

## **Μοντέλα στην Επιστήμη Τροφίμων 532Ε**

**Διδάσκοντες:**

**Κ. Κουτσουμανής, Ε. Κατσανίδης,**

**Ώρες Μαθήματος**

**Τρίτη 14.00-1600**



**Τομέας Επιστήμης και Τεχνολογίας Τροφίμων  
Σχολή Γεωπονίας, Α.Π.Θ**

## **Μοντέλα στην Επιστήμη Τροφίμων 532Ε**

**Στόχοι Μαθήματος**

- Κατανόηση βασικών αρχών στατιστικής ανάλυσης δεδομένων
- Εκμάθηση βασικών εφαρμογών στατιστικής ανάλυσης στο πρόγραμμα Excel
- Εφαρμογές στα τρόφιμα

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Πρόγραμμα Διαλέξεων

1. Στατιστικές έννοιες. Περιγραφική στατιστική. Μέσος όρος, διάμεσος, τυπική απόκλιση, διακύμανση, εύρος.
2. Περιγραφική στατιστική - εφαρμογές στον Η/Υ
3. Κατανομές πληθυσμών. Κανονική κατανομή, κατανομή  $t$ , κατανομή  $F$ , διωνυμική κατανομή.
4. Όρια εμπιστοσύνης.
5. Students Test ( $t$ -test)
6. Ανάλυση παραλλακτικότητας
7. Ανάλυση παραλλακτικότητας - εφαρμογές στον Η/Υ
8. Γραμμική παλινδρόμηση
9. Γραμμική παλινδρόμηση - εφαρμογές στον Η/Υ
10. Μη-γραμμική παλινδρόμηση
11. Πειραματικός σχεδιασμός
12. Παραγοντική ανάλυση
13. Ατελώς ομαδοποιημένα πειράματα, μεθοδολογία επιφάνειας απόκρισης.

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Υλικό Διδασκαλίας

- Διαφάνειες Διαλέξεων
- Διαδίκτυο

### Εξετάσεις

- Ασκήσεις, Εργασίες (20%)
- Τελικές Εξετάσεις (80%)

### Επικοινωνία

- Στέλτε e-mail στο [ekatsani@agro.auth.gr](mailto:ekatsani@agro.auth.gr) με τίτλο (subject) "532Ε"

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Εισαγωγή

Η Στατιστική είναι η επιστήμη των δεδομένων. Μπορεί να ορισθεί ως «η επιστήμη που έχει αντικείμενο την ανάπτυξη μεθόδων για τη συλλογή, παρουσίαση, ανάλυση και ερμηνεία δεδομένων» (Fisher).

Οι στατιστικές μέθοδοι, με κριτήριο τους στόχους που εξυπηρετούν, ταξινομούνται σε τρεις γενικές κατηγορίες:

1. Μέθοδοι για το σχεδιασμό της διαδικασίας συλλογής των δεδομένων
2. Μέθοδοι για τη συνοπτική και την αποτελεσματική παρουσίαση των δεδομένων
3. Μέθοδοι για την εξαγωγή συμπερασμάτων

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Εισαγωγή

Με βάση αυτή την ταξινόμηση, η Στατιστική διαρθρώνεται σε τρεις κλάδους:

1. Σχεδιασμός Πειραμάτων (Experimental Design) και Θεωρία Δειγματοληψίας (Sampling Theory)

Είναι κλάδοι της Στατιστικής που, σε γενικές γραμμές, έχουν ως αντικείμενο την ανάπτυξη μεθόδων για τη συλλογή δεδομένων μέσω της εκτέλεσης πειραμάτων ή μέσω δειγματοληψιών αντίστοιχα.

2. Περιγραφική Στατιστική (Descriptive Statistics)

Είναι ο κλάδος της Στατιστικής που έχει ως αντικείμενο την ανάπτυξη μεθόδων για τη συνοπτική και την αποτελεσματική παρουσίαση των δεδομένων

3. Στατιστική Συμπερασματολογία (Statistical Inference)

Είναι ο κλάδος της Στατιστικής που έχει ως αντικείμενο την ανάπτυξη μεθόδων για την ανάλυση των δεδομένων και την εξαγωγή συμπερασμάτων για τον πληθυσμό από τον οποίο προέρχονται.

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Σχεδιασμός Πειραμάτων και Θεωρία Δειγματοληψίας

#### Βασικές Έννοιες

- **Πληθυσμός (population)**. Είναι σύνολο υποκειμένων (ατόμων, αντικειμένων, τόπων και γενικότερα οποιονδήποτε οντοτήτων) τα οποία ενδιαφερόμαστε να μελετήσουμε ως προς ένα τουλάχιστον κοινό χαρακτηριστικό. Π.χ μια παρτίδα τροφίμων
- **Μονάδα (unit)** Κάθε στοιχείο του πληθυσμού Π.χ. Συσκευασία
- **Δείγμα (Sample)**. Ένα υποσύνολο του πληθυσμού που μελετάμε Π.χ. 10 συσκευασίες της παρτίδας
- **Μεταβλητή (variable)**. Οποιοδήποτε χαρακτηριστικό του οποίου η τιμή αλλάζει στα στοιχεία του πληθυσμού Π.χ οξύτητα, χρώμα, κλπ
- **Παρατηρήσεις (observations)**. Κάθε μετρούμενη τιμή της μεταβλητής Π.χ 5 μετρήσεις στην κάθε συσκευασία  $\times 10$  συσκ.=50 παρατηρήσεις

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Σχεδιασμός Πειραμάτων και Θεωρία Δειγματοληψίας

Για την εφαρμογή μεθόδων της Στατιστικής ο **πληθυσμός** πρέπει να είναι **καλά ορισμένος**. Πρέπει, δηλαδή, να μπορούμε να αποφανθούμε με σαφήνεια αν κάποια μονάδα ανήκει σε αυτόν ή όχι.

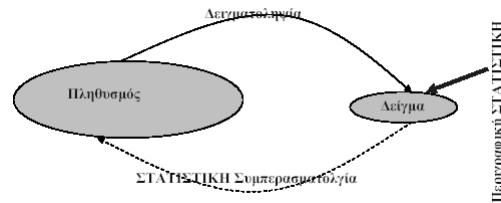
Οι πληθυσμοί διακρίνονται σε **πεπερασμένους** και σε **άπειρους**. Με κριτήριο το πλήθος των μονάδων ενός πληθυσμού, φυσικά, άπειροι πληθυσμοί δεν υπάρχουν. Όμως, σε περιπτώσεις πληθυσμών που έχουν μεγάλο πλήθος μονάδων, για τις ανάγκες της πιθανοθεωρητικής επεξεργασίας των δεδομένων, υποθέτουμε ότι αυτά προέρχονται από, πρακτικά, άπειρους πληθυσμούς.

Ένας τρόπος για να μελετήσουμε ένα φαινόμενο είναι να εξετάσουμε όλες τις μονάδες του πληθυσμού στον οποίο συμβαίνει. Αυτή η μέθοδος συλλογής δεδομένων, η εξέταση δηλαδή όλων των μονάδων του πληθυσμού που μελετάμε, ονομάζεται **ολική απογραφή ή απογραφή**.

Όμως, σε πολλές περιπτώσεις, η απογραφή είναι δύσκολη ή οικονομικά ασύμφορη ή και αδύνατη. Σε αυτές τις περιπτώσεις, όπως ήδη έχουμε αναφέρει, ο ερευνητής εξετάζει ένα μέρος των μονάδων του πληθυσμού, δηλαδή, επιλέγει ένα **δείγμα**, κάνει τις παρατηρήσεις του σε αυτό και στη συνέχεια **γενικεύει τα συμπεράσματά του** για ολόκληρο τον πληθυσμό με τη χρήση της **στατιστικής συμπεριστατολογίας**.

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Σχεδιασμός Πειραμάτων και Θεωρία Δειγματοληψίας



Μεταξύ της επί μέρους εμπειρίας (δείγμα) για τα αντικείμενα που έχουμε και στη γενίκευση πως αυτό που διαπιστώσαμε ισχύει για όλο το σύνολο (πληθυσμός) μεσολαβεί η έννοια της αντιπροσωπευτικότητας.

Έχουμε την απαίτηση το δείγμα να αποτελεί μια μικρογραφία όλων των ιδιοτήτων του πληθυσμού.

Αν το υποσύνολο/δείγμα που μελετήσαμε είναι αντιπροσωπευτικό του συνόλου/πληθυσμού τότε έχουμε το δικαίωμα να αποφανθούμε πως όσα παρατηρήσαμε στο υποσύνολο αυτό ισχύουν για όλο το σύνολο.

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Σχεδιασμός Πειραμάτων και Θεωρία Δειγματοληψίας

#### Επιλογή του Δείγματος

Λάθη στην επιλογή είναι αποτελούν αιτία μεροληψίας.

Μεροληψίες ή συστηματικά σφάλματα: Είναι συστηματικά σφάλματα προς την ίδια κατεύθυνση.

Για να μειωθεί η μεροληψία πρέπει να μειώσουμε την προσωπική επιλογή. Η απάντηση της Θεωρίας Δειγματοληψίας σε αυτό το πρόβλημα είναι η ανάπτυξη μεθόδων επιλογής τυχαίων δειγμάτων.

Τυχαία δειγματοληψία είναι η μέθοδος δειγματοληψίας σύμφωνα με την οποία στην επιλογή του δείγματος καταλήγουμε με δεδομένη πιθανότητα. Στην περίπτωση αυτή το δείγμα λέγεται τυχαίο.

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Σχεδιασμός Πειραμάτων και Θεωρία Δειγματοληψίας

#### Επιλογή του Δείγματος

Στην τυχαία δειγματοληψία ένα δείγμα μεγέθους  $n$  από έναν πληθυσμό λαμβάνεται με τέτοιο τρόπο ώστε κάθε δυνατό δείγμα μεγέθους  $n$  να έχει την ίδια πιθανότητα να εκλεγεί.

Δηλαδή, αν από ένα πληθυσμό μπορούμε να πάρουμε  $k$  το πλήθος δείγματα μεγέθους  $n$ , τότε

$$p_1 = p_2 = p_3 = \dots = p_k = \frac{1}{k}$$

(όπου,  $p$  η πιθανότητα να εκλεγεί το δείγμα  $i$ ).

Ένα δείγμα που επιλέγεται με τέτοιο τρόπο λέγεται απλό τυχαίο δείγμα.

Η τυχαία δειγματοληψία μπορεί να εξαλείψει τη μεροληψία που οφείλεται στη λάθος επιλογή δείγματος αλλά δεν εξαλείφει τη μεταβλητότητα. Η μεταβλητότητα από δείγμα σε δείγμα σε επαναλαμβανόμενα τυχαία δείγματα είναι αναπόφευκτη συνέπεια της μεταβλητότητας που υπάρχει στον πληθυσμό από τον οποίο προέρχονται.

