



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εθνικόν και Καποδιστριακόν
Πανεπιστήμιον Αθηνών

Μεθοδολογία των Επιστημών του Ανθρώπου: Στατιστική

Ενότητα 1: Περιγραφική Στατιστική

Βασίλης Γιαλαμάς
Σχολή Επιστημών της Αγωγής
Τμήμα Εκπαίδευσης και Αγωγής στην Προσχολική
Ηλικία

Περιεχόμενα ενότητας

Παρουσιάζονται βασικές έννοιες της περιγραφικής στατιστικής όπως πληθυσμός, δείγμα, μεταβλητή, κατανομή συχνοτήτων και τα χαρακτηριστικά της.



Επικρατούσα τιμή

- Σε περιπτώσεις, που διαφορετικές τιμές μιας μεταβλητής επαναλαμβάνονται περισσότερο από μια φορά, η επικρατούσα τιμή είναι η συχνότερη τιμή στα δεδομένα.
 - Για παράδειγμα στα δεδομένα 7, 7, 7, 8, 8, 8, 8, 10, 10, 10, 10, 10, 12, 12, 13 επικρατούσα τιμή είναι η 10.
- Όταν όλες οι τιμές έχουν την ίδια συχνότητα, δεν υπάρχει επικρατούσα τιμή.
- Αν δύο γειτονικές τιμές έχουν την ίδια συχνότητα, που συγχρόνως είναι η μεγαλύτερη απ' όλες, τότε παίρνουμε ως επικρατούσα τιμή το ημιάθροισμα των δύο τιμών.
- Στην περίπτωση μιας δικόρυφης κατανομής υπάρχουν δυο επικρατέστερες τιμές και είναι οι τιμές που βρίσκονται κάτω από τις δύο κορυφές της κατανομής.

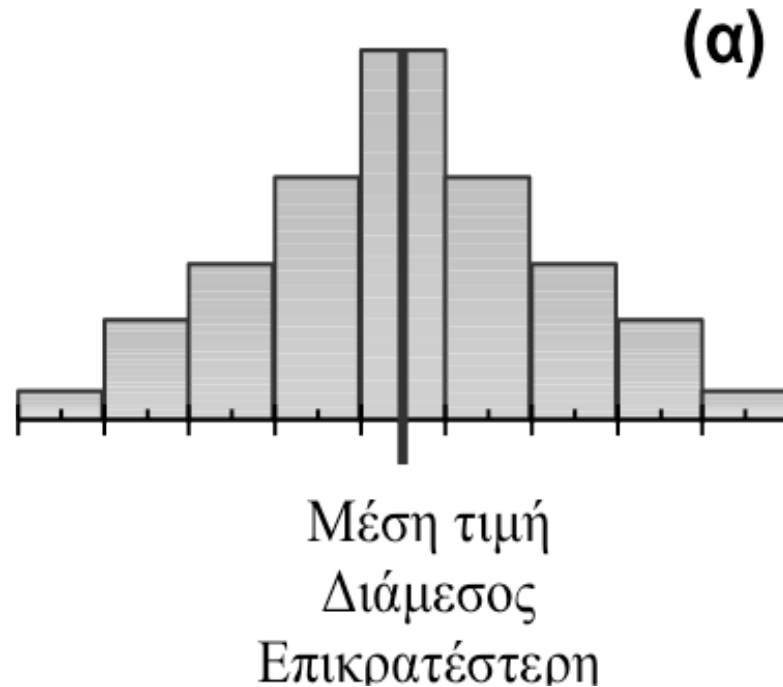


Σύγκριση μέσου, διαμέσου και επικρατούσης τιμής.

