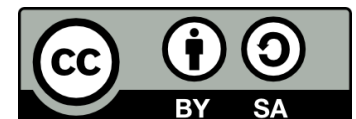




Στατιστική

2^ο Μάθημα: Εισαγωγή στη Στατιστική

Γεώργιος Μενεξές
Τμήμα Γεωπονίας



Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.
- Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδειας χρήσης, η άδεια χρήσης αναφέρεται ρητώς.



Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «Ανοικτά Ακαδημαϊκά Μαθήματα στο Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.





2^ο Μάθημα

Εισαγωγή στη Στατιστική

Η παρουσίαση βασίζεται σε υλικό από το έργο κυρίως των παρακάτω συγγραφέων:

- Παπαδημητρίου, Γ. (2001). *Περιγραφική Στατιστική*. Θεσσαλονίκη: Παρατηρητής.
- Κατσίλης, Ι. (1997). *Περιγραφική Στατιστική*. Αθήνα: Gutenberg.



Η Στατιστική του Κόσμου

Αν περιορίζαμε ολόκληρη την ανθρωπότητα σε ένα χωριό εκατό κατοίκων, αλλά διατηρούσαμε τις αναλογίες όλων των λαών τότε αυτό το χωριό θα αποτελείτο από:

- 57 Ασιάτες
- 21 Ευρωπαίους
- 14 Αμερικανούς
- 8 Αφρικανούς
- 52 γυναίκες
- 48 άνδρες
- 70 μη άσπρους



Η Στατιστική του Κόσμου (συνέχεια)

- 30 άσπρους
- 70 μη Χριστιανούς
- 30 Χριστιανούς
- 89 ετεροφυλόφιλους
- 11 ομοφυλόφιλους
- 6 πρόσωπα θα κατείχαν το 59% του παγκόσμιου πλούτου. Και τα έξι θα προέρχονταν από την Αμερική
- 80 θα είχαν μη ικανοποιητικές συνθήκες κατοικίας
- 70 θα ήταν αναλφάβητοι
- 50 θα ήταν υποσιτισμένοι



Η Στατιστική του Κόσμου (συνέχεια)

- 1 θα πέθαινε
- 2 θα γεννιόνταν
- 1 θα είχε κομπιούτερ
- 1 (μόνο ένας) θα είχε πτυχίο Πανεπιστημίου



Η Στατιστική του Κόσμου (συνέχεια)

- Αν σήμερα το πρωί ξυπνήσατε υγιείς και όχι άρρωστοι τότε είστε πιο τυχεροί από 1 εκατομμύριο ανθρώπους οι οποίοι δεν πρόκειται να ζουν την επόμενη βδομάδα.
- Αν δεν ζήσατε ποτέ ένα πόλεμο, ούτε νοιώσατε την μοναξιά της αιχμαλωσίας, την αγωνία του τραυματισμού και την πείνα τότε είστε τυχερότεροι από 500 εκατομμύρια ανθρώπους του κόσμου.
- Αν μπορείτε να πηγαίνετε στην εκκλησία χωρίς τον φόβο ότι θα σας απειλήσουν, θα σας συλλάβουν ή θα σας σκοτώσουν, τότε είστε πιο τυχεροί από 3 δισεκατομμύρια ανθρώπους του κόσμου.



Η Στατιστική του Κόσμου (συνέχεια)

- Αν έχετε φαγητό μέσα στο ψυγείο σας, αν είστε ντυμένοι, αν έχετε μια στέγη πάνω από το κεφάλι σας και ένα κρεβάτι, τότε είστε πλουσιότεροι από το 75% των κατοίκων αυτού του κόσμου.
- Αν έχετε λογαριασμό στην τράπεζα, λίγα λεφτά στο πορτοφόλι και λίγα ψιλά σε ένα κουμπαρά, τότε ανήκετε στο 8% των εύπορων ανθρώπων αυτού του κόσμου.



ΞΕΚΙΝΩΝΤΑΣ...

1. *“Know your topic”*
2. *“Know your people”*

Berdie, Anderson and Niebuhr (1986)



**Read, Study, Search
and Read again**



Παγκόσμια Έρευνα

Έγινε μια παγκόσμια έρευνα, η οποία περιλάμβανε την παρακάτω ερώτηση:

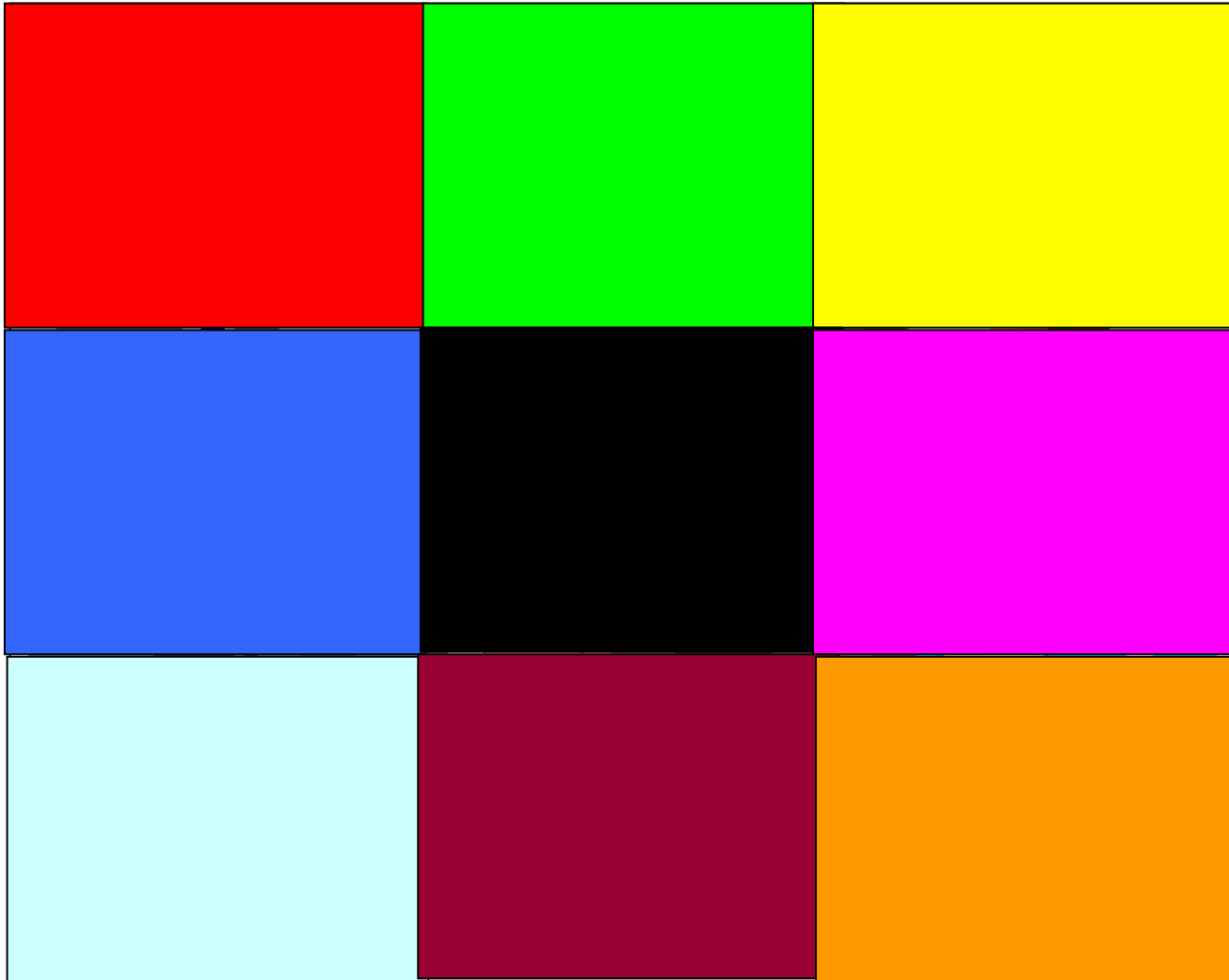
"Παρακαλούμε πείτε μας τη γνώμη σας για την έλλειψη φαγητού στον υπόλοιπο κόσμο."

Οι ερευνητές δεν μπόρεσαν να καταλήξουν σε κανένα συμπέρασμα, αφού αντιμετώπιστηκαν τα παρακάτω προβλήματα, κατά τη διάρκεια της έρευνας:

- Στην Αφρική κανείς δεν γνώριζε τι είναι το Φαγητό,
- Στη Δυτική Ευρώπη κανείς δεν γνώριζε τι είναι η Έλλειψη,
- Στην Ανατολική Ευρώπη κανείς δεν γνώριζε τι είναι η Γνώμη,
- Στην Νότιο Αμερική κανείς δεν γνώριζε τι είναι το Παρακαλούμε
- Στις ΗΠΑ κανείς δεν γνώριζε τι είναι ο υπόλοιπος κόσμος.



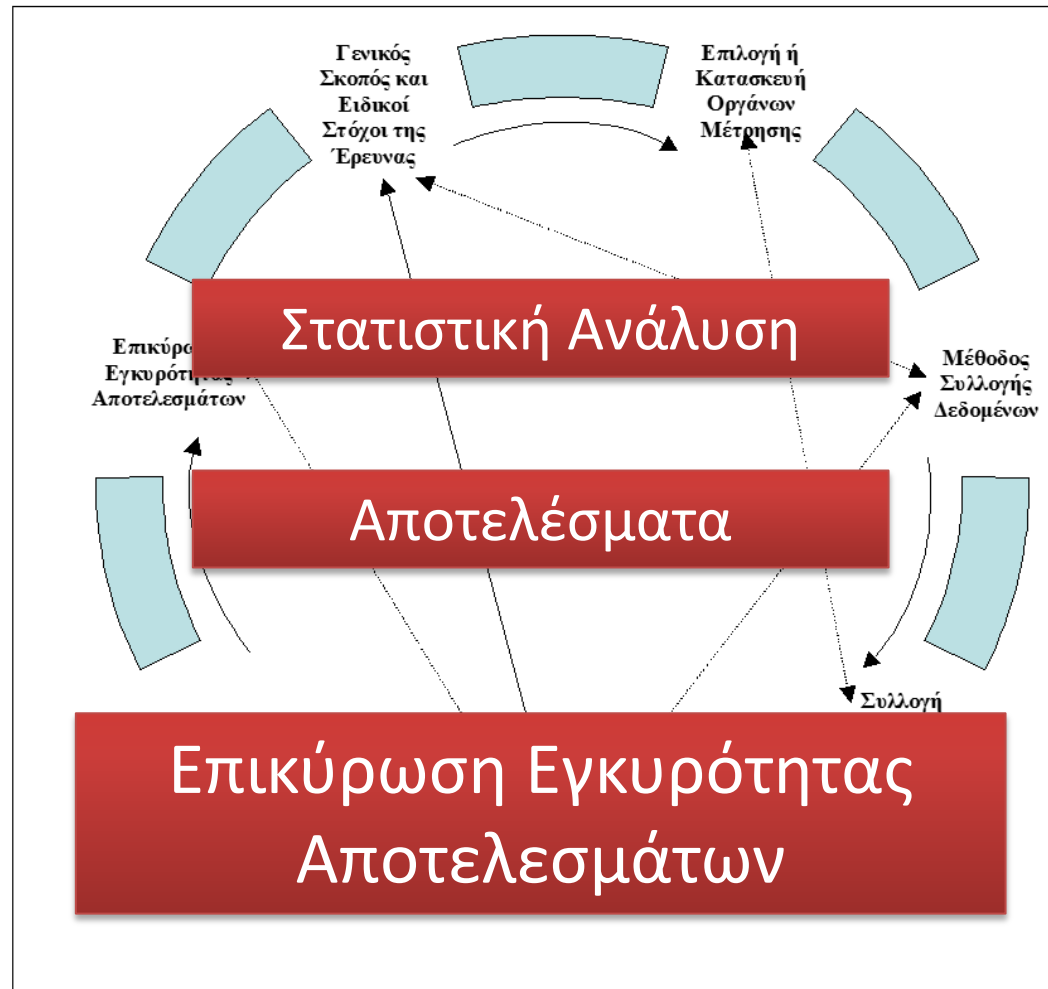
Η Επιστημονική Έρευνα ... και η Αναζήτηση της Αλήθειας