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Σχεδιασμός Πειραμάτων και Θεωρία Δειγματοληψίας

#### Επιλογή του Δείγματος

Η μεταβλητότητα μεταξύ τυχαίων δειγμάτων, δεν εξαλείφεται, αλλά η Στατιστική Συμπερασματολογία μας επιτρέπει να την περιγράψουμε, ανακοινώνοντας ένα περιθώριο σφάλματος με επιθυμητή πιθανότητα.

Έτσι, αν το δείγμα δείξει ότι ένα ποσοστό 45% των συσκευασιών είναι μολυσμένα με *Salmonella* και το περιθώριο σφάλματος με πιθανότητα 95% είναι 2% αυτό σημαίνει ότι: μπορούμε να έχουμε 95% εμπιστοσύνη ότι το ποσοστό μολυσμένων συσκευασιών στην παρτίδα (πληθυσμός) είναι ένας αριθμός 43% και 47% .

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Μέθοδοι Δειγματοληψίας

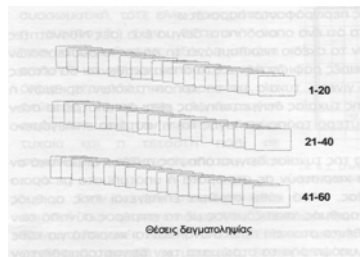
#### 1. Απλή τυχαία δειγματοληψία

Το δείγμα επιλέγεται με τη μέθοδο των τυχαίων αριθμών

Παράδειγμα

Θέλουμε να επιλέξουμε ένα δείγμα 10 συσκευασιών από έναν πληθυσμό 60 συσκευασιών που βρίσκονται σε 3 ράφια ενός ψυγείου σουπερ μάρκετ.

Αριθμούμε τις συσκευασίες



Επιλέγουμε με τη χρήση τυχαίων αριθμών



Τυχαιοί αριθμοί μπορούν να παραχθούν είτε από έτοιμους πίνακες τυχαίων αριθμών είτε με τη χρήση λογισμικών

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

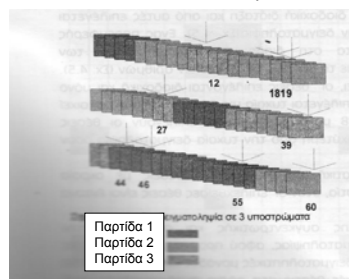
### Μέθοδοι Δειγματοληψίας

#### 2. Στρωματοποιημένη δειγματοληψία

Εφαρμόζεται όταν ο πληθυσμός περιέχει μονάδες με διαφορετικά χαρακτηριστικά (π.χ συσκευασίες από διαφορετικές παρτίδες). Ο πληθυσμός διαιρείται σε τμήματα (στρώματα) με ίδια χαρακτηριστικά. Από κάθε στρώμα επιλέγεται αναλογικός αριθμός δειγματοληπτικών μονάδων και αναλύεται στην αρχή ξεχωριστά και στη συνέχεια συνολικά.

Παράδειγμα

Ένας πληθυσμός 60 συσκευασιών που βρίσκονται σε 3 ράφια ενός ψυγείου σουπερ μάρκετ προέρχονται από 3 διαφορετικές παρτίδες των 20 συσκευασιών. Θέλουμε να επιλέξουμε ένα δείγμα 9 μονάδων.



Χωρίζουμε τον πληθυσμό σε 3 στρώματα για την κάθε παρτίδα και επιλέγουμε 3 μονάδες από το κάθε στρώμα με τη μέθοδο των τυχαίων αριθμών

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Μέθοδοι Δειγματοληψίας

#### 2. Στρωματοποιημένη δειγματοληψία

##### Πλεονεκτήματα.

- Ένα ανομοιογενές περιβάλλον μετατρέπεται σε ομάδες ομοιόμορφων τμημάτων με αποτέλεσμα οι πιθανές διαφορές στους πληθυσμούς των διαφορετικών ομάδων να γίνονται αντιληπτές
- Η διακύμανση ή αλλιώς μεταβλητότητα μεταξύ των δειγματοληπτικών μονάδων ανά στρώμα είναι σαφώς μικρότερη από εκείνη που παρατηρείται συνολικά σε όλες τις μονάδες
- Η συνολική διακύμανση που προέρχεται από την συνένωση των διακυμάνσεων όλων των στρωμάτων είναι μικρότερη από εκείνη μιας τυχαίας δειγματοληψίας

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

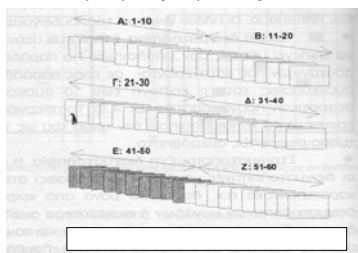
### Μέθοδοι Δειγματοληψίας

#### 3. Συγκεντρωτική ή Συσσωματική ή Ομαδοποιημένη δειγματοληψία

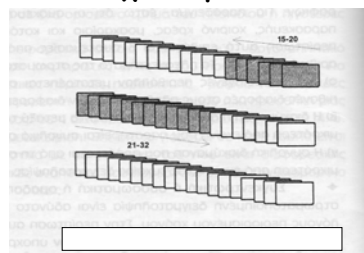
Πιο εύκολη στην πράξη μέθοδος δειγματοληψίας. Ο πληθυσμός χωρίζεται σε ομάδες ίσων μονάδων σε διαδοχική διάταξη και επιλέγεται τυχαία μία ομάδα όπου αναλύονται όλα τα δείγματα. Ένας, άλλος τρόπος εφαρμογής είναι η αρίθμηση όλων των μονάδων, η τυχαία επιλογή μίας μονάδας και η ανάλυση διαδοχικών μονάδων μετά από αυτή

##### Παράδειγμα

Διάρθρωση σε ομάδες και επιλογή μιας ομάδας τυχαία



Αρίθμηση μονάδων, επιλογή μίας μονάδας τυχαία και ανάλυση διαδοχικών μονάδων





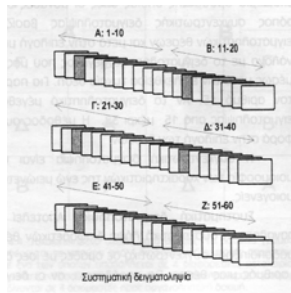
## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Μέθοδοι Δειγματοληψίας

#### 4. Συστηματική δειγματοληψία

Αποτελεί μορφή της συγκεντρωτικής. Ο πληθυσμός διαιρείται σε ομάδες με ίσες διαδοχικές μονάδες και επιλέγεται τυχαία ο αριθμός μίας θέσης

#### Παράδειγμα



Ο πληθυσμός διαιρείται σε 6 ομάδες 10 μονάδων. Επιλέγεται τυχαία μία θέση (θέση 3 στο παράδειγμα)

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Μέθοδοι Δειγματοληψίας

#### 5. Δειγματοληψία σταθερών θέσεων (σταθμών)

Οι μονάδες λαμβάνονται από σταθερές θέσεις η επιλογή των οποίων γίνεται με βάση μία από τις παραπάνω μεθόδους

#### Παράδειγμα

Στους διαδρόμους παραγωγής τυποποιημένων προϊόντων όπου σε σταθερά σημείο και σε καθορισμένο χρονικό διάστημα λαμβάνονται δείγματα.

Σε γραμμές συνεχούς παραγωγής όπου λαμβάνονται μία μονάδα για κάθε X δείγματα που παράγονται

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Κατηγορίες Μεταβλητών και Δεδομένων

Μεταβλητή (variable). Οποιοδήποτε χαρακτηριστικό του οποίου η τιμή αλλάζει στα στοιχεία του πληθυσμού

#### Ανεξάρτητες και Εξαρτημένες μεταβλητές

Τις περισσότερες φορές ο στόχος της έρευνας είναι να δημιουργήσει μία σχέση αιτίου-αποτελέσματος ανάμεσα σε δύο μεταβλητές. Με άλλα λόγια, ο στόχος σε μία έρευνα είναι να αποδειχθεί ότι οι αλλαγές που εμφανίζονται στην τιμή μιας μεταβλητής οφείλονται στην αλλαγή της τιμής μιας άλλης μεταβλητής. Για να γίνει αυτό εφικτό, η επιστημονική έρευνα χρησιμοποιεί δείγματα και προσεκτικά σχεδιασμένα πειράματα, όπου ο ερευνητής μπορεί να ελέγχει την τιμή μιας μεταβλητής και ταυτόχρονα να παρατηρεί τις τιμές της άλλης μεταβλητής, χωρίς η διαδικασία του πειράματος να επηρεάζεται από εξωτερικούς παράγοντες ή μεταβλητές. Η μεταβλητή που ελέγχει ο ερευνητής ονομάζεται ανεξάρτητη μεταβλητή (independent variable), ενώ αυτή που αποτελεί το αντικείμενο παρατήρησης ονομάζεται εξαρτημένη μεταβλητή (dependent variable). Το ζητούμενο για τον ερευνητή είναι να εξετάσει την επίδραση των διαφορετικών τιμών της ανεξάρτητης μεταβλητής στις τιμές της εξαρτημένης.

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Κατηγορίες Μεταβλητών και Δεδομένων

Μεταβλητή (variable). Οποιοδήποτε χαρακτηριστικό του οποίου η τιμή αλλάζει στα στοιχεία του πληθυσμού

#### Ανεξάρτητες και Εξαρτημένες μεταβλητές

#### Παράδειγμα

1. Μελέτη του μικροβιακού πληθυσμού σε ένα τρόφιμο σε σχέση με τη θερμοκρασία και το χρόνο συντήρησης

Ανεξάρτητες μεταβλητές: Θερμοκρασία, Χρόνος

Εξαρτημένη μεταβλητή: Μικροβιακός πληθυσμός

2. Μελέτη της υγρασίας ενός τυριού σε σχέση με συγκέντρωση αλατιού και το χρόνο ωρίμανσης

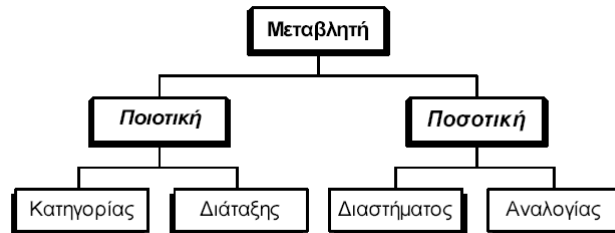
Ανεξάρτητες μεταβλητές: Συγκέντρωση αλατιού, Χρόνος ωρίμανσης

Εξαρτημένη μεταβλητή: Υγρασία

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Κατηγορίες Μεταβλητών και Δεδομένων

**Μεταβλητή (variable).** Οποιοδήποτε χαρακτηριστικό του οποίου η τιμή αλλάζει στα στοιχεία του πληθυσμού



## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Κατηγορίες Μεταβλητών και Δεδομένων

**Ποιοτικές Μεταβλητές.** οι τιμές τους δεν έχουν αριθμητικές ιδιότητες και συνεπώς οι γνωστές πράξεις είναι άνευ νοήματος.