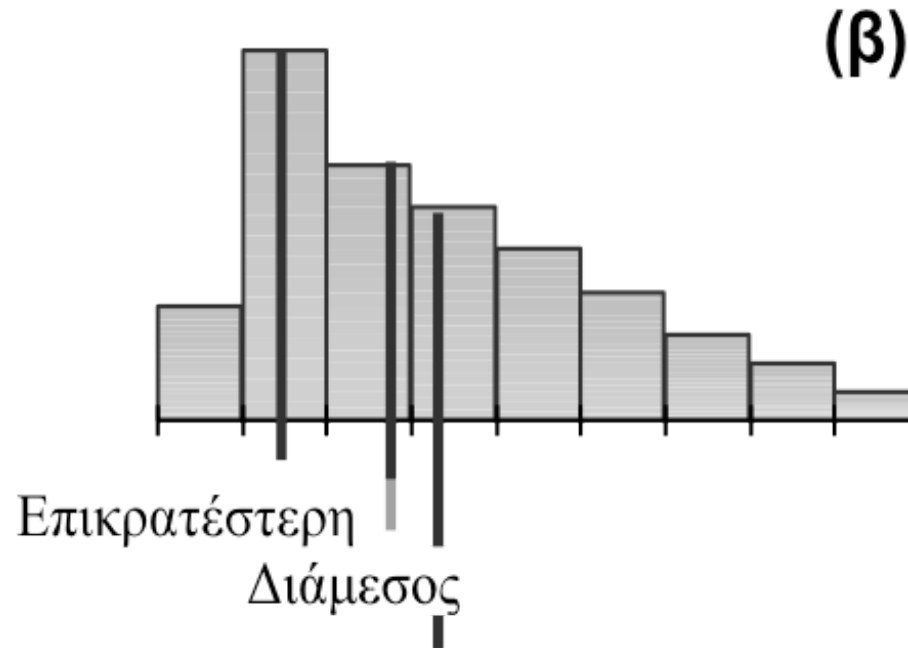
- Στο μέσο όρο, όλες οι τιμές συμμετέχουν κατά τον υπολογισμό του και είναι καθαρό μέτρο κεντρικής τάσης.
- Η διάμεσος στηρίζεται στη σειρά των τιμών. Όπως είδαμε, αν οι τιμές διαταχθούν, διάμεσος είναι η κεντρική τιμή.
- Η επικρατούσα δεν εξαρτάται από το μέγεθος των τιμών, ούτε από τη σειρά τους, αλλά από τη συχνότητά τους.
- Μπορούμε να συγκρίνουμε τα τρία μέτρα, τοποθετώντας τους στη γραφική παράσταση της κατανομής συχνοτήτων μιας μεταβλητής. Ο μέσος είναι το κέντρο βάρους της κατανομής. Η κάθετη ευθεία στον άξονα των τιμών, στο σημείο που βρίσκεται η διάμεσος, χωρίζει το εμβαδόν που ορίζει η κατανομή σε 2 ίσα μέρη. Η επικρατούσα είναι η τιμή που αντιστοιχεί στο ψηλότερο σημείο της καμπύλης.



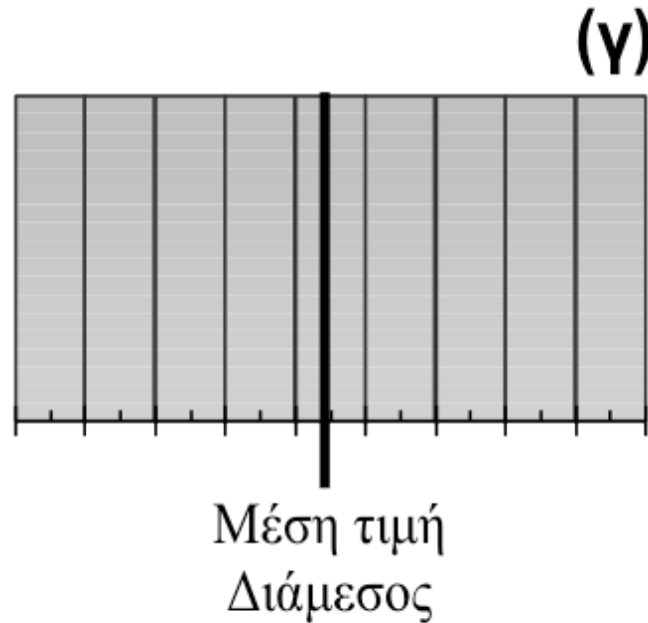
Μορφές της κατανομής συχνοτήτων ποσοτικών μεταβλητών και σχετικές θέσεις των μέτρων κεντρικής θέσης (1 από 5)