Δομημένη Προσέγγιση στη Μεθοδολογία Έρευνας



Δομημένη Προσέγγιση στη Μεθοδολογία Έρευνας



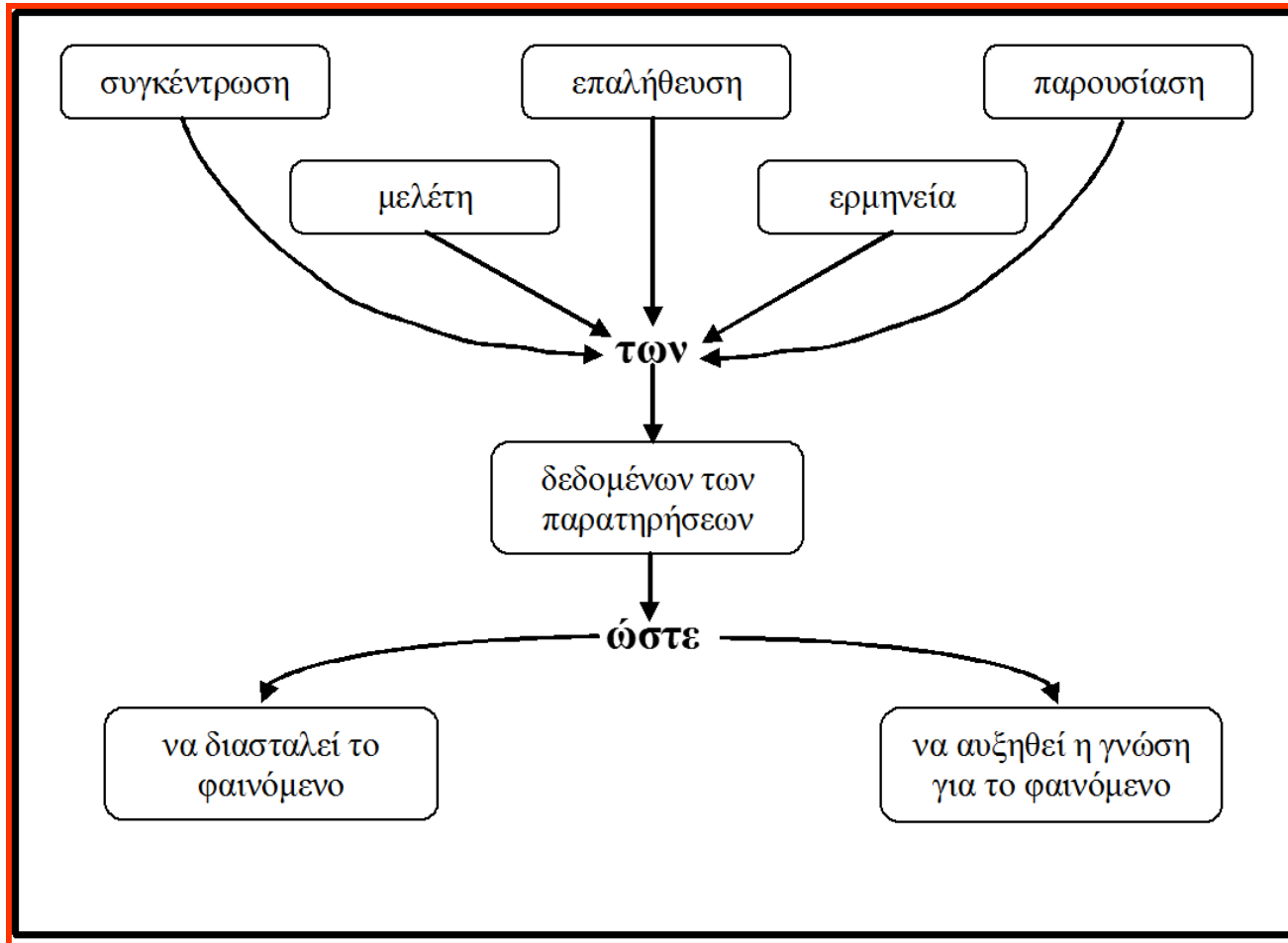
Βασικές Έννοιες και Ορισμοί (1)

Στατιστική:

Η επιστήμη που έχει ως αντικείμενο την αποτελεσματική αξιοποίηση δεδομένων μετά από κατάλληλη κατά περίπτωση **συλλογή, επεξεργασία, οργάνωση, παρουσίαση και ανάλυση** αριθμητικών στοιχείων με σκοπό τη σωστή λήψη μέτρων και αποφάσεων.



Διαδικασίες της Στατιστικής



Πηγή: Παπαδημητρίου (2001).

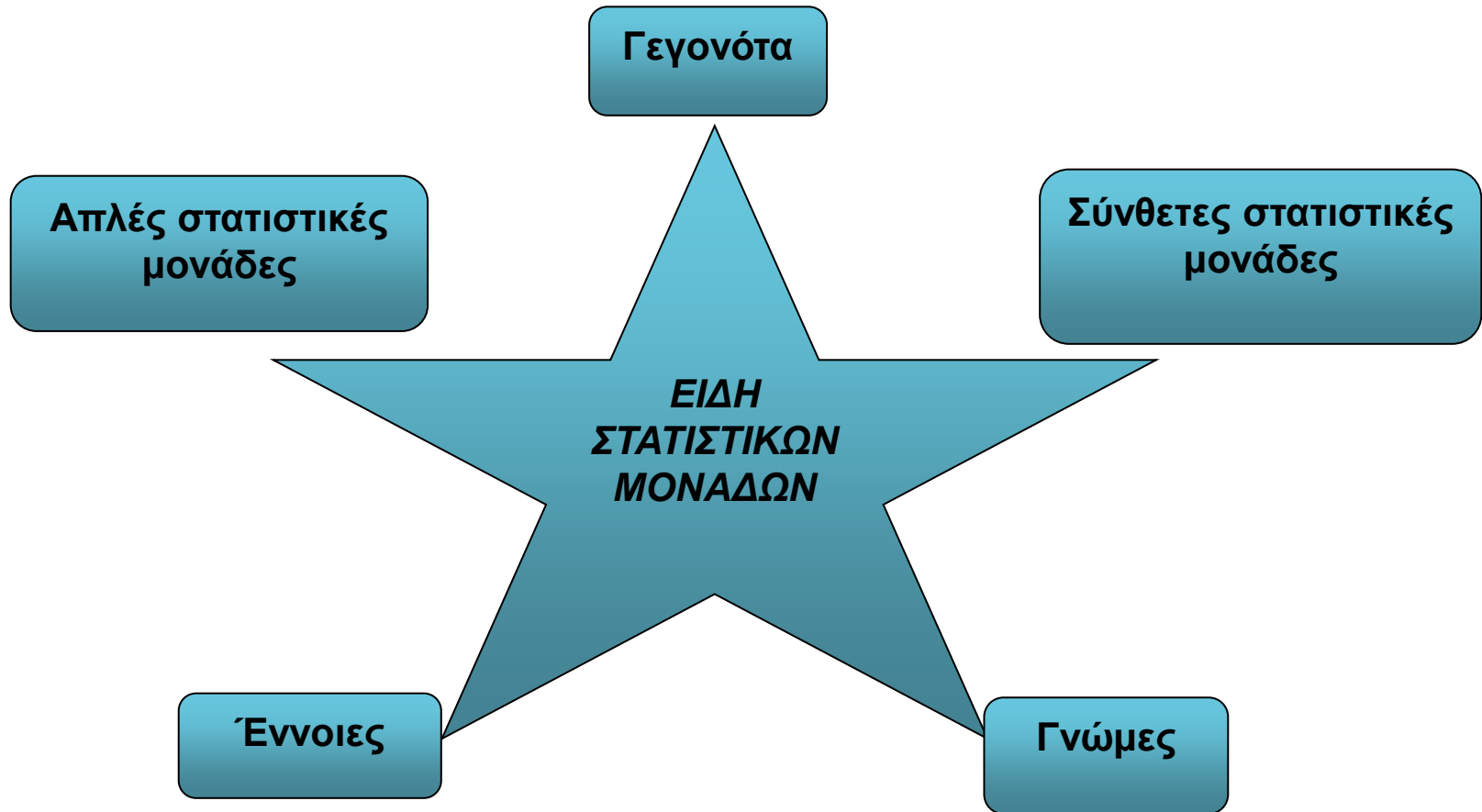


Βασικές Έννοιες και Ορισμοί (2)

- **Πληθυσμός:** Τα στοιχεία (μετρήσεις-παρατηρήσεις) για όλους τους ανθρώπους (υποκείμενα) ή τα πράγματα (αντικείμενα) μιας μελέτης. Το σύνολο των στοιχείων-δεδομένων που μας ενδιαφέρει.
- **Στατιστικός Πληθυσμός:** Το σύνολο των παρατηρήσεων που αναφέρονται σε ένα χαρακτηριστικό ή ιδιότητα των μονάδων του υπό εξέταση πληθυσμού.
- Κάθε στοιχείο του πληθυσμού ονομάζεται **μονάδα** ή **άτομο**.



Στατιστικές Μονάδες



Πηγή: Παπαδημητρίου (2001).



Συγκέντρωση Δεδομένων



Πηγή: Παπαδημητρίου (2001).



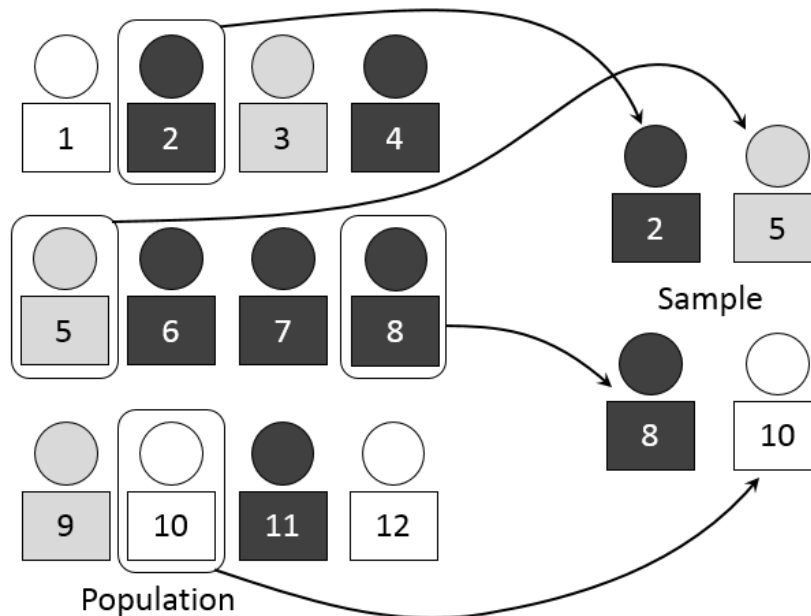
Βασικές Έννοιες και Ορισμοί (3)

- **Δείγμα:** Ένα μέρος (μια υπο-ομάδα) του πληθυσμού. Ένα υποσύνολο δεδομένων από τον πληθυσμό.
- **Τυχαία δειγματοληψία:** Η τεχνική της επιλογής μιας υπο-ομάδας του πληθυσμού έτσι ώστε κάθε δυνατή υπο-ομάδα του ίδιου μεγέθους να έχει την ίδια πιθανότητα επιλογής.
- Στην **απλή τυχαία δειγματοληψία** όλα τα δείγματα μεγέθους n έχουν την ίδια πιθανότητα να επιλεγούν.