Διακρίνονται με βάση την κλίμακα απόδοσης τιμών σε:

#### **1. Μεταβλητές κατηγορίας (nominal)**

Στις μεταβλητές κατηγορίας, η μόνη σχέση μεταξύ διαφορετικών τιμών, είναι η ύπαρξη διαφοράς. Δηλαδή, μια κλίμακα κατηγορίας, επιτρέπει μόνο την κατηγοριοποίηση των υποκειμένων (π.χ. διαφορετική κατηγορία χρώματος, παραγωγού, προέλευσης).

#### **2. Μεταβλητές διάταξης (ordinal)**

Στις κλίμακες διάταξης, μεταξύ διαφορετικών τιμών υπάρχει σχέση διάταξης. Δηλαδή, μια κλίμακα διάταξης, δεν επιτρέπει μόνο την κατηγοριοποίηση των υποκειμένων αλλά και μια σχέση διάταξης-ιεράρχησής τους. Για παράδειγμα, στη μεταβλητή οργανοληπτικής αξιολόγησης ενός προϊόντος, μπορούν να αποδοθούν τιμές με βάση την κλίμακα: πολύ θετική, θετική, αδιάφορη, αρνητική, πολύ αρνητική. Προφανώς, ίσες διαφορές μεταξύ τιμών δε συνεπάγονται και ίσες διαφορές στο χαρακτηριστικό που εκφράζει η μεταβλητή αφού, οι τιμές αυτές δεν ποσοτικοποιούν το χαρακτηριστικό. Η διάταξη εκφράζει π.χ το «καλύτερο» ή το «προτιμότερο» όχι όμως το «πόσο καλύτερο» ή το «πόσο προτιμότερο». Δηλαδή, 1ος-2ος δεν είναι ίσο με 5ος-6ος.

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Κατηγορίες Μεταβλητών και Δεδομένων

**Ποσοτικές Μεταβλητές.** οι τιμές τους έχουν αριθμητικές ιδιότητες και συνεπώς οι γνωστές πράξεις έχουν νόημα.

Διακρίνονται με βάση την κλίμακα απόδοσης τιμών σε:

#### 1. Μεταβλητές διαστήματος (interval)

Στις μεταβλητές διαστήματος, γίνεται ποσοτικοποίηση του χαρακτηριστικού που εκφράζει η μεταβλητή. Έτσι, μια κλίμακα διαστήματος επιτρέπει, όχι μόνο τη διάταξη-ιεράρχηση των υποκειμένων, αλλά και τον προσδιορισμό επακριβώς της διαφοράς τους. Ίσες διαφορές μεταξύ τιμών συνεπάγονται και ίσες διαφορές στο χαρακτηριστικό που εκφράζει η μεταβλητή. Όμως, δεν έχουν νόημα οι αναλογίες. Αυτό συμβαίνει, διότι στις κλίμακες διαστήματος μετράμε μεταβλητές στις οποίες δεν εμφανίζεται παντελής έλλειψη του χαρακτηριστικού που εκφράζουν. Δηλαδή, το μηδέν δεν εμφανίζεται εγγενώς στο διάστημα τιμών τους (όπως εμφανίζεται για παράδειγμα στις τιμές της μεταβλητής απόσταση δύο σημείων όπου μηδέν σημαίνει ταύτιση των σημείων δηλαδή ότι δεν υπάρχει απόσταση), αλλά ορίζεται συμβατικά-αυθαίρετα.

Για παράδειγμα, θερμοκρασία 0 οC δε συνεπάγονται παντελή έλλειψη του παράγοντα θερμοκρασίας.

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Κατηγορίες Μεταβλητών και Δεδομένων

**Ποσοτικές Μεταβλητές.** οι τιμές τους έχουν αριθμητικές ιδιότητες και συνεπώς οι γνωστές πράξεις έχουν νόημα.

Διακρίνονται με βάση την κλίμακα απόδοσης τιμών σε:

#### 2. Μεταβλητές αναλογίας (ratio)

Μεταβλητές αναλογίας των οποίων οι τιμές αντιστοιχούν αναλογικά στην ποσότητα του φυσικού μεγέθους που μετρούν. Αυτό σημαίνει ότι το μηδέν (0) ανήκει οπωσδήποτε στο διάστημα μεταβολής τους, και δηλώνει την πλήρη απουσία.

Για παράδειγμα, η συγκέντρωση ενός αντιμικροβιακού παράγοντα. Συγκέντρωση 0% δηλώνει παντελή έλλειψη του παράγοντα.

# Μοντέλα στην Επιστήμη Τροφίμων 532Ε

## Περιγραφική Στατιστική

Στόχος της Περιγραφικής Στατιστικής είναι, «η ανάπτυξη μεθόδων για τη συνοπτική και την αποτελεσματική παρουσίαση των δεδομένων».

Τα εργαλεία της Περιγραφικής Στατιστικής είναι

1. Μέθοδοι πινακοποίησης των δεδομένων

2. Μέθοδοι γραφικής παρουσίασης των δεδομένων

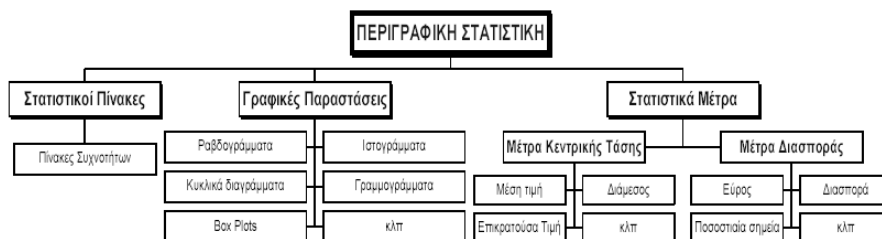
3. Αριθμητικά περιγραφικά μέτρα (Στατιστικά Μέτρα). Είναι αριθμητικά μεγέθη τα οποία βοηθούν στην περιγραφή της κατανομής των δεδομένων. Πολλά από αυτά χρησιμοποιούνται και στη στατιστική συμπερασματολογία. Τα αριθμητικά περιγραφικά μέτρα για τον πληθυσμό ονομάζονται παράμετροι (parameters) ενώ για το δείγμα ονομάζονται στατιστικά (summary statistics).

**ΠΡΟΣΟΧΗ** Τα εργαλεία της Περιγραφικής Στατιστικής που θα χρησιμοποιήσουμε εξαρτώνται από τη φύση των δεδομένων (κατηγορία μεταβλητής)

# Μοντέλα στην Επιστήμη Τροφίμων 532Ε

## Περιγραφική Στατιστική

Τα εργαλεία της Περιγραφικής Στατιστικής



Τα δεδομένα περιγράφονται με πίνακες και σχήματα

Τα δεδομένα περιγράφονται με αριθμούς

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Στατιστικοί Πίνακες

Η κατηγοριοποίηση (ομαδοποίηση, ταξινόμηση) ή γενικότερα η κατανομή των πρωτογενών δεδομένων σε κατηγορίες (ομάδες, τάξεις, διαστήματα) αποτελεί την πρώτη προσέγγιση στο πρόβλημα :

- του περιορισμού του όγκου
- της συστηματικής παρουσίας

κάθε τύπου μεταβλητής (ποιοτικής ή ποσοτικής) από την (Περιγραφική) Στατιστική.

Οι όροι πίνακες συχνοτήτων και κατανομές συχνοτήτων χρησιμοποιούνται ταυτόσημα για να αποδώσουν την τεχνική οργάνωσης των τιμών μιας μεταβλητής σε κατηγορίες αλλά και την τελική παρουσίασή τους.

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Στατιστικοί Πίνακες (Πίνακες Συχνοτήτων)

Η γενική μορφή ενός πίνακα συχνοτήτων

Κατηγορίες (ή τάξεις ή διαστήματα τιμών)	Συχνότητα	Σχετική Συχνότητα (%)	Αθροιστική Συχνότητα	Σχετική Αθροιστ. Συχνότητα

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Στατιστικοί Πίνακες

**Συχνότητα (Frequency):** μίας τιμής  $x_i$  της μεταβλητής  $X$  μεγέθους  $n \times N$  είναι ο φυσικός αριθμός  $v_i \times N$ ,  $v_i \leq n$ , που δείχνει το πλήθος εμφάνισης της τιμής  $x_i$ .  
 Ισχύει προφανώς:  $v_1 + v_2 + \dots + v_m = n$

**Σχετική συχνότητα :** μιας τιμής  $x_i$  της μεταβλητής  $X$  είναι ο αριθμός  $f_i = \frac{v_i}{n}$   
 και  $i = 1, 2, \dots, m$  όπου  $m \leq n$

Ισχύει προφανώς  $f_1 + f_2 + \dots + f_m = 1$

Η σχετική συχνότητα εκφράζεται και επί τοις εκατό, οπότε  $f_i \% = 100 \times f_i$

**Αθροιστική συχνότητα  $N_i$ :** είναι το πλήθος των παρατηρήσεων που είναι μικρότερες ή ίσες της τιμής  $x_i$ .

Ομοίως και σχετική αθροιστική συχνότητα  $F_i$ .

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Στατιστικοί Πίνακες (Πίνακες Συχνοτήτων)

Για Ποιοτικές Μεταβλητές

Π.χ Κατηγορία μεγέθους αυγών: 1(Α), 2(Β), 3(Γ), 4(Δ)

Δείγμα 37 αυγών

3 3 2 2 3 2 2 3 2 3 4 2 4 2 2 4 2 4 2  
 2 4 2 2 3 2 3 2 2 2 4 2 2 3 2 3 2 2

Προκύπτουν άμεσα, από όλες τις δυνατές τιμές που μπορεί να πάρει η μεταβλητή.