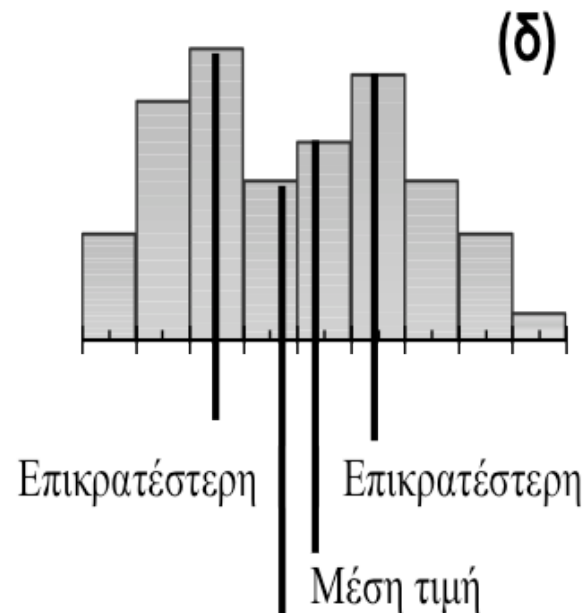
Μορφές της κατανομής συχνοτήτων ποσοτικών μεταβλητών και σχετικές θέσεις των μέτρων κεντρικής θέσης (2 από 5)



Μορφές της κατανομής συχνοτήτων ποσοτικών μεταβλητών και σχετικές θέσεις των μέτρων κεντρικής θέσης (3 από 5)



Μορφές της κατανομής συχνοτήτων ποσοτικών μεταβλητών και σχετικές θέσεις των μέτρων κεντρικής θέσης (4 από 5)



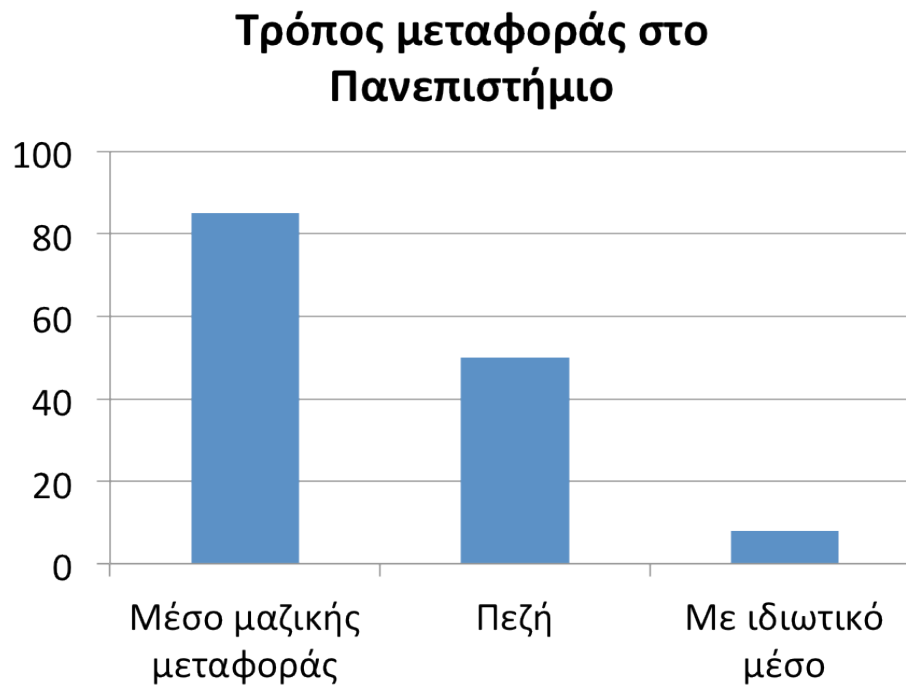
Μορφές της κατανομής συχνοτήτων ποσοτικών μεταβλητών και σχετικές θέσεις των μέτρων κεντρικής θέσης (5 από 5)

Στο ερώτημα, **ποιο μέτρο χρησιμοποιείται συνήθως**, η απάντηση εξαρτάται από:

- Τον τύπο των δεδομένων
- Από τη μορφή της κατανομής των τιμών
 - Η κατανομή εμφανίζει μεγάλη ασυμμετρία (δ)
 - Υπάρχουν ιδιαίτερα ακραίες τιμές (α)
 - Κάποια υποκείμενα παίρνουν απροσδιόριστες τιμές (β)
 - Η κατανομή είναι ανοικτών ορίων (γ)