Απλή Τυχαία Δειγματοληψία

- Όταν κάθε στατιστική μονάδα του πληθυσμού που μελετούμε έχει την ίδια πιθανότητα να επιλεγεί σαν μονάδα του δείγματος, τότε έχουμε **απλή τυχαία δειγματοληψία**.



Πηγή: en.wikipedia.org/wiki



Συστηματική Τυχαία Δειγματοληψία



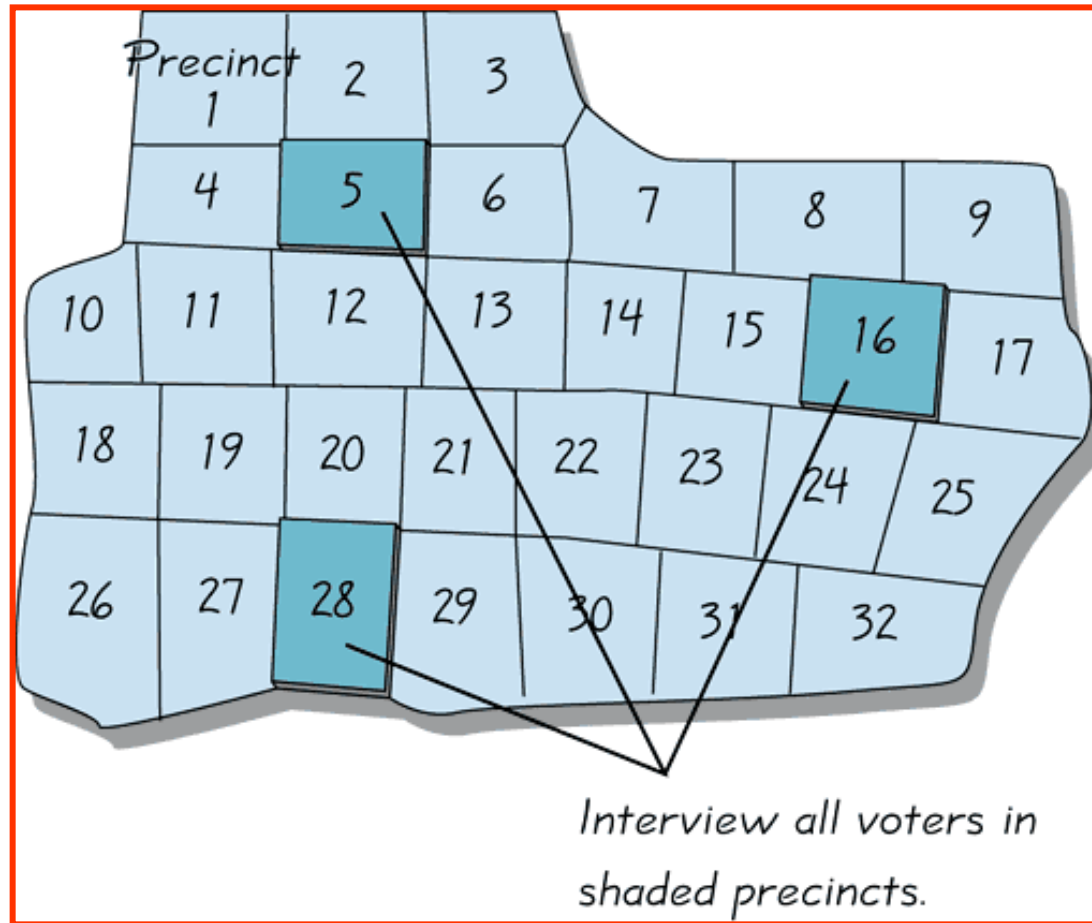
Στρωματοποιημένη Τυχαία Δειγματοληψία



Πηγή: complex.elte.hu/.../ES1.pdf



Κατά Συστάδες Τυχαία Δειγματοληψία



Πηγή: <https://quizlet.com/2784476/htl-exam-i-flash-cards/>



Συμπτωματική ή “Βολική-Ευκολίας” Δειγματοληψία



Πηγή: [Zamir \(...\)](#)



Βασικές Έννοιες και Ορισμοί (4)

- **Σταθερή:** Ένα χαρακτηριστικό που έχει την ίδια τιμή για όλες τις μονάδες του πληθυσμού.
- **Μεταβλητή:** Ένα χαρακτηριστικό που παίρνει διαφορετικές τιμές στον πληθυσμό.
- **Τυχαία μεταβλητή:** Κάθε χαρακτηριστικό ή ιδιότητα των στατιστικών μονάδων.
- **Στατιστικός Πληθυσμός:** Το σύνολο των ιδεατών μετρήσεων ή παρατηρήσεων που αναφέρονται σε κάποια μεταβλητή του πληθυσμού.



Βασικές Έννοιες και Ορισμοί (5)

- **Παράμετρος:** Τιμή-δείκτης που υπολογίζεται από δεδομένα του πληθυσμού.
- **Στατιστικό:** Τιμή-δείκτης που υπολογίζεται από δεδομένα του δείγματος.

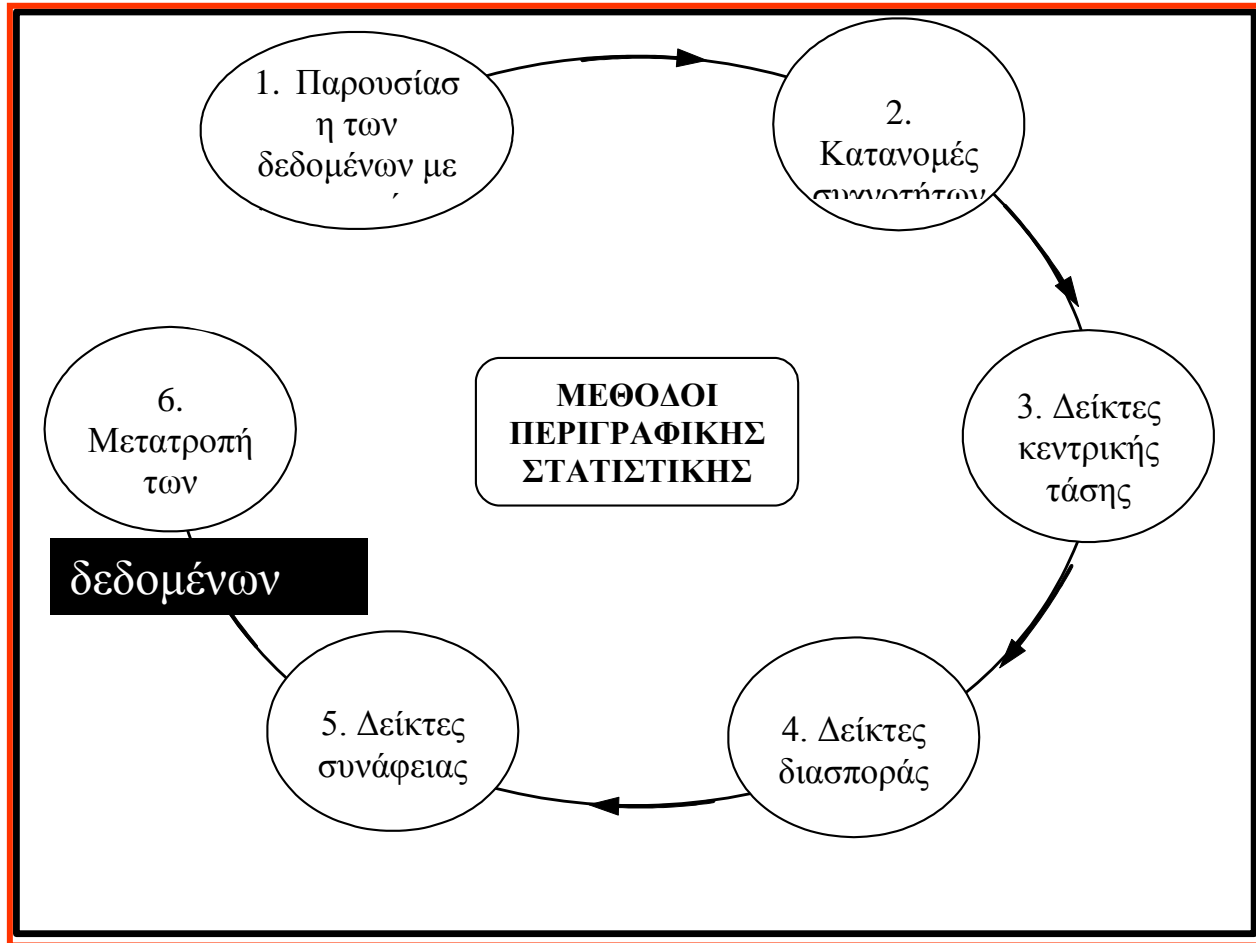


Βασικές Έννοιες και Ορισμοί (6)

- **Περιγραφική Στατιστική:** Στατιστικά που περιγράφουν την κατανομή μιας ή περισσότερων μεταβλητών.
 - Προσφέρει μεθόδους εξαγωγής συμπερασμάτων που αναφέρονται αποκλειστικά στον ερευνηθέντα πληθυσμό χωρίς γενικεύσεις ή προβλέψεις.
- **Επαγωγική Στατιστική:** Στατιστικά και έλεγχοι που χρησιμοποιούνται για γενικεύσεις από το δείγμα στον πληθυσμό.



Περιγραφική Στατιστική



Πηγή: Παπαδημητρίου (2001)



Βασικές Έννοιες και Ορισμοί (7)

- **Μέτρηση:** Η πρόσαψη αριθμών ή ονομασιών σε αντικείμενα ή γεγονότα σύμφωνα με κάποια προκαθορισμένα κριτήρια.
- Τα **όργανα μέτρησης** είναι τα "κριτήρια" που χρησιμοποιούμε στη μέτρηση.



Βασικές Έννοιες και Ορισμοί (8)

- **Εγκυρότητα:** Το να μετρά ένα όργανο μέτρησης αυτό που υποτίθεται ότι μετρά.
- **Αξιοπιστία:** Το να δίνει ένα όργανο μέτρησης τα ίδια αποτελέσματα, όταν μετρά το ίδιο πράγμα σε διαφορετικές περιπτώσεις.



Βασικές Έννοιες και Ορισμοί (9)

- Κλίμακες μέτρησης:
 - Ονομαστική
 - Διάταξης
 - Ίσων Διαστημάτων
 - Αναλογική
 - Ειδική περίπτωση οι **δίτιμες** κλίμακες που παίρνουν μόνο δύο τιμές (0, 1).



Τυπολογία Κλιμάκων Μέτρησης (I)






Τυπολογία Κλιμάκων Μέτρησης (II)

- **Ονομαστικές** (*nominal*): Το σύνολο τιμών τους δηλώνει μόνο διαφοροποίηση (π.χ. χρώμα ματιών, τόπος γέννησης, φύλο).
- **Διάταξης** (*ordinal*): Στο σύνολο τιμών τους μπορούμε να ορίσουμε σχέση διάταξης (π.χ. σειρά κατάταξης σε ένα αγώνισμα, μορφωτικό επίπεδο, κλάσεις ηλικιών, κλάσεις εισοδήματος, βαθμίδες ιεραρχίας).
- **Διαστήματος** (*interval*): Ίσες διαφορές μεταξύ των τιμών τους συνεπάγονται και ίσες διαφορές για το αντίστοιχο χαρακτηριστικό που μετρά η κλίμακα (π.χ. θερμοκρασία, ηλικία). Το μηδέν δεν είναι καλά ορισμένο.
- **Αναλογίας** (*ratio*): Οι τιμές τους αντιστοιχούν αναλογικά στην ποσότητα του χαρακτηριστικού που μετρούν (π.χ. ταχύτητα, χρόνος, μήκος, βάρος, εισόδημα). Το μηδέν είναι καλά ορισμένο.