Κατηγορίες	Συχνότητα	Σχετική Συχνότητα (%)	<del>Αθροιστική Συχνότητα</del>	<del>Σχετική Αθροιστ. Συχνότητα</del>

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Στατιστικοί Πίνακες (Πίνακες Συχνοτήτων)

Για Ποιοτικές Μεταβλητές

Π.χ Κατηγορία μεγέθους αυγών: 1(Α), 2(Β), 3(Γ), 4(Δ)

Δείγμα 37 αυγών

3 3 2 2 3 2 2 3 2 3 4 2 4 2 2 4 2 4 2  
2 4 2 2 3 2 3 2 2 2 4 2 2 3 2 3 2 2

Κατηγορία	Συχνότητα	Σχετ. Συχνότητα %
2 (Β)	22	59.5
3 (Γ)	9	24.3
4 (Δ)	6	16.2
	37	100

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Στατιστικοί Πίνακες (Πίνακες Συχνοτήτων)

Για Ποσοτικές Μεταβλητές

Π.χ Βάρος συσκευασιών τροφίμων (γρ)

Δείγμα 37 συσκευασιών

129 126 126 115 123 130 149 144 152 135 122 146 138 116 116 117 140 120 132  
129 132 132 113 139 140 122 104 142 132 150 140 143 138 119 114 112 138

Η γενική μορφή ενός πίνακα συχνοτήτων

Κατηγορίες (ή τάξεις ή διαστήματα τιμών)	Συχνότητα	Σχετική Συχνότητα (%)	Αθροιστική Συχνότητα	Σχετική Αθροιστ. Συχνότητα

Προσδιορισμός Κατηγοριών



## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Στατιστικοί Πίνακες (Πίνακες Συχνοτήτων)

Για Ποσοτικές Μεταβλητές

Π.χ Βάρος συσκευασιών τροφίμων (γρ)

Δείγμα 37 συσκευασιών

129 126 126 115 123 130 149 144 152 135 122 146 138 116 116 117 140 120 132  
129 132 132 113 139 140 122 104 142 132 150 140 143 138 119 114 112 138

### Προσδιορισμός Κατηγοριών

Οι κατηγορίες (που τώρα ονομάζονται διαστήματα τιμών) πρέπει να οριστούν.

i) Βρίσκουμε το εύρος των παρατηρήσεων:  $R = X_{\max} - X_{\min}$

ii) Καθορίζουμε το πλήθος των ομάδων (διαστήματα τιμών):  $k = 1 + 3.32 \log_{10}(n)$

iii) Υπολογίζουμε το πλάτος της κάθε ομάδας:  $d = R/k$

εδώ είναι

$R = 152 - 104 = 48$

$k = 6.2 (7)$

$d = 48/7 = 6.8 (7)$  (στρογγυλοποιούμε πάντα προς τα πάνω)

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Στατιστικοί Πίνακες (Πίνακες Συχνοτήτων)

Για Ποσοτικές Μεταβλητές

Π.χ Βάρος συσκευασιών τροφίμων (γρ)

Δείγμα 37 συσκευασιών

129 126 126 115 123 130 149 144 152 135 122 146 138 116 116 117 140 120 132  
129 132 132 113 139 140 122 104 142 132 150 140 143 138 119 114 112 138

Ομάδες	Κέντρο	Κλειστό αριστερά, ανοικτό δεξιά			
		Συχνότητα	Σχετική Συχνότητα	Αθροιστική Συχνότητα	Αθροιστική Σχετ. Συχνότητα
[104 – 111)	107.5				
[111 – 118)	114.5				
[118 – 125)	121.5				
[125 – 132)	128.5				
[132 – 139)	135.5				
[139 – 146)	142.5				
[146 – 153)	149.5				

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Στατιστικοί Πίνακες (Πίνακες Συχνοτήτων)

Για Ποσοτικές Μεταβλητές

Π.χ Βάρος συσκευασιών τροφίμων (γρ)

Δείγμα 37 συσκευασιών

129 126 126 115 123 130 149 144 152 135 122 146 138 116 116 117 140 120 132  
129 132 132 113 139 140 122 104 142 132 150 140 143 138 119 114 112 138

Ομάδες	Κέντρο	Συχνότητα	Σχετική Συχνότητα	Αθροιστική Συχνότητα	Αθροιστική Σχετ. Συχνότητα
[104 – 111)	107.5	1	2.7	1	2.7
[111 – 118)	114.5	7	18.92	8	21.62
[118 – 125)	121.5	5	13.51	13	35.14
[125 – 132)	128.5	5	13.51	18	48.65
[132 – 139)	135.5	8	21.62	26	70.27
[139 – 146)	142.5	7	18.92	33	89.19
[146 – 153)	149.5	4	10.81	37	100.00
		37	100.00		

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Γραφικές Παραστάσεις

- Πριν την κατασκευή τους θα πρέπει να έχουμε στα χέρια μας τον αντίστοιχο πίνακα συχνοτήτων.
- Δεν προσφέρουν περισσότερη πληροφορία από εκείνη που περιέχεται στους αντίστοιχους πίνακες συχνοτήτων.
- Η αμεσότητα της εικόνας διευκολύνει στον εντοπισμό των διαφοροποιήσεων που υπάρχουν.
- Δίνουν μια χοντρική άποψη του προτύπου που (πιθανώς) υπάρχει στον πληθυσμό.

### Ποιοτικών Μεταβλητών

- ραβδόγραμμα
- κυκλικό διάγραμμα

### Ποσοτικών Μεταβλητών

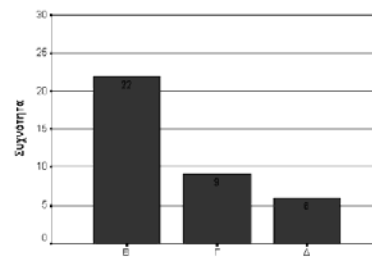
- ιστόγραμμα
  - πολύγωνο συχνοτήτων
  - πολύγωνο αθροιστικών συχνοτήτων
- φυλογράφημα (stem and leaf)
- θηκόγραμμα (box-plot διάγραμμα)

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Γραφικές Παραστάσεις Το Ραβδόγραμμα

- Κατασκευάζουμε τον πίνακα συχνοτήτων της ποιοτικής μεταβλητής
- Θεωρούμε το σύστημα των ορθογώνιων αξόνων.
- Κατά μήκος του οριζώντιου άξονα απεικονίζουμε τις κατηγορίες της μεταβλητής. Μεταξύ των κατηγοριών μεσολαβεί κενό διάστημα
- Κατά μήκος του κατακόρυφου άξονα καταγράφουμε τις (σχετικές) συχνότητες της κάθε κατηγορίας.
- Κατασκευάζουμε ορθογώνια πάνω από κάθε κατηγορία της μεταβλητής με ύψος ίσο με τη συχνότητά της.

Κατηγορία	Συχνότητα	Σχετική Συχνότητα
2	22	59.5
3	9	24.3
4	6	16.2
	37	100

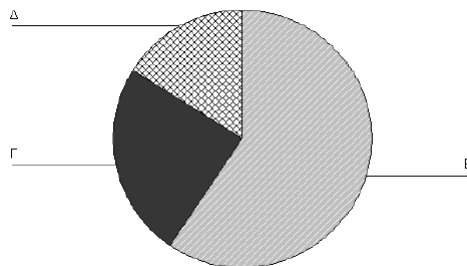


## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Γραφικές Παραστάσεις Το Κυκλικό Διάγραμμα

- Το σύνολο των δεδομένων αντιστοιχεί στο εμβαδό ενός κύκλου
- Κάθε κατηγορία ορίζει έναν κυκλικό τομέα εμβαδού  $\epsilon_1 = 360^\circ \times f_i$

Κατηγορία	Συχνότητα	Σχετική Συχνότητα
2	22	59.5
3	9	24.3
4	6	16.2
	37	100



$$\epsilon_1 = 360^\circ \times f_1 = 360^\circ \times 0 = 0$$

$$\epsilon_2 = 360^\circ \times f_2 = 360^\circ \times 0,595 = 214^\circ$$

.

.

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Γραφικές Παραστάσεις

#### Το Ιστόγραμμα

- Κατασκευάζουμε τον πίνακα συχνοτήτων της ποσοτικής μεταβλητής.
- Θεωρούμε το σύστημα των ορθογωνίων αξόνων.
- Κατά μήκος του οριζόντιου άξονα απεικονίζουμε τις ομάδες της μεταβλητής. Οι ομάδες είναι τοποθετημένες η μία συνεχόμενη της άλλης χωρίς κενά για να φαίνεται η συνέχεια των δεδομένων.
- Κατά μήκος του κατακόρυφου άξονα καταγράφουμε τις (σχετικές) συχνότητες της κάθε ομάδας.
- Κατασκευάζουμε ορθογώνια πάνω από κάθε ομάδα της μεταβλητής με ύψος ίσο με τη (σχετική) συχνότητα.



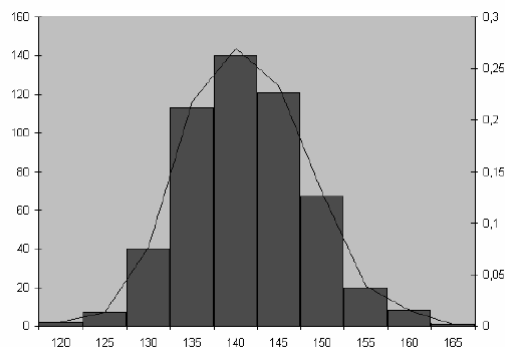
## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Γραφικές Παραστάσεις

#### Το Ιστόγραμμα

(Το σχετικό πολύγωνο συχνοτήτων)

- Κατασκευάζουμε το ιστόγραμμα της ποσοτικής μεταβλητής.
- Ενώνουμε τα μέσα της άνω πλευράς των ορθογωνίων με μια τεθλασμένη γραμμή.



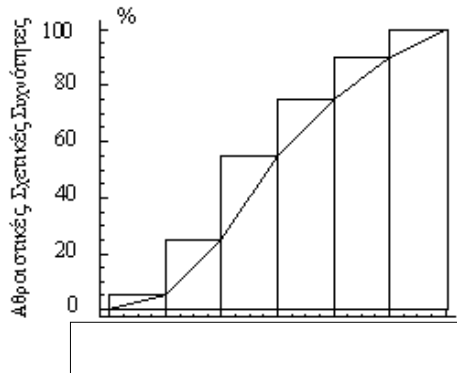
## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

Γραφικές Παραστάσεις

Το Ιστόγραμμα

(Ιστόγραμμα και πολύγωνο αθροιστικών συχνοτήτων)

➤ Το εμβαδόν (και το ύψος) κάθε ορθογωνίου είναι ίσο με την αθροιστική σχετική συχνότητα της αντίστοιχης κλάσης

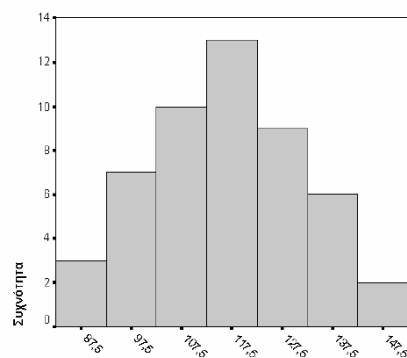
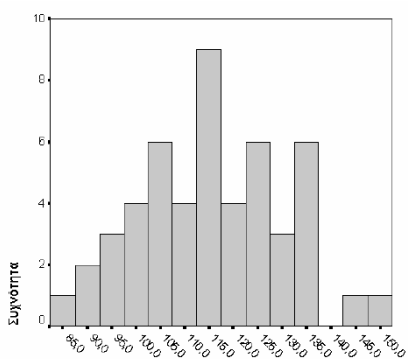


## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

Γραφικές Παραστάσεις

Το Ιστόγραμμα

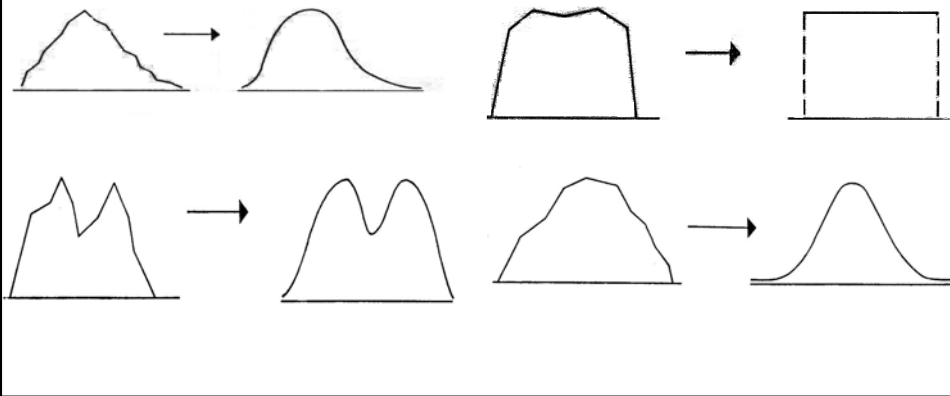
Προβλήματα από την επιλογή του πλήθους των ομάδων.



## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Γραφικές Παραστάσεις

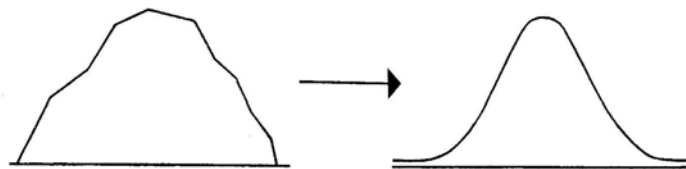
➤ Οι καμπύλες συχνοτήτων, πέραν της προφανούς χρησιμότητάς τους στο πλαίσιο της Περιγραφικής Στατιστικής, έχουν μεγάλη σπουδαιότητα στη Θεωρία Πιθανοτήτων και στη Στατιστική Συμπερασματολογία. Οι καμπύλες συχνοτήτων μπορεί να έχουν διάφορες μορφές όπως:



## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Γραφικές Παραστάσεις

➤ Όταν μια καμπύλη συχνοτήτων είναι συμμετρική ως προς τον κατακόρυφο άξονα που διέρχεται από την κορυφή της κατανομής, όπως η πρώτη από τις παραπάνω, τότε η κατανομή είναι συμμετρική. Τα δύο άκρα της καμπύλης λέγονται ουρές της κατανομής και πλησιάζουν ασυμπτωτικά τον άξονα των τιμών. Προφανώς, σε μια συμμετρική κατανομή, δεξιά και αριστερά του άξονα συμμετρίας βρίσκεται το ίδιο ποσοστό παρατηρήσεων (50%).

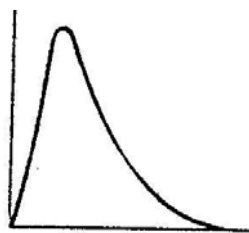


## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

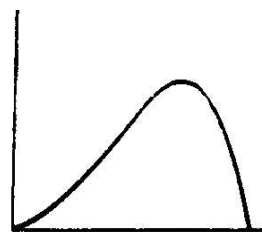
### Γραφικές Παραστάσεις

➤ Όταν η καμπύλη συχνοτήτων δεν είναι συμμετρική, δηλαδή, όταν δεξιά και αριστερά του κατακόρυφου άξονα που περνάει από την κορυφή δε βρίσκεται το ίδιο ποσοστό παρατηρήσεων, τότε η κατανομή είναι ασύμμετρη. Υπάρχουν δύο ειδών ασυμμετρίες: Θετικές και αρνητικές. Μια καμπύλη συχνοτήτων παρουσιάζει θετική ασυμμετρία όταν οι περισσότερες παρατηρήσεις βρίσκονται δεξιά της κορυφής, ενώ, παρουσιάζει αρνητική ασυμμετρία όταν οι περισσότερες παρατηρήσεις βρίσκονται αριστερά της κορυφής.

Θετική ασυμμετρία



Αρνητική ασυμμετρία



## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Γραφικές Παραστάσεις

➤ Τέλος, οι καμπύλες συχνοτήτων, ανάλογα με το βαθμό συγκέντρωσης των παρατηρήσεων στο μέσο και στα άκρα της κατανομής, διακρίνονται σε μεσόκυρτες, λεπτόκυρτες, και πλατύκυρτες

Μεσόκυρτη



Λεπτόκυρτη



Πλατύκυρτη

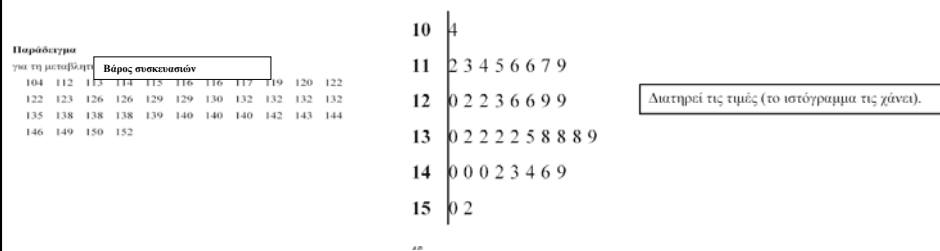


# Μοντέλα στην Επιστήμη Τροφίμων 532Ε

## Γραφικές Παραστάσεις

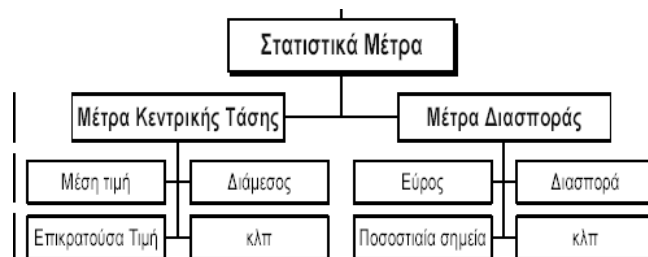
### Το Φυλλόγραμμα

- Διατάσσουμε τις παρατηρήσεις.
- Κάθε παρατήρηση χωρίζεται σε δύο μέρη: (τα οδηγούντα ψηφία -stems- και στα επόμενα leaves-).
- Διατάσσονται τα οδηγούντα ψηφία σε μια στήλη αρχίζοντας από τη μικρότερη τιμή.
- Δίπλα στη γραμμή που αντιστοιχεί στο καθένα από τα οδηγούντα ψηφία γράφονται τα επόμενα ψηφία για κάθε παρατήρηση που έχει stem αυτό της γραμμής.



# Μοντέλα στην Επιστήμη Τροφίμων 532Ε

## Στατιστικά Μέτρα



Με τα μέτρα κεντρικής τάσης προσπαθούμε να προσδιορίσουμε το 'τυπικό', το 'πιο συνηθισμένο' στα δεδομένα.

Τα μέτρα μεταβλητότητας βοηθούν στον εντοπισμό των 'διαφορών' στα δεδομένα.

Η ύπαρξη μόνο των μέτρων κεντρικής τάσης χωρίς αναφορά στα αντίστοιχα μέτρα μεταβλητότητας είναι παραπλανητική και χωρίζαξια.



## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Μέτρα κεντρικής τάσης

Τα μέτρα θέσης-κεντρικής τάσης μας δίνουν πληροφορίες για τη θέση της κατανομής των παρατηρήσεων. Τα πλέον χρησιμοποιούμενα είναι η μέση τιμή, η διάμεσος, η κορυφή και τα ποσοστημόρια.

#### Αριθμητικός μέσος ή μέση τιμή (First moment or Mean or Average)

Η μέση τιμή ενός πληθυσμού συμβολίζεται με  $\mu$  και η μέση τιμή ενός δείγματος με  $\bar{X}$

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$$

εάν οι τιμές  $x_i$  έχουν συχνότητες  $v_i, i=1, 2, \dots, m$  τότε η μέση τιμή δίνεται από τον τύπο:

$$\bar{X} = \frac{v_1 \cdot x_1 + v_2 \cdot x_2 + \dots + v_m \cdot x_m}{n} = \frac{1}{n} \cdot \sum_{i=1}^m v_i \cdot x_i = \sum_{i=1}^m x_i \cdot f_i$$

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Μέτρα κεντρικής τάσης

#### Αριθμητικός μέσος ή μέση τιμή (First moment or Mean or Average)

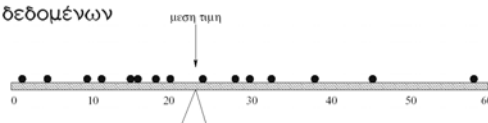
Από τον ορισμό της μέσης τιμής, είναι φανερό ότι αν οι τιμές είναι όλες μεταξύ τους ίσες, θα είναι ίσες με τη μέση τιμή τους. Ξαίνεται, δηλαδή, ότι με τη μέση τιμή επιδιώκεται να ορισθεί ένας «τυπικός εκπρόσωπος» των παρατηρήσεων. Εύκολα αποδεικνύεται ότι εάν στις τιμές της μεταβλητής  $X$  προσθέσουμε έναν αριθμό  $a$  τότε και η μέση τιμή μεταβάλλεται κατά  $a$ .

Δηλαδή  $\bar{X} + a = \overline{X + a}$

Επίσης εάν οι τιμές της  $X$  πολλαπλασιασθούν επί  $k$  τότε και η μέση τιμή πολλαπλασιάζεται επί  $k$ .

Δηλαδή  $k \cdot \bar{X} = \overline{k \cdot X}$

Η μέση τιμή είναι το 'κέντρο ισοροπίας' των δεδομένων



## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Μέτρα κεντρικής τάσης

Αριθμητικός μέσος ή μέση τιμή (First moment or Mean or Average)

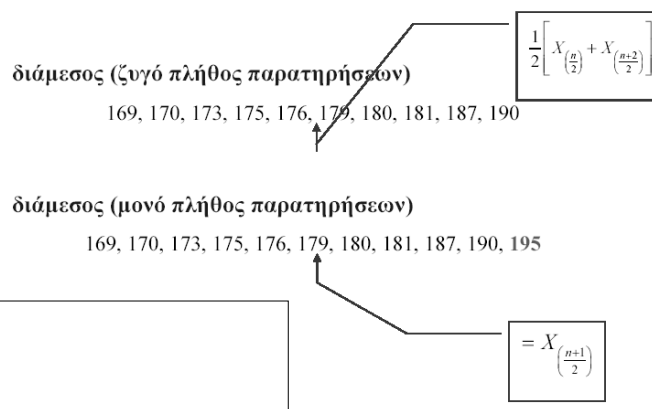
$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$$

Πλεονεκτήματα	Μειονεκτήματα
<ul style="list-style-type: none"> <li>• Για τον υπολογισμό της χρησιμοποιούνται όλες οι τιμές.</li> <li>• Είναι μοναδική για κάθε σύνολο δεδομένων.</li> <li>• Είναι εύκολα κατανοητή.</li> <li>• Ο υπολογισμός της είναι σχετικά εύκολος.</li> <li>• Αξιοποιείται στη στατιστική συμπερασματολογία</li> </ul>	<ul style="list-style-type: none"> <li>• Επηρεάζεται πολύ από ακραίες τιμές.</li> <li>• Μπορεί να μην αντιστοιχεί σε δυνατή τιμή της μεταβλητής.</li> <li>• Δεν υπολογίζεται για ποιοτικά δεδομένα.</li> <li>• Είναι δύσκολος ο υπολογισμός της σε ομαδοποιημένα δεδομένα με ανοικτές τις ακραίες κλάσεις.</li> </ul>

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Μέτρα κεντρικής τάσης

**Διάμεσος δ (Median)** ενός δείγματος παρατηρήσεων, οι οποίες έχουν διαταχθεί κατά αύξουσα τάξη, είναι η μεσαία παρατήρηση εάν το πλήθος των παρατηρήσεων είναι περιττό ή ο μέσος όρος των δύο μεσαίων παρατηρήσεων εάν το πλήθος είναι άρτιο.



## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

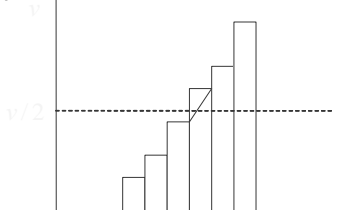
### Μέτρα κεντρικής τάσης

#### Διάμεσος δ (Median)

Στην περίπτωση ομαδοποιημένων μεταβλητών η διάμεσος βρίσκεται από το ιστόγραμμα των αθροιστικών συχνοτήτων. Αλγεβρικά η διάμεσος δίνεται από τον τύπο:

$$\delta = l_i + \frac{\frac{v}{2} - N_{i-1}}{v_i} \cdot c_i$$

όπου  $l_i$  το κατώτερο όριο της κλάσης που περιέχει την διάμεσο,  $v_i$  η συχνότητα και  $c_i$  το πλάτος της κλάσης αντίστοιχα,  $N_{i-1}$  η αθροιστική συχνότητα της προηγούμενης κλάσης και  $v$  το μέγεθος του δείγματος. Δηλαδή η διάμεσος είναι το σημείο τομής του ευθυγράμμου τμήματος, που ενώνει το άνω δεξί άκρο της κλάσης που περιέχει τη διάμεσο με το άνω δεξί άκρο της προηγούμενης κλάσης, και της ευθείας  $y=v/2$



## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Μέτρα κεντρικής τάσης

#### Διάμεσος δ (Median)

Η διάμεσος έχει, μεταξύ άλλων, και την ακόλουθη ιδιότητα:

$$\sum_{i=1}^v |x_i - \delta| < \sum_{i=1}^v |x_i - \lambda|, \quad \forall \lambda$$

Δηλαδή, το άθροισμα των απολύτων αποκλίσεων των παρατηρήσεων από τη διάμεσό τους είναι μικρότερο από το άθροισμα των απολύτων αποκλίσεών τους από οποιαδήποτε άλλη τιμή.

Πλεονεκτήματα	Μειονεκτήματα
<ul style="list-style-type: none"><li>Είναι εύκολα κατανοητή.</li><li>Δεν επηρεάζεται από ακραίες τιμές.</li><li>Υπολογίζεται και στην περίπτωση που οι ακραίες κλάσεις είναι ανοικτές.</li><li>Ο υπολογισμός της είναι απλός.</li><li>Είναι μοναδική σε κάθε σύνολο δεδομένων.</li></ul>	<ul style="list-style-type: none"><li>Δε χρησιμοποιούνται όλες οι τιμές για τον υπολογισμό της.</li><li>Είναι δύσκολη η αξιοποίησή της στη στατιστική συμπερασματολογία.</li><li>Δεν υπολογίζεται για κατηγορικά δεδομένα.</li><li>Για τον υπολογισμό της μπορεί να χρειαστεί παρεμβολή.</li></ul>

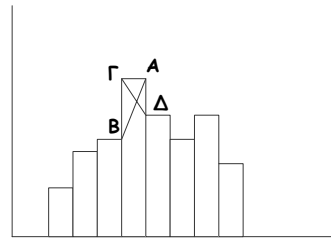
## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Μέτρα κεντρικής τάσης

**Επικρατούσα τιμή  $M_0$  (Mode):** ορίζεται ως η παρατήρηση με την μεγαλύτερη συχνότητα.

Εάν έχουμε ομαδοποιημένη μεταβλητή τότε είναι το σημείο τομής των ευθυγράμμων τμημάτων τα οποία ορίζονται από α) ΑΒ, όπου Α το άνω δεξί άκρο της κλάσης με τη μεγαλύτερη συχνότητα και Β το άνω δεξί άκρο της προηγούμενης κλάσης. β) ΓΔ, όπου Γ το άνω αριστερό άκρο της κλάσης με τη μεγαλύτερη συχνότητα και Δ το άνω αριστερό άκρο της επόμενης κλάσης.

Πλεονεκτήματα	Μειονεκτήματα
<ul style="list-style-type: none"> <li>Υπολογίζεται εύκολα</li> <li>Είναι εύκολα κατανοητή.</li> <li>Υπολογίζεται και από ελλιπή δεδομένα.</li> <li>Δεν επηρεάζεται από ακραίες τιμές.</li> <li>Υπολογίζεται και για ποιοτικά δεδομένα.</li> </ul>	<ul style="list-style-type: none"> <li>Δε χρησιμοποιούνται όλες οι τιμές για τον υπολογισμό της.</li> <li>Στη στατιστική συμπερασματολογία έχει περιορισμένη σημασία.</li> <li>Δεν ορίζεται πάντα μονοσήμαντα. Δηλαδή, μπορεί να υπάρχουν περισσότερες από μία ή και καθόλου.</li> </ul>



## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Μέτρα κεντρικής τάσης

#### Ποσοστιαία σημεία ή Ποσοστημόρια (quantiles)

Τα ποσοστημόρια του δείγματος συμβολίζονται με  $P_\alpha$ . Αποτελούν γενίκευση της διαμέσου και δίνουν αναλυτικότερη περιγραφή της θέσης της κατανομής των παρατηρήσεων. Το ποσοστημόριο  $P_\alpha$  είναι το σημείο της κατανομής για το οποίο το  $\alpha\%$  των παρατηρήσεων είναι μικρότερες ή ίσες από αυτό και το υπόλοιπο  $(1-\alpha)\%$  των παρατηρήσεων είναι μεγαλύτερες ή ίσες από αυτό. Ειδικότερα έχουμε:

**Εκατοστημόρια (percentiles)**  $p_1, p_2, \dots, p_{99}$

**Δεκατημόρια (deciles)** αν  $p_{10}, p_{20}, \dots, p_{90}$

**Τεταρτημόρια (quartiles)**  $p_{25} = Q_1, p_{50} = Q_2 = \delta, p_{75} = Q_3$

α ποσοστημόρια σε ομαδοποιημένες παρατηρήσεις μπορούν να υπολογισθούν από τον τύπο:

$$p_\alpha = L_i + \frac{\frac{\alpha}{100} \cdot v - N_{i-1}}{v_i} \cdot c_i$$

όπου,  $L_i$  είναι το κάτω άκρο της κλάσης στην οποία ανήκει η παρατήρηση με σειρά  $\alpha/100 \cdot v$ ,  $c_i$  είναι το πλάτος της,  $v_i$  είναι η συχνότητά της και  $N_{i-1}$  είναι η αθροιστική συχνότητα της προηγούμενης κλάσης.

# Μοντέλα στην Επιστήμη Τροφίμων 532Ε

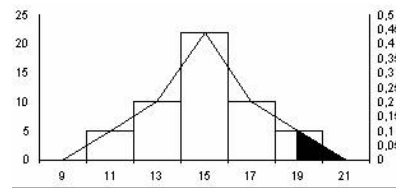
## Μέτρα κεντρικής τάσης

### Ποσοστιαία σημεία ή Ποσοστημόρια (quantiles)

#### Παράδειγμα

Στον παρακάτω πίνακα συχνοτήτων δίνεται η κατανομή της βαθμολογίας 50 μαθητών Λυκείου. Αν στο 5% των μαθητών με την υψηλότερη βαθμολογία δοθεί υποτροφία, τι βαθμό πρέπει να έχει ένας μαθητής για να πάρει υποτροφία;

Βαθμολογία	Συχνότητα	Αθροιστική Συχνότητα
9-10	5	5
11-12	5	10
13-14	10	20
15-16	20	40
17-18	10	50
19-20	5	55
21-22	5	60



$$p_{95} = L_i + \frac{95 \cdot V - N_{i-1}}{V_i} \cdot C_i = 18 + \frac{0.95 \cdot 50 - 45}{5} \cdot 2 \approx 19$$

$L_i$  είναι το κάτω άκρο της κλάσης στην οποία ανήκει η παρατήρηση με σειρά  $a/100 \cdot n$ ,  $C_i$  είναι το πλάτος της,  $V_i$  είναι η *συχνότητά* της και  $N_{i-1}$  είναι η *αθροιστική συχνότητα* της προηγούμενης κλάσης.

# Μοντέλα στην Επιστήμη Τροφίμων 532Ε

## Μέτρα κεντρικής τάσης

Σύγκριση μέσης τιμής, διαμέσου και επικρατούσας τιμής

### QUIZ

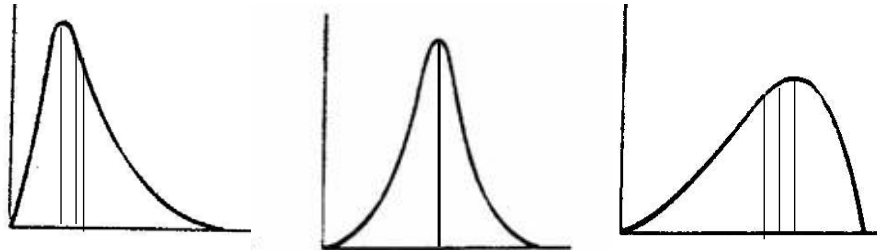
Ποια είναι η διάταξή τους στα παρακάτω ιστογράμματα?



## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Μέτρα κεντρικής τάσης

Σύγκριση μέσης τιμής, διαμέσου και επικρατούσας τιμής  
Ποια είναι η διάταξή τους στα παρακάτω ισογράμματα?