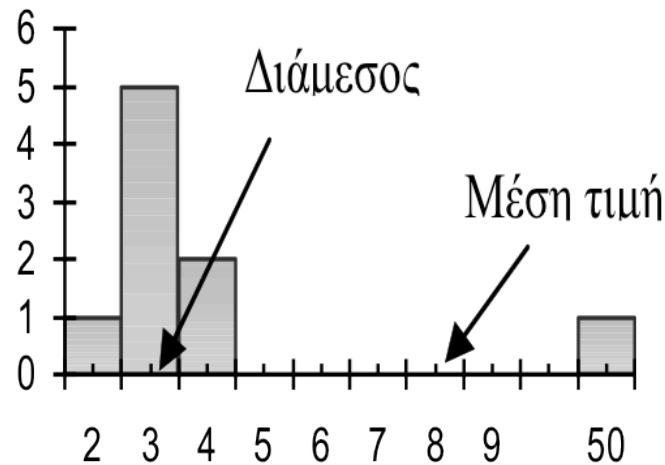
Ποιο μέτρο κεντρικής θέσης είναι κατάλληλο; (1 από 4)



Κατηγορικά δεδομένα



Ποιο μέτρο κεντρικής θέσης είναι κατάλληλο; (2 από 4)



Σχήμα 4.2 Κατανομή συχνοτήτων οικογενειακού εισοδήματος

Ασυμμετρία ή/και παρουσία ακραίων τιμών



Ποιο μέτρο κεντρικής θέσης είναι κατάλληλο; (3 από 4)

μαθητής	Χρόνος (min)
1	7
2	10
3	14
4	15
5	Δεν τέλειωσε



Χρόνος που χρειάστηκε ο μαθητής για να ολοκληρώσει ένα έργο στα πλαίσια μιας 45λεπτης διδακτικής ώρας.

Κάποια υποκείμενα παίρνουν απροσδιόριστες τιμές.



Ποιο μέτρο κεντρικής θέσης είναι κατάλληλο; (4 από 4)

Αριθμός παιδιών	f
0	4
1	8
2	7
3	3
4 και άνω	2

Κατανομή του αριθμού παιδιών σε ένα σύνολο 24 οικογενειών.

Η κατανομή είναι ανοικτών ορίων.



Μέτρα Διασποράς

Έννοια της μεταβλητότητας

Είναι φανερή η διαφορά διακύμανσης στις τρεις παρακάτω ομάδες δεδομένων A, B και Γ παρά το ότι έχουν την ίδια μέση τιμή $\bar{X} = 17$.

A: 14, 16, 17, 18, 20

B: 1, 5, 14, 25, 40

Γ: 17, 17, 17, 17, 17



Εύρος

- **Το εύρος** (L) μιας ομάδας δεδομένων, ορίζεται ως διαφορά της μικρότερης από τη μεγαλύτερη τιμή. Από τον πίνακα 2.2, για τις τιμές WPPSI έχουμε $L = 25 - 2 = 23$.
- Το εύρος του 50% των μεσαίων τιμών, ονομάζεται **ενδοτεταρτημοριακό εύρος (ΕΤΕ)**.
- **ΕΤΕ = $Q_3 - Q_1$** όπου:
 - Q_1 το 1^ο τεταρτημόριο, δηλ το 25% των περιπτώσεων έχουν τιμή $\leq Q_1$
 - Q_3 το 3^ο τεταρτημόριο δηλ το 75% των περιπτώσεων έχουν τιμή $\leq Q_3$



Υπολογισμός ΕΤΕ σε Ομαδοποιημένη Κατανομή Συχνοτήτων (1 από 2)