Τυπολογία Κλιμάκων Μέτρησης (III)

Παράδειγμα

Ονομαστική	Χώρα προέλευσης οδηγού			
		Αγγλία	Γαλλία	Βραζιλία
Διάταξης	Κατάταξη στο αγώνισμα	3ος	2ος	1ος
Διαστήματος	Βαθμολογία απόδοσης στο παγκόσμιο πρωτάθλημα (κλίμακα 0-100)	80	90	98
Αναλογίας	Επίδοση-Χρόνος Τερματισμού σε hours	1,4	1,3	1,2

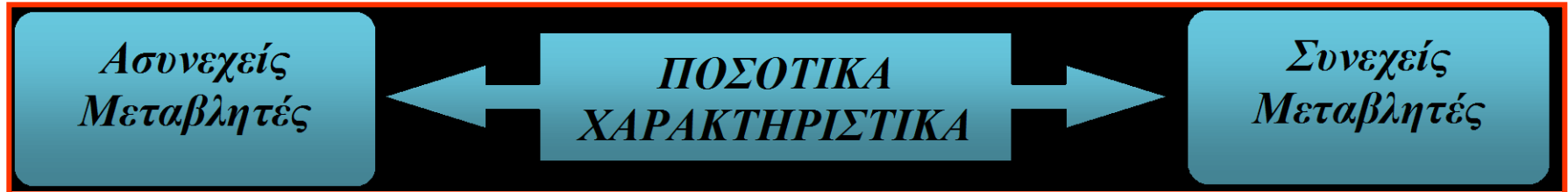


Μεταβλητές

- Τα χαρακτηριστικά των στατιστικών μονάδων, που θέλουμε να μελετήσουμε σε μία συγκεκριμένη περίπτωση, ονομάζονται **μεταβλητές**.
- Τα χαρακτηριστικά στα οποία μπορούμε να αντιστοιχίσουμε έναν αριθμό μιας ορισμένης κλίμακας ονομάζονται ποσοτικά ή **ποσοτικές** μεταβλητές.
- **Ποιοτικές** μεταβλητές είναι εκείνες των οποίων τις τιμές δεν μπορούμε να τις μετρήσουμε.



Βασικές Έννοιες και Ορισμοί (10)



- **Συνεχής μεταβλητή:** Μια μεταβλητή που μπορεί να πάρει απεριόριστο αριθμό τιμών. Θεωρητικά μπορεί να πάρει οποιαδήποτε τιμή ενός διαστήματος του συνόλου R .
- **Ασυνεχής ή απαριθμητή μεταβλητή:** Μια μεταβλητή που μπορεί να πάρει μόνο ορισμένο αριθμό τιμών. Δέχεται μόνο τιμές που ανήκουν στο σύνολο Z .

Πηγή εικόνας: Παπαδημητρίου (2001)



Ποιοτικές Μεταβλητές

- Τέτοιες μεταβλητές είναι το **φύλο**, το **επίπεδο μόρφωσης**, ο **τόπος καταγωγής**, η **γνώμη** ενός ατόμου για ένα προϊόν.
- Παρά το ότι οι ποιοτικές μεταβλητές δεν μετρούνται, είμαστε υποχρεωμένοι να δώσουμε σ' αυτές ορισμένες τιμές (μία σε κάθε στατιστική μονάδα). Έτσι για τη μεταβλητή φύλο δίνουμε την τιμή 1 όταν πρόκειται για άνδρα και την τιμή 2 για γυναίκα. Πρέπει να τονίσουμε ότι οι τιμές 1 και 2 δεν έχουν την έννοια της διάταξης (το 2 είναι μεγαλύτερο του 1) αλλά της **διαφορετικότητας**.
- Οι μέθοδοι **Βέλτιστης Κλιμάκωσης** (*Optimal Scaling*) οδηγούν σε μια τελείως διαφορετική οπτική και αντιμετώπιση των κλιμάκων μέτρησης των μεταβλητών.



Βασικές Έννοιες και Ορισμοί (11)

- Οι **κατανομές συχνοτήτων** μπορούν να χρησιμοποιηθούν για την περιγραφή μεταβλητών ανεξάρτητα από την κλίμακα μέτρησης.
- Οι κατανομές συχνοτήτων παρουσιάζονται είτε υπό μορφή **πινάκων ή/και με διαγράμματα**.



Διάρθρωση Στατιστικών Πινάκων



Παράδειγμα Διάρθρωσης Πίνακα

Παράδειγμα διάρθρωσης στατιστικού πίνακα

Επικεφαλίδα - Τίτλος ← Πίνακας 7.3: Παραγωγή Ηλεκτρικού Ρεύματος στην Ελλάδα

Μονάδες και έτος αναφοράς } σε εκατομμύρια ΩΧΒ έτη 1959 - 1964

Κύριο Σώμα {

Έτος	Παραγωγή*
1959	1968
1960	2198
1961	2484
1962	2735
1963	3113
1964	3716
Σύνολο	13214

Κορμός (από την παραγωγή του 1961)

Στήλη (από την παραγωγή του 1961)

Πηγή - Υποσημειώσεις {

Πηγή: ΕΣΥΕ

* Υποσημείωση: Δεν περιλαμβάνεται η παραγωγή του εργοστασίου για δική τους χρήση



Πίνακας Απόλυτων Συχνοτήτων

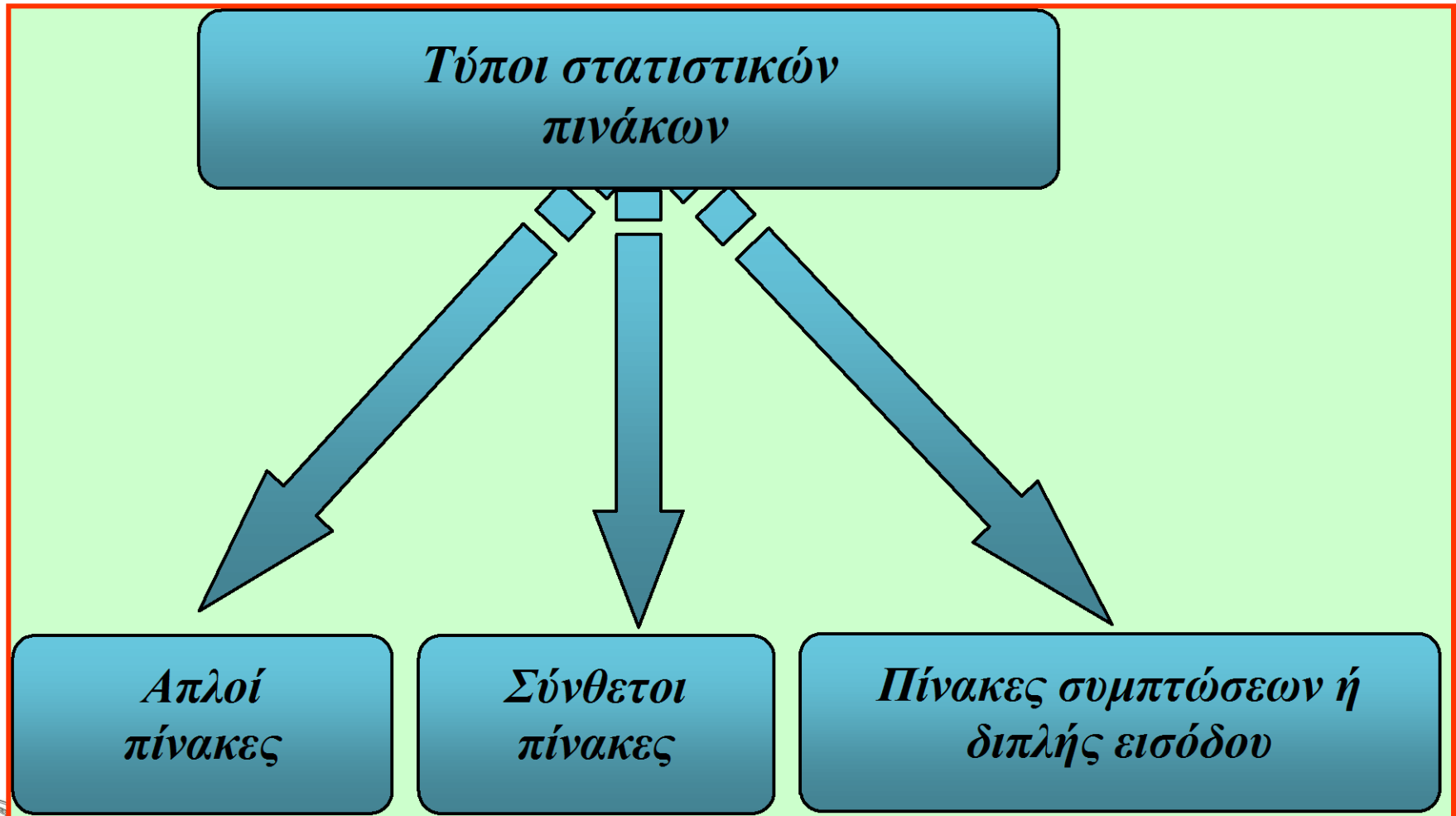
Πίνακας 4.1: Κατανομή των πρωτοετών φοιτητών των ετών 1980,85,90 και 1995 ως προς το επίπεδο μόρφωσης της μητέρας

Επίπεδο Μόρφωσης μητέρας	Πρωτοετείς 1980	Πρωτοετείς 1985	Πρωτοετείς 1990	Πρωτοετείς 1995	→ Τίτλοι
Α' βάρθια	2.300	2.400	2.200	2.000	→ Στήλες
Β' βάρθια	2.800	3.800	3.500	3.200	
Γ' βάρθια	1.400	1.800	2.000	2.200	
Σύνολο	6.500	8.000	7.700	7.400	→ Σύνολο