$M_0 \delta \bar{x}$

$\bar{x} = \delta = M_0$

$\bar{x} \delta M_0$

Όταν η καμπύλη συχνοτήτων της κατανομής είναι συμμετρική ισχύει:

$$\bar{x} = \delta = M_0 \quad \delta = \frac{Q_1 + Q_3}{2}$$

Όταν η καμπύλη συχνοτήτων της κατανομής παρουσιάζει θετική ασυμμετρία ισχύει:

$$\bar{x} > \delta > M_0 \quad \delta < \frac{Q_1 + Q_3}{2}$$

Όταν η καμπύλη συχνοτήτων της κατανομής παρουσιάζει αρνητική ασυμμετρία ισχύει: και

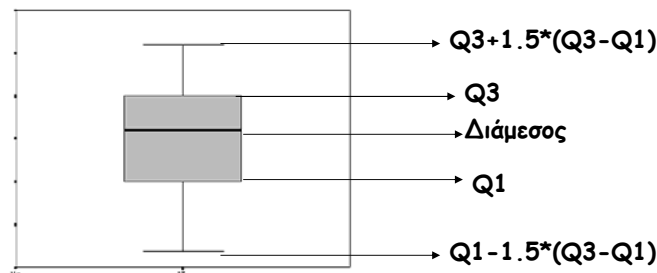
$$\bar{x} < \delta < M_0 \quad \delta > \frac{Q_1 + Q_3}{2}$$

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Γραφικές Παραστάσεις

#### Θηκόγραμμα

Το θηκόγραμμα είναι γνωστό και ως το διάγραμμα των πέντε αριθμών. Πρόκειται για ένα ορθογώνιο με δύο κεραίες (whiskers) το οποίο κατασκευάζεται ως εξής: η κάτω βάση του ορθογωνίου βρίσκεται στο  $Q_1$  και η πάνω στο  $Q_3$ . Η διάμεσος αναπαριστάνεται με ένα οριζόντιο ευθύγραμμο τμήμα μέσα στο ορθογώνιο. Το μήκος των βάσεων του ορθογωνίου λαμβάνεται αυθαίρετα. Η πάνω και η κάτω κεραία που έχουν τη μορφή T και ανεστραμμένου T αντίστοιχα, εκτείνονται μέχρι τις οριακές τιμές που μπορεί να είναι η μεγαλύτερη παρατήρηση που είναι μικρότερη ή ίση από το ανώτερο εσωτερικό φράγμα και η μικρότερη παρατήρηση που είναι μεγαλύτερη ή ίση από το κατώτερο εσωτερικό φράγμα



## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Γραφικές Παραστάσεις Θηκόγραμμα (Box Plot)

Παράδειγμα

Διάμεσος: 132

25% ποσοστιαίο σημείο:  $Q_1 = 120$

75% ποσοστιαίο σημείο:  $Q_3 = 140$

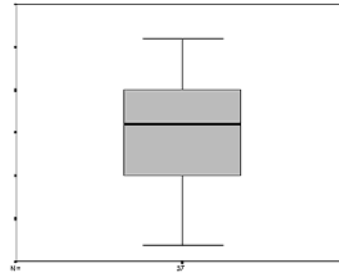
ενδοτεταρτομοριακό εύρος:  $d_F = Q_3 - Q_1 = 20$

κάτω\_φρ: η παρατήρηση με τιμή μεγαλύτερη του  $120 - (1.5)20 = 90$

(104)

πάνω\_φρ: η παρατήρηση με τιμή μικρότερη του  $140 + (1.5)20 = 170$

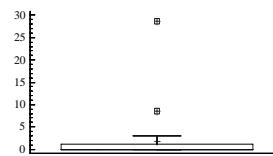
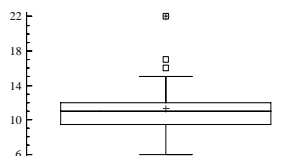
(152)



## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Γραφικές Παραστάσεις Θηκόγραμμα (Box Plot)

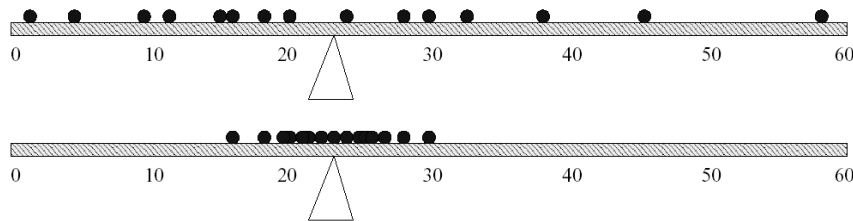
Παράτυπα σημεία (outliers)



## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Μέτρα Διασποράς

Η Μεταβλητότητα ή διασπορά είναι ένα άλλο πολύ σημαντικό χαρακτηριστικό του δείγματος. Τα μέτρα διασπορά περιγράφουν τη μεταβλητότητα του δείγματος και χρησιμοποιούνται για τη εξαγωγή συμπερασμάτων σχετικά με τον πληθυσμό



## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Μέτρα Διασποράς

#### Εύρος (range) και Ενδοτεταρτημοριακή Απόκλιση (interquartile deviation)

Εύρος ορίζεται η διαφορά της μικρότερης από τη μεγαλύτερη παρατήρηση ( ) του δείγματος.

$$R = \chi_{\max} - \chi_{\min}$$

Είναι το πιο απλό μέτρο διασποράς και έχει τα ακόλουθα πλεονεκτήματα και μειονεκτήματα:

Πλεονεκτήματα	Μειονεκτήματα
<ul style="list-style-type: none"><li>• Είναι πολύ απλό στον υπολογισμό.</li><li>• Χρησιμοποιείται αρκετά στον έλεγχο ποιότητας.</li><li>• Μπορεί να χρησιμοποιηθεί για την εκτίμηση της τυπικής απόκλισης.</li></ul>	<ul style="list-style-type: none"><li>• Δε θεωρείται αξιόπιστο μέτρο διασποράς, επειδή βασίζεται μόνο στη μικρότερη και στη μεγαλύτερη παρατήρηση και συνεπώς είναι ευαίσθητο σε έκτροπες τιμές.</li><li>• Δε χρησιμοποιείται για περαιτέρω στατιστική ανάλυση.</li></ul>



## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Μέτρα Διασποράς

Εύρος (range) και Ενδοτεταρτημοριακή Απόκλιση (interquantile deviation)

Δείγμα I	Δείγμα II	Δείγμα III	Δείγμα IV
8	4	1	1
9	7	5	3
10	10	10	10
11	13	15	17
12	16	19	19

Αν χρησιμοποιήσουμε το εύρος για την αριθμητική περιγραφή της μεταβλητότητας στα τέσσερα δείγματα του παραπάνω παραδείγματος, βλέπουμε ότι ενώ ανιχνεύει τη διαφορά στη μεταβλητότητα μεταξύ π.χ. των δειγμάτων I και II (το I έχει εύρος  $12-8 = 4$  ενώ το II έχει εύρος  $16-4 = 12$ ) εντούτοις, δεν ανιχνεύει τη διαφορά που υπάρχει στη μεταβλητότητα μεταξύ των δειγμάτων III και IV (και το III και το IV έχουν εύρος  $19-1 = 18$ ). Δηλαδή, υπάρχουν κατανομές που έχουν ίσες μέσες τιμές, ίσες διαμέσους και ίδιο εύρος και εντούτοις, διαφέρουν σημαντικά. Δεν αρκεί επομένως το εύρος για να αποτυπωθεί αριθμητικά η μεταβλητότητα μιας κατανομής. Είναι φανερό ότι αυτό οφείλεται στο ότι στον υπολογισμό του εμπλέκονται μόνο δυο παρατηρήσεις.

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Μέτρα Διασποράς

Εύρος (range) και Ενδοτεταρτημοριακή Απόκλιση (interquantile deviation)

Δείγμα I	Δείγμα II	Δείγμα III	Δείγμα IV
8	4	1	1
9	7	5	3
10	10	10	10
11	13	15	17
12	16	19	19

Για να αντιμετωπίσουμε αυτό το πρόβλημα μπορούμε να χρησιμοποιήσουμε ως μέτρο της μεταβλητότητας τη διαφορά  $Q3-Q1$  για τον υπολογισμό της οποίας συμμετέχουν σαφώς περισσότερες παρατηρήσεις (όσες συμμετέχουν στον υπολογισμό των  $Q1$  και  $Q3$ ). Η διαφορά αυτή ονομάζεται ενδοτεταρτημοριακή απόκλιση (interquantile deviation). Επειδή μεταξύ των  $Q1$  και  $Q3$  βρίσκεται το 50% των παρατηρήσεων είναι φανερό ότι όσο μικρότερη είναι η ενδοτεταρτημοριακή απόκλιση τόσο μικρότερη είναι η μεταβλητότητα των παρατηρήσεων. Επίσης, σε αντίθεση με την τυπική απόκλιση και τη διασπορά, η ενδοτεταρτημοριακή απόκλιση δεν επηρεάζεται από ακραίες τιμές. Αν χρησιμοποιήσουμε την ενδοτεταρτημοριακή απόκλιση για την αριθμητική περιγραφή της μεταβλητότητας στα τέσσερα δείγματα του παραδείγματός μας, βλέπουμε ότι πλέον ανιχνεύονται όλες οι υπάρχουσες διαφορές μεταξύ των τεσσάρων δειγμάτων.

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Μέτρα Διασποράς

#### Τυπική απόκλιση (standard deviation) και Διασπορά (variance)

Η τυπική απόκλιση του πληθυσμού συμβολίζεται με  $\sigma$  και του δείγματος με  $s$ .

$$s = \sqrt{\frac{1}{v-1} \sum_{i=1}^v (\chi_i - \bar{\chi})^2} = \sqrt{\frac{1}{v-1} \left( \sum_{i=1}^v \chi_i^2 - v \cdot \bar{\chi}^2 \right)}$$

Εύκολα αποδεικνύεται ότι η τυπική απόκλιση έχει τις ακόλουθες ιδιότητες:

• Αν οι παρατηρήσεις είναι μεταξύ τους ίσες τότε η τυπική απόκλιση τους είναι μηδέν

Αν  $\omega_i = \chi_i + \beta$  τότε  $s_{\omega} = s$ . Δηλαδή, αν στις παρατηρήσεις προσθέσουμε μια σταθερή ποσότητα (θετική ή αρνητική), τότε η τυπική τους απόκλιση δεν μεταβάλλεται.

Αν  $\omega_i = \alpha \cdot \chi_i$  τότε  $s_{\omega} = |\alpha| \cdot s$ . Δηλαδή, αν οι παρατηρήσεις πολλαπλασιασθούν με την ίδια ποσότητα, τότε η τυπική τους απόκλιση θα πολλαπλασιασθεί με την το απόλυτο της ίδιας ποσότητας.

Γενικά, αν  $\omega_i = \alpha \cdot \chi_i + \beta$  τότε  $s_{\omega} = |\alpha| \cdot s$ .

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Μέτρα Διασποράς

#### Τυπική απόκλιση (standard deviation) και Διασπορά (variance)

#### QUIZ

Είναι δυνατόν να συγκριθούν οι τυπικές αποκλίσεις δύο δειγμάτων για να βγάλουμε συμπεράσματα για τη μεταβλητότητά τους?

Δύο δείγματα με την ίδια τυπική απόκλιση παρουσιάζουν την ίδια μεταβλητότητα?

