα ποσοστημόρια του κάθε δείγματος όπως φαίνεται και στην παρακάτω Εικόνα 1.20.



Εικόνα 1.20:  
Σύγκριση δυο **Συναρτήσεων Αθροιστικών Κατανομών** ( $x_1$ ,  $x_2$ ) σε ένα κοινό γράφημα Q-Q και κοινά σημεία σύγκρισης A,B,Γ



dimitrisk@aegean.gr

facebook

Dimitris Kavroudakis



Dimitris Kavroudakis



dimitris\_k



Instagram

d\_kavroudakis



dimitriskavroudakis

**Δρ. Δημήτρης Καβρουδάκης**  
Αν. Καθηγητής Γεωγραφικής  
Ανάλυσης

[www.dimitrisk.gr](http://www.dimitrisk.gr)