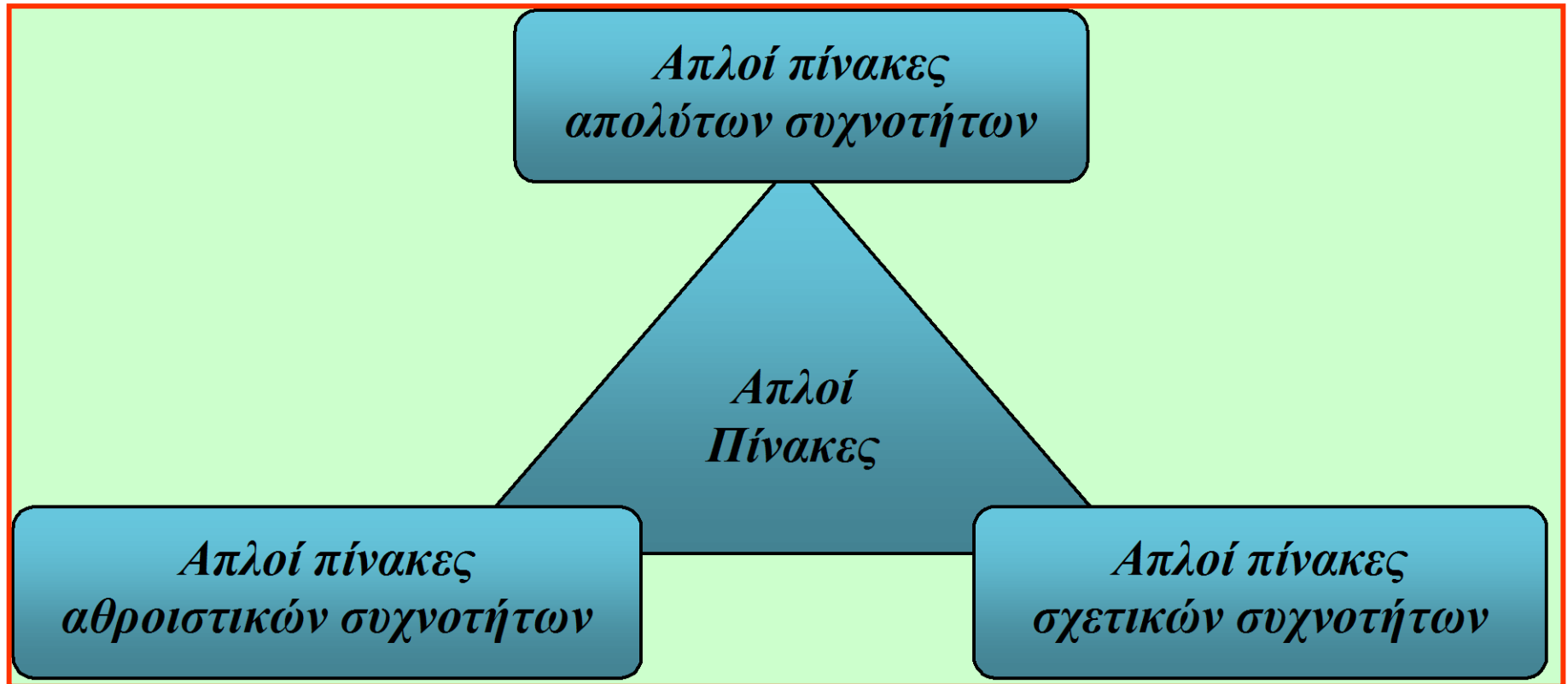
Πηγή: Παπαδημητρίου (2001)



Τύποι Στατιστικών Πινάκων



Απλοί Πίνακες Συχνοτήτων



Πηγή: Παπαδημητρίου (2001)



Παράδειγμα (Π1)

Πίνακας 4.4 : Κατάσταση των φορτηγών εμπορικών πλοίων με ελληνική σημαία το έτος 1935 ανάλογα με την ηλικία τους.

Ηλικία πλοίου	Αριθμός πλοίων
Μέχρι 5 ετών	0
5 – 15	8
15 – 20	149
20 – 25	121
25 – 40	170
40 - 50	34
Πάνω από 50	24
Σύνολο	506

Πηγή: Στατ. Επετηρίδα της Ελλάδας 1936, σελ. 261

Πηγή: Παπαδημητρίου (2001)



Παράδειγμα (Π2)

Παραδείγματα απλών πινάκων σχετικών συχνοτήτων

Πίνακας 4.7: Κατανομή 1200 εργαζομένων ως προς την Οικογενειακή Κατάστασή τους

Οικ. Κατ.	Απόλ. Συχν. n_i		Σχετικ. Συχν. P_i
Άγαμος	300	(300/1200=)	0,25
Έγγαμος	700	(700/1200)	0,58
Διαζευγ μένος	50	(50/1200)	0,04
Χήρος	150	(150/1200)	0,13
Σύνολο	1200		1,00

Πίνακας 4.8: Κατανομή 1200 εργαζομένων ως προς την ηλικία τους

ηλικία	Απόλ. Συχν. n_i		Σχετικ. Συχν. P_i
[18 – 25)	200	(200/1200)	0,17
[25 – 35)	500	(500/1200)	0,42
[35 – 50)	400	(400/1200)	0,33
[50 – 65]	100	(100/1200)	0,08
Σύνολο	1200		1,00



Παράδειγμα (Π3)

Πίνακας 4.9: Απόλυτη αθροιστική συχνότητα της μεταβλητής X

X	Απόλυτη Συχνότητα n_i		Αθροιστική Συχνότητα φ_i	
Κλάσεις {	C_1	n_1	$(n_1 = \varphi_1) \longrightarrow$	$\varphi_1 = n_1$
	C_2	n_2	$(n_1+n_2 = \varphi_2) \longrightarrow$	φ_2
	C_3	n_3	$(n_1+n_2+n_3 = \varphi_3) \longrightarrow$	φ_3
	C_4	n_4	$(n_1+n_2+n_3+n_4 = \varphi_4) \longrightarrow$	$\varphi_4 = N$
Σύνολο	N			

Πηγή: Παπαδημητρίου (2001)



Παράδειγμα (Π4)

Πίνακας 4.10: Κατανομή 1200 εργαζομένων ως προς το επίπεδο μόρφωσης

Επίπεδο μόρφωσης	n_i	Φ_i	P_i	F_i
Αγράμματος	50	50	0,042	0,042
Δημοτικό	100	150	0,083	0,125
Γυμνάσιο	300	450	0,250	0,375
Λύκειο	500	950	0,417	0,792
ΑΕΙ ή ΤΕΙ	250	1200	0,208	1,00
Σύνολο	1200		1,00	

n_i = απόλυτη συχνότητα (ποιοτικών μεταβλητών)

Φ_i = αθροιστική συχνότητα $\Phi_i = \sum_{j=1}^i n_j = n_1 + n_2 + \dots + n_i$

P_i = σχετική συχνότητα $P_i = n_i / N$

F_i = αθροιστική σχετική συχνότητα $F_i = \sum_{j=1}^i P_j = p_1 + p_2 + \dots + p_i = \frac{\Phi_i}{N}$



Παράδειγμα (Π5)

Πίνακας 4.12: Έγγαμοι κατά φύλο και ηλικία

Ηλικία	Άνδρες	Γυναίκες	Σύνολο
10-19	4260	36908	41168
20-29	161852	332504	494356
30-44	776608	856316	1632924
45-64	819176	732548	1551724
65-79	291100	190800	481900
80 και άνω	42632	18820	61452
Σύνολο	2095628	2167896	4263524

Πηγή: Στατιστική Επετηρίδα Ε.Σ.Υ.Ε. 1980

Πηγή: Παπαδημητρίου (2001)



Παράδειγμα (Π6)

Παράδειγμα: Κατανομή 400 ατόμων ως προς ηλικία και επίπεδο μόρφωσης

Πίνακας 4.13: Απόλυτες Συχνότητες

Ηλικία	Επίπεδο μόρφωσης			Σύνολο
	A' βάθμια	B' βάθμια	Γ' βάθμια	
15 – 25	150	250	100	500
25 – 45	100	300	400	800
45 – 65	150	450	100	700
Σύνολο	400	1000	600	2000

Πίνακας 4.14: Σχετικές Συχνότητες (κατά στήλες)

Ηλικία	Επίπεδο μόρφωσης		
	A' βάθμια	B' βάθμια	Γ' βάθμια
15 – 25	0,375	0,250	0,167
25 – 45	0,250	0,300	0,667
45 – 65	0,375	0,450	0,166
Σύνολο	1,000	1,000	1,000

Πίνακας 4.15: Σχετικές Συχνότητες (κατά γραμμές)

Ηλικία	Επίπεδο μόρφωσης			Σύνολο
	A' βάθμια	B' βάθμια	Γ' βάθμια	
15 – 25	0,300	0,500	0,200	1,000
25 – 45	0,125	0,375	0,500	1,000
45 – 65	0,214	0,643	0,143	1,000



Παράδειγμα (Π7)

Πίνακας 4.16: Πίνακας συμπτώσεων των μεταβλητών ηλικία και επίπεδο μόρφωσης

Ηλικία	Επίπεδο μόρφωσης			Σύνολο
	A' βάθμια	B' βάθμια	Γ' βάθμια	
18 – 25	10	110	30	150
25 – 35	20	320	320	660
35 – 50	50	350	280	680
50 – 65	120	220	170	510
Σύνολο	200	1000	800	2000

Πίνακας 4.17: Πίνακας συμπτώσεων σχετικών συχνοτήτων με ερμηνεία κατά στήλες

Ηλικία	Επίπεδο μόρφωσης			Σύνολο
	A' βάθμια	B' βάθμια	Γ' βάθμια	
18 – 25	0,050	0,110	0,038	0,075
25 – 35	0,100	0,320	0,400	0,330
35 – 50	0,250	0,350	0,350	0,340
50 – 65	0,600	0,220	0,212	0,255
Σύνολο	1,000	1,000	1,000	1,000

Πίνακας 4.18: Πίνακας συμπτώσεων σχετικών συχνοτήτων με ερμηνεία κατά γραμμές

Ηλικία	Επίπεδο μόρφωσης			Σύνολο
	A' βάθμια	B' βάθμια	Γ' βάθμια	
18 – 25	0,067	0,733	0,200	1,000
25 – 35	0,030	0,485	0,485	1,000
35 – 50	0,074	0,515	0,411	1,000
50 – 65	0,235	0,432	0,333	1,000
Σύνολο	0,100	0,500	0,400	1,000



Γενικευμένοι Πίνακες Συμπτώσεων (*Burt*)

	ΗΛΙΚΙΑ X				ΑΠΟΔΟΧΕΣ 15/θημέρου Y			
	(15-25]	(25-35]	(35-45]	(45-55]	(100-140]	(140-200]	(200-260]	(260-300]
(15-25]	50				40	5	5	0
(25-35]		150			30	50	60	10
(35-45]			200*		20	75	75	30
(45-55]				100	10	20	10	60
(100-140]	40	30	20	10	100			
(140-200]	5	50	75	20		150		
(200-260]	5	60	75	10			150	
(260-300]	0	10	30	60				100



Γενικευμένοι Πίνακες Συμπτώσεων (*Burt*)

	ΗΛΙΚΙΑ X				ΑΠΟΔΟΧΕΣ 15/θημέρου Y			
	(15-25]	(25-35]	(35-45]	(45-55]	(100-140]	(140-200]	(200-260]	(260-300]
(15-25]	50→0,10				40→0,40	5→0,03	5→0,03	0→0,0
(25-35]		150→0,30			30→0,30	50→0,33	60→0,40	10→0,10
(35-45]			200→0,40		20→0,20	75→0,50	75→0,50	30→0,30
(45-55]				100→0,20	10→0,10	20→0,13	10→0,07	60→0,60
(100-140]	40→0,80	30→0,20	20→0,10	10→0,10	100→0,20			
(140-200]	5→0,10	50→0,33	75→0,375	20→0,20		150→0,30		
(200-260]	5→0,10	60→0,40	75→0,375	10→0,10			150→0,30	
(260-300]	0→0,0	10→0,07	30→0,15	60→0,60				100→0,20



Άσκηση

Σε πρόσφατη έρευνα ρωτήθηκαν κάποια άτομα σχετικά με το ποιο θεωρούν ως το σημαντικότερο πρόβλημα σήμερα. Οι απαντήσεις στην ερώτηση αυτή διασταυρώθηκαν με τις διάφορες ηλικιακές κατηγορίες των ατόμων που συμμετείχαν στην έρευνα. Μέρος των αποτελεσμάτων δίνεται στον παρακάτω πίνακα:

Πηγή: Μπεχράκης (1999)



Πίνακας Συμπτώσεων "Ηλικιακής Κατηγορίας" με "Σημαντικότερο Πρόβλημα"

	Τζόγος	Στεγαστικό	Σπουδές	Οικονομικό	Υγεία	Σύνολα
18-24	24	4	23	82	12	145
25-34	42	2	20	146	13	223
35-44	?	0	12	136	18	220
45-54	40	1	6	?	9	177
55-64	37	1	7	68	15	?
65+	20	5	3	61	13	?
Σύνολα	217	?	71	614	80	