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Μέτρα Διασποράς

Τυπική απόκλιση (standard deviation) και Διασπορά (variance)

Δείγμα I	Δείγμα II	Δείγμα III	Δείγμα IV
8	4	1	1
9	7	5	3
10	10	10	10
11	13	15	17
12	16	19	19

Αν για καθένα από τα τέσσερα δείγματα του Πίνακα 2, υπολογίσουμε την τυπική απόκλιση του, παίρνουμε, 1,6 4,7 7,3 και 8,1 αντίστοιχα. Μπορούμε να ισχυρισθούμε ότι η μεταβλητότητα, π.χ. του δείγματος IV είναι μεγαλύτερη από τη μεταβλητότητα του δείγματος I επειδή είναι  $8,1 > 1,6$ ;  
Η απάντηση είναι ναι, γιατί τα δείγματα έχουν την ίδια μέση τιμή.

Αν, όμως, επιχειρήσουμε να συγκρίνουμε τις μεταβλητότητες δύο ή περισσότερων δειγμάτων που έχουν άνισες μέσες τιμές, με βάση μόνο τις τυπικές αποκλίσεις τους, τότε είναι πολύ πιθανό να οδηγηθούμε σε λάθος συμπεράσματα. Αν, για παράδειγμα, σε δύο δείγματα είναι  $\bar{x}=5, s=1$  και  $\bar{x}=150, s=12$  αντίστοιχα, μπορούμε να ισχυρισθούμε ότι το δεύτερο δείγμα παρουσιάζει μεγαλύτερη μεταβλητότητα από το πρώτο επειδή  $12 > 1$ ; Φυσικά όχι, αφού «άλλο 1 στα 5 και άλλο 12 στα 150».

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Μέτρα Διασποράς

Τυπική απόκλιση (standard deviation) και Διασπορά (variance)

Είναι, επομένως, λογικό να αναζητήσουμε ένα μέτρο το οποίο να εκφράζει την τυπική απόκλιση των παρατηρήσεων ως ποσοστό της μέσης τιμής τους. Δηλαδή, ένα μέτρο σχετικής μεταβλητότητας. Ένα τέτοιο μέτρο, είναι ο συντελεστής μεταβλητότητας (coefficient of variation):

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

Έτσι, αν συγκρίνουμε συντελεστές μεταβλητότητας των δύο δειγμάτων,

$$\bar{x}=5, s=1$$

$$\bar{x}=150, s=12$$

παρατηρούμε ότι: Για το πρώτο δείγμα είναι  $CV=20\%$  και για το δεύτερο δείγμα είναι  $CV=8\%$ . Δηλαδή, στο πρώτο δείγμα η τυπική απόκλιση είναι το 20% της μέσης τιμής του ενώ στο δεύτερο δείγμα η τυπική απόκλιση είναι το 8% της μέσης τιμής του. Συνεπώς, η μεγαλύτερη μεταβλητότητα παρουσιάζεται στο πρώτο και όχι στο δεύτερο δείγμα.

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Μέτρα Διασποράς

#### Τυπική απόκλιση (standard deviation) και Διασπορά (variance)

Η τυπική απόκλιση των παρατηρήσεων μπορεί να χρησιμοποιηθεί ως μονάδα μέτρησης της απόστασης των παρατηρήσεων από τη μέση τιμή τους.

Αν μετασχηματίσουμε κάθε τιμή  $x_i$ , σε  $\frac{x_i - \bar{x}}{s}$ , δημιουργούμε μια νέα κατανομή .

$$\frac{x_1 - \bar{x}}{s}, \frac{x_2 - \bar{x}}{s}, \dots, \frac{x_n - \bar{x}}{s}$$

Ας συμβολίσουμε τις νέες τιμές με  $z$ , δηλαδή,  $z_i = \frac{x_i - \bar{x}}{s}$ . Οι  $z_i$ -τιμές έχουν τις ακόλουθες, πολύ ενδιαφέρουσες, ιδιότητες:

1. Η  $z_i$ -τιμή μιας τιμής  $x_i$ , εκφράζει, σε μονάδες τυπικής απόκλισης, την απόσταση της  $x_i$  από τη μέση τιμή .
2. Αν μια  $z_i$ -τιμή είναι θετική αυτό σημαίνει ότι η τιμή  $x_i$  είναι μεγαλύτερη από τη μέση τιμή ενώ αν είναι αρνητική σημαίνει ότι η τιμή  $x_i$  είναι μικρότερη από τη μέση τιμή.
3. Η μέση τιμή των  $z_i$ -τιμών είναι πάντα 0 και η τυπική τους απόκλιση είναι πάντα 1.

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Μέτρα Διασποράς

#### Τυπική απόκλιση (standard deviation) και Διασπορά (variance)

Οι  $z_i$ -τιμές έχουν τις ακόλουθες, πολύ ενδιαφέρουσες, ιδιότητες:

4. Ίσες αποστάσεις  $z_i$ -τιμών μιας κατανομής, έχουν ταυτόσημο νόημα. Για παράδειγμα, η διαφορά μεταξύ των  $z$ -τιμών 2 και 2,5 είναι ταυτόσημη με τη διαφορά μεταξύ των  $z$ -τιμών 3 και 3,5. Και οι δύο διαφορές δείχνουν μια απόσταση μισής τυπικής απόκλισης.
5. Στις  $z$ -τιμές το 0 έχει νόημα, δηλαδή, δεν ορίζεται συμβατικά-αυθαίρετα. Η  $z$ -τιμή 0 σημαίνει «έλλειψη απόστασης», δηλαδή, η τιμή συμπίπτει με τη μέση τιμή. Η μορφή της κατανομής των  $z$ -τιμών είναι όμοια με τη μορφή της κατανομής των τιμών (διατηρούνται π.χ. οι ασυμμετρίες ή η συμμετρία). Έτσι, αν η κατανομή των τιμών έχει μορφή κανονικής κατανομής, τότε και η κατανομή των  $z_i$ -τιμών θα έχει μορφή κανονικής κατανομής.
6. Οι  $z$ -τιμές μπορούν να χρησιμοποιηθούν για τη σύγκριση τιμών που ανήκουν σε διαφορετικές κατανομές.

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Μέτρα Διασποράς

#### Τυπική απόκλιση (standard deviation) και Διασπορά (variance)

Ο κανόνας του Chebyshev.

Για κάθε αριθμό  $k > 1$ , τουλάχιστον το  $100\left(1 - \frac{1}{k^2}\right)\%$  των παρατηρήσεων βρίσκεται στο διάστημα  $(\bar{x} - kS, \bar{x} + kS)$  (ποσοστό των τιμών που βρίσκονται σ' ένα διάστημα με κέντρο τη μέση τιμή).

Ο εμπειρικός κανόνας.

Χρειαζόμαστε κατανομή στην οποία η διάμεσος, η μέση τιμή και η επικρατούσα τιμή να συμπίπτουν (περίπου).

$(\bar{x} - S, \bar{x} + S)$  68%

$(\bar{x} - 2S, \bar{x} + 2S)$  95%

$(\bar{x} - 3S, \bar{x} + 3S)$  99%

## Μοντέλα στην Επιστήμη Τροφίμων 532Ε

### Μέτρα Διασποράς

#### Τυπική απόκλιση (standard deviation) και Διασπορά (variance)

Το τετράγωνο της τυπικής απόκλισης των παρατηρήσεων ονομάζεται διασπορά και συμβολίζεται με  $\sigma^2$  για τον πληθυσμό και με  $s^2$  για το δείγμα. Δηλαδή η διασπορά δίνεται από τον τύπο:

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

Η διασπορά, ως μέτρο μεταβλητότητας, δε διαφέρει ουσιαστικά από την τυπική απόκλιση. Παρότι, έχει το μειονέκτημα ότι δεν εκφράζεται στις ίδιες μονάδες με την μεταβλητή της οποίας τη μεταβλητότητα μετράει, εντούτοις, χρησιμοποιείται ευρύτατα στη Στατιστική Συμπερασματολογία για τις καλές της μαθηματικές ιδιότητες. Αυτός είναι και ο βασικός λόγος που επέβαλε τον ορισμό της.

# Μοντέλα στην Επιστήμη Τροφίμων 532Ε

## Μέτρα Διασποράς

### Τυπική απόκλιση (standard deviation) και Διασπορά (variance)

. Συνοπτικά, η διασπορά και η τυπική απόκλιση έχουν τα ακόλουθα πλεονεκτήματα και μειονεκτήματα:

Πλεονεκτήματα	Μειονεκτήματα
<ul style="list-style-type: none"><li>• Για τον υπολογισμό τους, λαμβάνονται υπόψη όλες οι παρατηρήσεις.</li><li>• Έχουν μεγάλη εφαρμογή στη στατιστική συμπερασματολογία</li><li>• Με βάση την τυπική απόκλιση και τη μέση τιμή, μπορούν να ορισθούν διαστήματα στα οποία βρίσκεται γνωστό ποσοστό παρατηρήσεων</li></ul>	<ul style="list-style-type: none"><li>• Το κυριότερο μειονέκτημα της διασποράς είναι ότι δεν εκφράζεται στις ίδιες μονάδες με τη μεταβλητή. Το μειονέκτημα αυτό παύει να υπάρχει με τη χρησιμοποίηση της τυπικής απόκλισης</li><li>• Απαιτούνται περισσότερες αλγεβρικές πράξεις για τον υπολογισμό τους απ' ό,τι στα άλλα μέτρα.</li></ul>

# Μοντέλα στην Επιστήμη Τροφίμων 532Ε

## Μέτρα Διασποράς

### Μέτρα Ασυμμετρίας

συντελεστής λοξότητας  $\alpha_3 = \frac{\sum(X_i - \bar{X})^3}{nS^3}$  τιμές στο διάστημα (-3, 3)

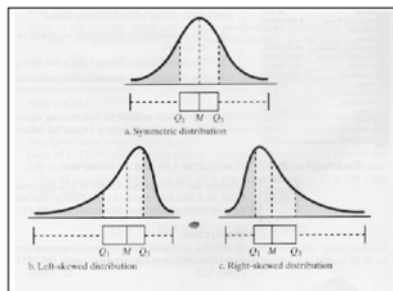
⇒ συμμετρική κατανομή ( $\alpha_3 = 0$ )

⇒ ασύμμετρη αριστερά (αρνητικός συντελεστής λοξότητας)

οι περισσότερες παρατηρήσεις αριστερά της επικρατούσας τιμής

⇒ ασύμμετρη δεξιά (θετικός συντελεστής λοξότητας)

οι περισσότερες παρατηρήσεις δεξιά της επικρατούσας τιμής



# Μοντέλα στην Επιστήμη Τροφίμων 532Ε

## Μέτρα Διασποράς

### Μέτρα Ασυμμετρίας

$$\text{συντελεστής κέρτωσης } \alpha_4 = \frac{\sum (X_i - \bar{X})^4}{nS^4}$$

⊖ < 3 (πλατύκυρτη)

⊖ > 3 (λεπτόκυρτη)