Πηγή: Μπεχράκης (1999)



Ζητούμενα

- Να συμπληρωθούν τα στοιχεία που λείπουν (δηλώνονται με ?).
- Πόσα άτομα συμμετείχαν στο δείγμα;
- Ποιο είναι το ποσοστό % των ατόμων ηλικίας 35-44 ετών που θεωρούν ως σημαντικότερο πρόβλημα το τζόγο;
- Από τα άτομα που θεωρούν ως σημαντικότερο πρόβλημα το Οικονομικό πόσο τοις εκατό ανήκει στην ηλικιακή κατηγορία 45-54 ετών;
- Ποιο είναι το ποσοστό % των ατόμων ηλικίας 45-64 ετών που θεωρούν ως σημαντικότερο πρόβλημα την υγεία;
- Στο σύνολο του δείγματος, ποιο είναι το ποσοστό των ατόμων που ανήκουν στην ηλικιακή κατηγορία 18-24 και ταυτόχρονα θεωρούν τις σπουδές ως το σημαντικότερο πρόβλημα;



Ιστογράμματα (1)

Πίνακας 5.14 : Κατανομή κατά ηλικία των ανέργων Ελλήνων του 1996

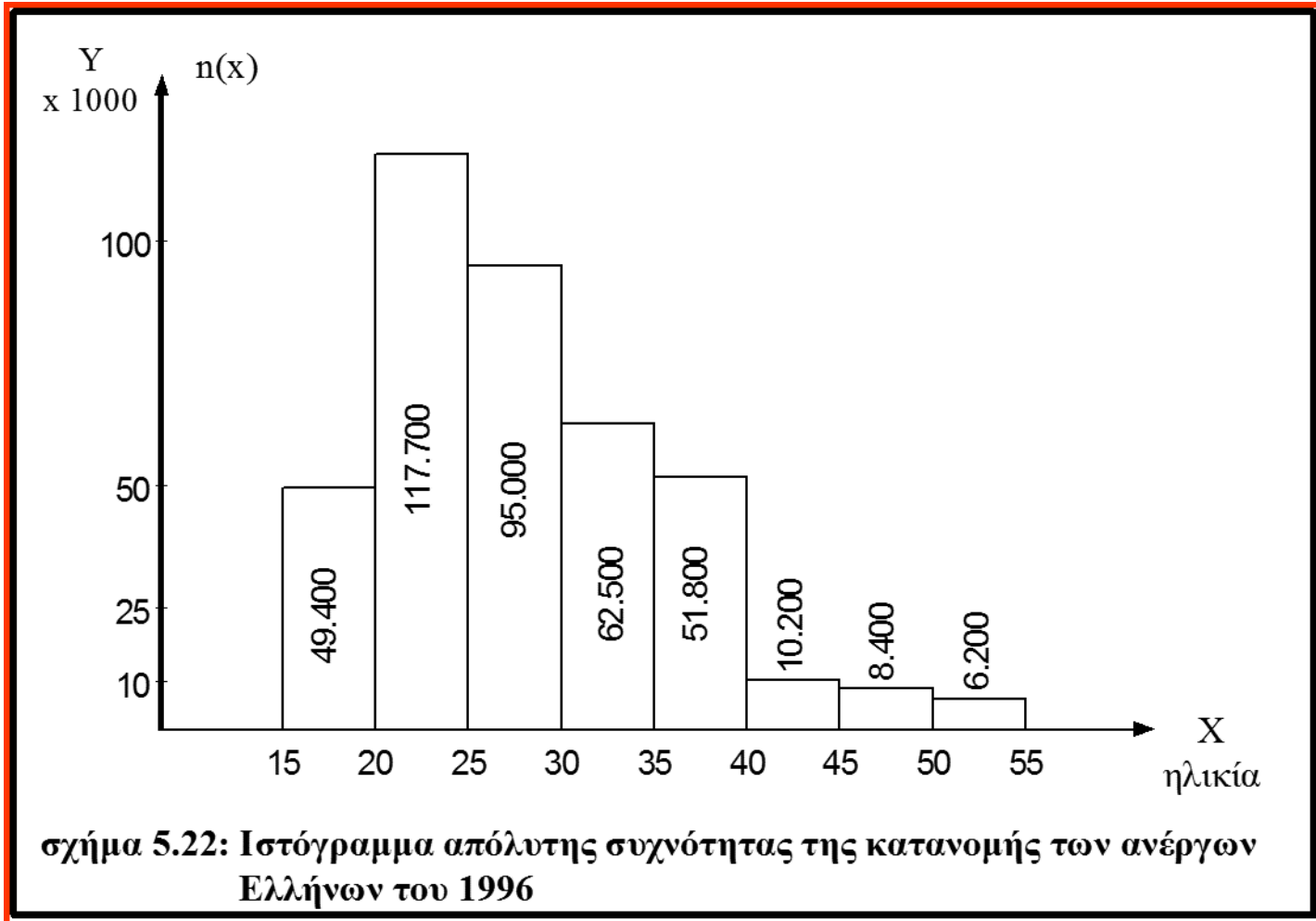
Ηλικία (X)	Αριθμός ανέργων $n(x)$	Σχετική συχνότητα $f(x)$
[15,20)	49.400	12.3
[20,25)	117.700	29.3
[25,30)	95.000	23.7
[30,35)	62.500	15.6
[35,40)	51.800	13.0
[40,45)	10.200	2.5
[45,50)	8.400	2.1
[50,55)	6.200	1.5
Σύνολο	401.200	100.0

Πηγή : Στατιστική Επετηρίδα ΕΣΥΕ 1998

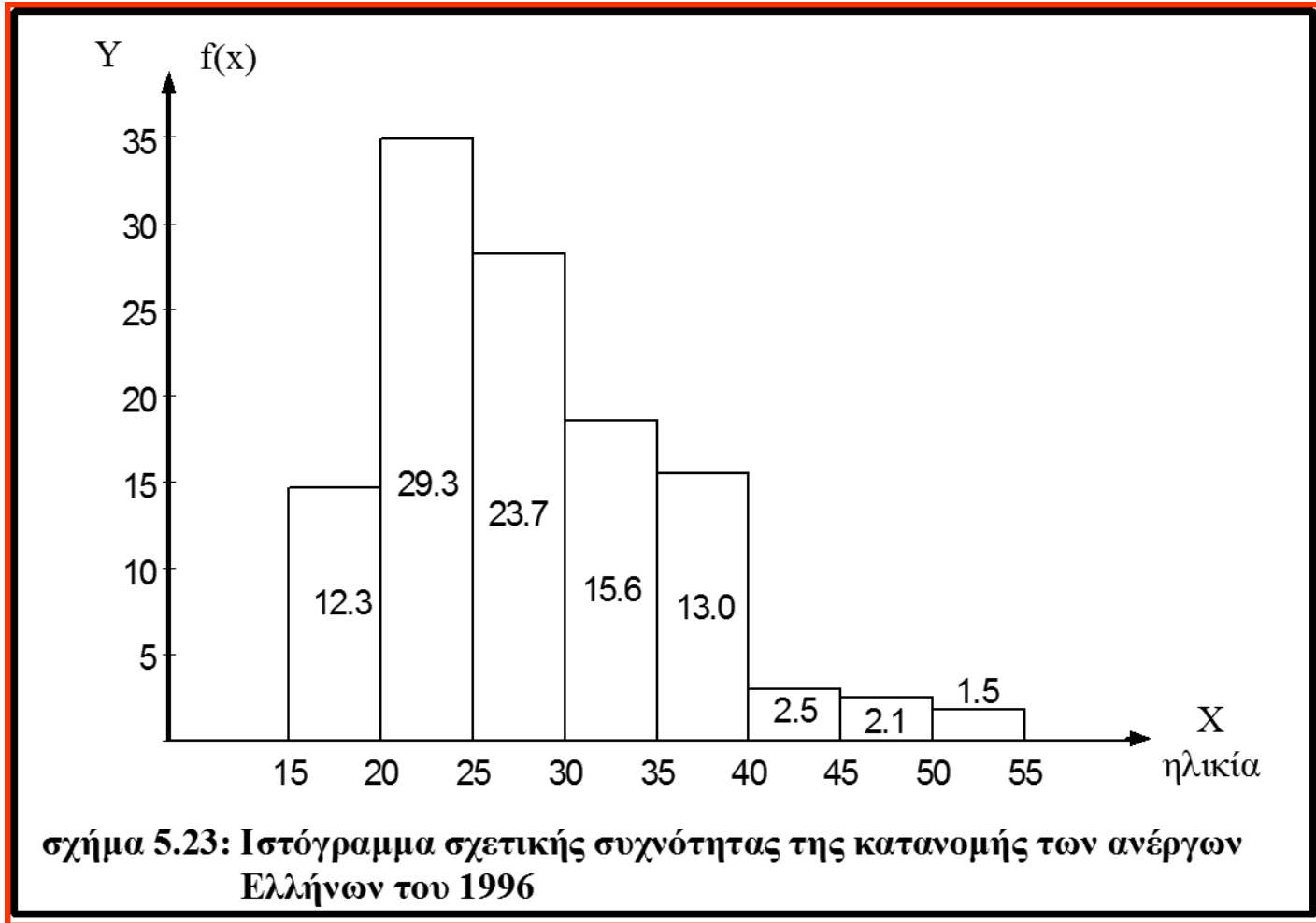
Πηγή: Παπαδημητρίου (2001)



Ιστογράμματα (2)

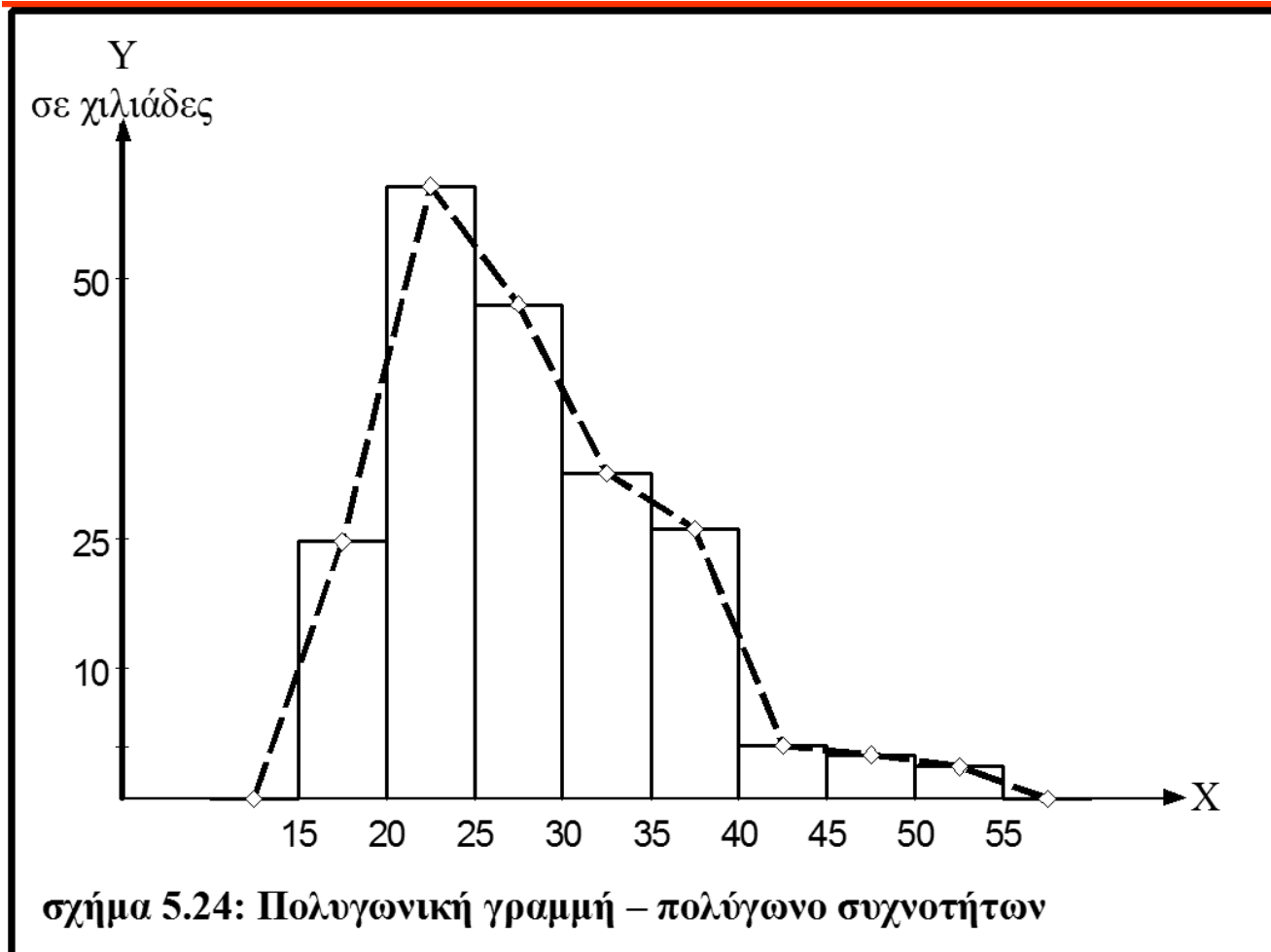


Ιστογράμματα (3)



Πηγή: Παπαδημητρίου (2001)

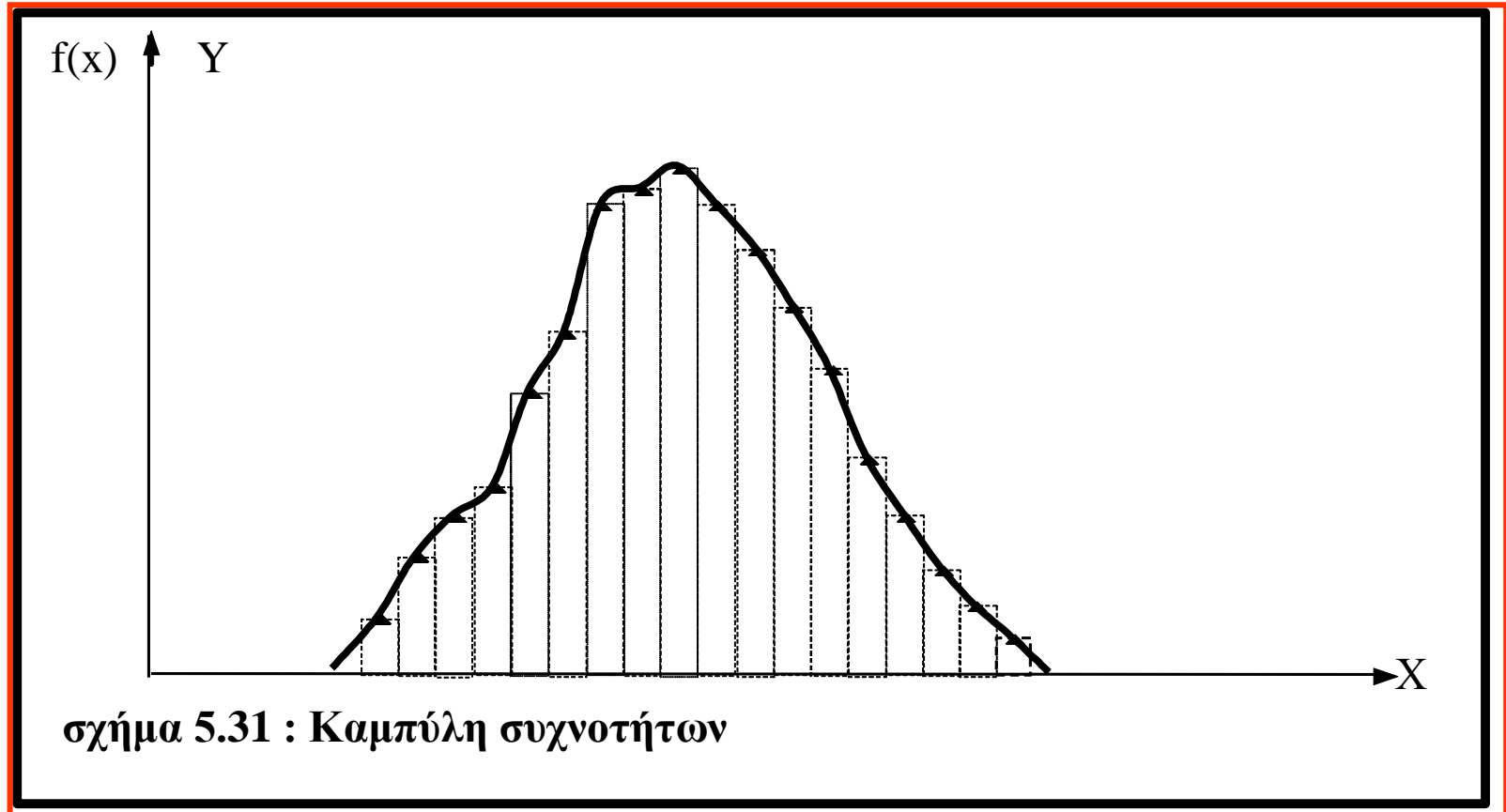
Ιστογράμματα (4)



Πηγή: Παπαδημητρίου (2001)



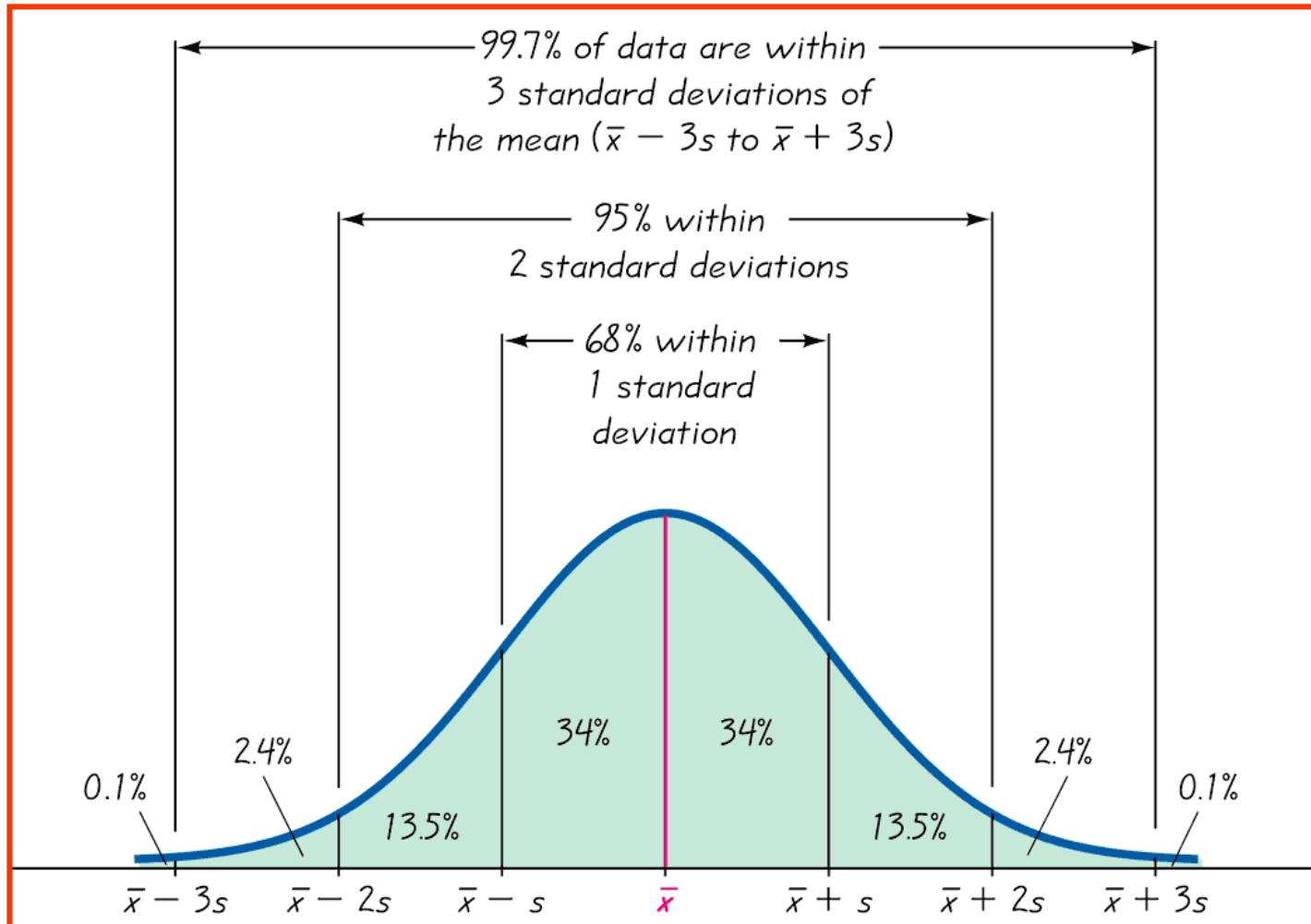
Ιστογράμματα (5)



Πηγή: Παπαδημητρίου (2001)



Κανονική Κατανομή



Βασικές Έννοιες και Ορισμοί (12)

- Στους πίνακες μπορούμε να χρησιμοποιήσουμε τις **απόλυτες συχνότητες**, τις **σχετικές συχνότητες** ή τις **αθροιστικές συχνότητες**.
- Στα διαγράμματα μπορούμε να χρησιμοποιήσουμε **ραβδοδιαγράμματα**, **πολύγωνα συχνοτήτων**, **κυκλικά διαγράμματα (πίτες)** κ.ά.



Βασικές Έννοιες και Ορισμοί (13)

- Όταν η μεταβλητή που θέλουμε να περιγράψουμε έχει πολλές κατηγορίες ή τιμές, η περιγραφή βελτιώνεται αν **ομαδοποιήσουμε** τα δεδομένα πριν υπολογίσουμε τις συχνότητες.
- Για να ομαδοποιήσουμε τα δεδομένα δημιουργούμε νέες κατηγορίες κάθε μία από τις οποίες μπορεί να περιέχει ίσο ή άνισο αριθμό από τις αρχικές τιμές.



Χωρισμός των τιμών των Ποσοτικών Μεταβλητών σε Κλάσεις (I)

- Αν δεν υπάρχει ή δεν είναι επιθυμητό να εφαρμοστεί κάποιο τέτοιο λογικό κριτήριο κατηγοριοποίησης των τιμών του εισοδήματος θα μπορούσε να χρησιμοποιηθεί ένα από τα παρακάτω στατιστικά κριτήρια:
- Η κατασκευή **διαστημάτων** του τύπου «**από-έως**» με όρια των διαστημάτων τα τεταρτημόρια (*quartiles*) (Q_{25} , Q_{50} και Q_{75}) της κατανομής του εισοδήματος.
- Η κατασκευή **διαστημάτων με ίσο περιεχόμενο (%)** των διαθέσιμων παρατηρήσεων. Για παράδειγμα, μπορούμε να κατασκευάσουμε τρία διαστήματα που το καθένα να περιλαμβάνει το 33,3% των παρατηρήσεων του εισοδήματος. Αν υποθέσουμε ότι έχουμε 300 μετρήσεις του εισοδήματος, το κάθε ένα από τα τρία διάστημα να περιλαμβάνει 100 παρατηρήσεις.
- Η κατασκευή **διαστημάτων ίσου εύρους**. Αν υποθέσουμε ότι το εύρος του εισοδήματος είναι r τότε μπορούμε να κατασκευάσουμε d διαστήματα με εύρος r/d το καθένα.



Χωρισμός των τιμών των Ποσοτικών Μεταβλητών σε Κλάσεις (II)

- Η κατασκευή δύο διαστημάτων του τύπου «κάτω από-πάνω από» με βάση κάποια οριακή κρίσιμη τιμή που να έχει ιδιαίτερη σημασία στο θεωρητικό πλαίσιο στο οποίο θα ερμηνευτούν τα αποτελέσματα της έρευνας.
- Η κατασκευή διαστημάτων με την εφαρμογή μεθόδων της ασαφούς λογικής (*fuzzy logic*). Όμως, οι τεχνικές αυτές είναι αρκετά πολύπλοκες και απαιτούν τη χρήση ειδικών λογισμικών.
- Εφαρμογή Μεθόδων Βέλτιστης Κατηγοριοποίησης.

Στην πράξη, οι τέσσερις πρώτες μέθοδοι αποδεικνύονται απλές και λειτουργικές.

Η επιλογή του στατιστικού κριτηρίου κατηγοριοποίησης των μεταβλητών πρέπει να είναι τέτοια ώστε οι κλάσεις που θα προκύψουν να έχουν πρακτική σημαντικότητα και ερμηνευτική λειτουργικότητα στο θεωρητικό πλαίσιο που διέπει την εκάστοτε έρευνα.



Βασικές Έννοιες και Ορισμοί (14)

- Όταν οι κατανομές συχνοτήτων περιγράφουν μια μόνο μεταβλητή κάθε φορά, ονομάζονται **μονομεταβλητές** συχνότητες.
- Όταν περιγράφουν την κατανομή μιας μεταβλητής ανά κατηγορία κάποιας άλλης, λέγονται **διμεταβλητές**.



Βασικές Έννοιες και Ορισμοί (15)

- Οι πίνακες που χρησιμοποιούνται για την παρουσίαση των διμεταβλητών συχνοτήτων έχουν τόσες σειρές (r) όσες οι κατηγορίες της μιας μεταβλητής και τόσες στήλες (c) όσες οι κατηγορίες της άλλης μεταβλητής.
- Στα κελιά που δημιουργούνται από το συνδυασμό σειρών και στηλών ($r \times c$) γράφουμε τις συχνότητες που αντιστοιχούν.



Βασικές Έννοιες και Ορισμοί (16)

- Στις διμεταβλητές συχνότητες τα ποσοστά (και εκατοστιαία ποσοστά) μπορούν να υπολογιστούν κατά τρεις διαφορετικούς τρόπους:
 - α) διαιρώντας την απόλυτη συχνότητα του κελιού διά του συνόλου της σειράς,
 - β) διαιρώντας την απόλυτη συχνότητα του κελιού διά του συνόλου της στήλης,
 - γ) διαιρώντας την απόλυτη συχνότητα του κελιού διά του γενικού συνόλου.



Βασικές Έννοιες και Ορισμοί (17)

- ✓ Τα μέτρα **κεντρικής τάσης** και **διασποράς** χρησιμοποιούνται για την περιγραφή μεταβλητών.
- ✓ Υπάρχουν μέτρα για όλες τις κλίμακες μέτρησης, αλλά τα πιο σημαντικά μέτρα (ο **αριθμητικός μέσος** και η **τυπική απόκλιση**) απαιτούν **ποσοτικές** (ίσων διαστημάτων, αναλογικές) κλίμακες.



Βασικές Έννοιες και Ορισμοί (18)

- Τα κυριότερα μέτρα κεντρικής τάσης είναι: Η επικρατούσα τιμή, η διάμεσος και ο αριθμητικός μέσος.
- Η επικρατούσα τιμή είναι η τιμή ή η κατηγορία με τη μεγαλύτερη συχνότητα.
- Η διάμεσος είναι η τιμή ή η κατηγορία που χωρίζει την κατανομή στη μέση, δηλαδή, οι μισές από τις τιμές ή κατηγορίες είναι χαμηλότερες και άλλες μισές ψηλότερες από τη διάμεσο.



Βασικές Έννοιες και Ορισμοί (19)

- Ο **αριθμητικός μέσος** (ο μέσος όρος) είναι η τιμή που ισοζυγεί την κατανομή.
- Όταν η κατανομή είναι κανονική, όποιο μέτρο και να διαλέξουμε θα μας δώσει την ίδια τιμή.
- Όταν η κατανομή είναι **λοξή**, τότε οι τιμές διαφέρουν.



Βασικές Έννοιες και Ορισμοί (20)

- Τα κυριότερα μέτρα διασποράς είναι η "διασπορά-παραλλακτικότητα-διακύμανση" και η τυπική απόκλιση.
- Η διασπορά είναι η μέση τετραγωνισμένη απόκλιση των τιμών από τον αριθμητικό μέσο.
- Η τυπική απόκλιση είναι η τετραγωνική ρίζα της διασποράς.



Βασικές Έννοιες και Ορισμοί (21)

- Η μετατροπή κατανομών διευκολύνει τη σύγκριση δύο τιμών από διαφορετικές κατανομές.
- Όταν θέλουμε να αλλάξουμε τον αριθμητικό μέσο μιας κατανομής, προσθέτουμε (ή αφαιρούμε) μια σταθερή σε όλες τις τιμές της μεταβλητής.



Βασικές Έννοιες και Ορισμοί (22)

- Όταν θέλουμε να αλλάξουμε την τυπική απόκλιση μιας κατανομής, **πολλαπλασιάζουμε** (ή διαιρούμε) όλες τις τιμές της μεταβλητής επί μια σταθερή.
- Όταν θέλουμε να αλλάξουμε την τυπική απόκλιση και τον αριθμητικό μέσο μιας κατανομής, **πρώτα πολλαπλασιάζουμε** (ή διαιρούμε) για να αλλάξουμε την τυπική απόκλιση και μετά **προσθέτουμε** (ή αφαιρούμε) την επιθυμητή ποσότητα για να αλλάξουμε τον αριθμητικό μέσο.



Βασικές Έννοιες και Ορισμοί (23)

- Η τυπική κατανομή έχει αριθμητικό μέσο 0 και τυπική απόκλιση 1.
- Οι τυπικές τιμές (Z) εκφράζουν τις αρχικές τιμές σε τυπικές αποκλίσεις. Δηλαδή, μια τυπική τιμή 3 δείχνει ότι η αρχική τιμή ήταν τρεις τυπικές αποκλίσεις πάνω από το μέσο όρο και μια τυπική τιμή -2 δείχνει ότι η αρχική τιμή ήταν δύο τυπικές αποκλίσεις κάτω από το μέσο όρο.



Βασικές Έννοιες και Ορισμοί (24)

- Η κανονική τυπική κανονική μας επιτρέπει να την πιθανότητα να πάρει μια μεταβλητή τιμές πάνω ή κάτω από ένα δεδομένο όριο ή μεταξύ δύο ορίων.
- Σε γενικές γραμμές, σε μια κανονική τυπική κανονική 68% όλων των τιμών βρίσκονται μεταξύ -1 και $+1$ και 95% μεταξύ -2 και $+2$.



Βιβλιογραφία (1)

- **Παπαδημητρίου, Γ. (2001).** *Περιγραφική Στατιστική*. Θεσσαλονίκη: Παρατηρητής.
- **Κατσιλής, Ι. (1997).** *Περιγραφική Στατιστική*. Αθήνα: Gutenberg.
- **Μενεξές, Γ. (2008).** Μια Δομημένη Προσέγγιση στην Πολυμεταβλητή Στατιστική Ανάλυση Βιολογικών, Περιβαλλοντικών, Κοινωνικών και Οικονομικών Δεδομένων. Στο *Φυσικοί Πόροι, Περιβάλλον και Ανάπτυξη*, (σσ. 519-534). Επιμέλεια: Γ. Αραμπατζής και Σ. Πολύζος. Θεσσαλονίκη: Εκδόσεις Τζιόλα.
- **Μπεχράκης Θ.Ε. (1999).** *Πολυδιάστατη Ανάλυση Δεδομένων: Μέθοδοι και Εφαρμογές*, Αθήνα: Εκδόσεις ΝΕΑ ΣΥΝΟΡΑ-Α. Α. ΛΙΒΑΝΗΣ.
- **Φωτιάδης, Ν. (1995).** *Εισαγωγή στη Στατιστική για Βιολογικές Επιστήμες*. Θεσσαλονίκη: University Studio Press.



Βιβλιογραφία (2)

- **Carmines, E. & Zeller, R. (1979).** *Reliability and Validity Assessment*. Newbury Park: Sage Publications.
- **Berdie, D., Anderson, J. & Niebuhr, M. (1986).** *Questionnaires: Design and Use*. London: The Scarecrow Press, Inc.
- **Τσάντας, Ν., Μωϋσιάδης, Χ., Μπαγιάτης, Ν., και Χατζηπαντελής, Θ., (1999).** *Ανάλυση Δεδομένων με τη βοήθεια Στατιστικών Πακέτων*. Εκδόσεις ΖΗΤΗ, Θεσσαλονίκη.
- **Κάλλας, Γ. (2002).** *Ζητήματα Σχεδιασμού Εμπειρικών Ερευνών: Αξιοποίηση Μεθόδων της Πληροφορικής Τεχνολογίας*. Αθήνα: Εθνικό Κέντρο Κοινωνικών Ερευνών, Εκδόσεις Νεφέλη.



Σημείωμα Αναφοράς

Copyright Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, Γεώργιος Μενεξές.
«Στατιστική. Εισαγωγή στη Στατιστική». Έκδοση: 1.0. Θεσσαλονίκη 2014.
Διαθέσιμο από τη δικτυακή διεύθυνση:
<https://opencourses.auth.gr/courses/OCRS484/>



Σημείωμα Αδειοδότησης

Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά - Μη Εμπορική Χρήση - Όχι Παράγωγα Έργα 4.0 [1] ή μεταγενέστερη, Διεθνής Έκδοση. Εξαιρούνται τα αυτοτελή έργα τρίτων π.χ. φωτογραφίες, διαγράμματα κ.λ.π., τα οποία εμπεριέχονται σε αυτό και τα οποία αναφέρονται μαζί με τους όρους χρήσης τους στο «Σημείωμα Χρήσης Έργων Τρίτων».



Ο δικαιούχος μπορεί να παρέχει στον αδειοδόχο ξεχωριστή άδεια να χρησιμοποιεί το έργο για εμπορική χρήση, εφόσον αυτό του ζητηθεί.

Ως **Μη Εμπορική** ορίζεται η χρήση:

- που δεν περιλαμβάνει άμεσο ή έμμεσο οικονομικό όφελος από την χρήση του έργου, για το διανομέα του έργου και αδειοδόχο
- που δεν περιλαμβάνει οικονομική συναλλαγή ως προϋπόθεση για τη χρήση ή πρόσβαση στο έργο
- που δεν προσπορίζει στο διανομέα του έργου και αδειοδόχο έμμεσο οικονομικό όφελος (π.χ. διαφημίσεις) από την προβολή του έργου σε διαδικτυακό τόπο

[1] <http://creativecommons.org/licenses/by-nc-nd/4.0/>





Τέλος Ενότητας

Επεξεργασία: Μαρία Αλεμπάκη
Θεσσαλονίκη, Εαρινό Εξάμηνο 2014-2015