Η Q_1 βρίσκεται στο πρώτο διάστημα (k), για το οποίο $cf > N/4$ και υπολογίζουμε:

$$Q_1 = L_k + \frac{\frac{N}{4} - cf_{k-1}}{f_k} \cdot h$$

Η Q_3 βρίσκεται στο πρώτο διάστημα (k), για το οποίο $cf > 3N/4$ και υπολογίζουμε:

$$Q_3 = L_k + \frac{\frac{3N}{4} - cf_{k-1}}{f_k} \cdot h$$

Διάστημα	Κέντρο διαστήματος	Συχνότητα		Αθροιστική Συχνότητα	
		Απόλυτη	Σχετική	Απόλυτη	Σχετική
2 – 4	3	1	1,7	1	1,7
5 – 7	6	9	15,0	10	16,7
8 – 10	9	19	31,7	29	48,4
11 – 13	12	11	18,3	40	66,7
14 – 16	15	6	10,0	46	76,7
17 – 19	18	7	12,6	53	88,3
20 – 22	21	5	8,3	58	96,6
23 – 25	24	2	3,4	60	101,0
Σύνολο		60	101,0		

Υπολογισμός ΕΤΕ σε Ομαδοποιημένη Κατανομή Συχνοτήτων (2 από 2)

$$\begin{aligned}Q_1 &= 7,5 + \frac{60/4 - 10}{19} \cdot 3 \\ &= 7,5 + 0,79 \\ &= 8,29\end{aligned}$$

$$\begin{aligned}Q_3 &= 13,5 + \frac{(3 \cdot 60)/4 - 40}{6} \cdot 3 \\ &= 13,5 + 2,5 \\ &= 16\end{aligned}$$

$$\begin{aligned}\text{ΕΤΕ} &= Q_3 - Q_1 \\ &= 16 - 8,29 \\ &= 7,71\end{aligned}$$



Διακύμανση και τυπική απόκλιση (1 από 4)

Έστω για παράδειγμα οι τιμές από κάποια μέτρηση: 8, 13, 18, 24, 27, 30.

Αν θέλουμε τη μεταβλητότητα των τιμών της ομάδας, γύρω από τη μέση τιμή τους που είναι 20, μπορούμε να σχηματίσουμε τις αποκλίσεις x_i από τη μέση τιμή:

$$d_i = X_i - \bar{X}$$

-12, -7, -2, +4, +7, +10.

Η συνολική μεταβλητότητα μπορεί να εκφραστεί ως άθροισμα των αποκλίσεων:

$$\sum_{i=1}^N d_i$$

Και η μέση απόκλιση μια τιμής:

$$\frac{\sum_{i=1}^N d_i}{N}$$



Διακύμανση και τυπική απόκλιση (2 από 4)

Όμως η τιμή $\sum_{i=1}^N d_i$ είναι πάντα ίση με 0, π.χ:

$$\frac{(-12)+(-7)+(-2)+4+7+10}{6} = \frac{0}{6} = 0$$

Το πρόβλημα μπορεί να διορθωθεί αν χρησιμοποιηθούν οι απόλυτες τιμές των αποκλίσεων ως μέτρο μεταβλητότητας

Μέση απόλυτη απόκλιση $\frac{\sum |d_i|}{N}$

$$\frac{|-12|+|-7|+|-2|+|4|+|7|+|10|}{6} = 7$$



Διακύμανση και τυπική απόκλιση (3 από 4)

Διακύμανση της X σε πληθυσμό: $\sigma^2 = \frac{\sum_{i=1}^n (X_i - \mu)^2}{n}$

Τυπική απόκλιση της X σε πληθυσμό: $\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^n (X_i - \mu)^2}{n}}$

X_i = Οι τιμές της μεταβλητής

μ = Η μέση τιμή του πληθυσμού

n = Το μέγεθος του πληθυσμού

$$\sigma^2 = \frac{(-12)^2 + (-7)^2 + (-2)^2 + (4)^2 + (7)^2 + (10)^2}{6} = \frac{362}{6} = 60,3$$



Διακύμανση και τυπική απόκλιση (4 από 4)

Διακύμανση και τυπική απόκλιση δείγματος:

$$S^2 = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N - 1}, \quad S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N - 1}}$$

X_i = Οι τιμές της μεταβλητής

\bar{X} = Η μέση τιμή του δείγματος

N = Το μέγεθος του δείγματος



Υπολογιστικοί τύποι για διακύμανση

Από απλό πίνακα τιμών:
$$S^2 = \frac{\sum_{i=1}^N X_i^2 - \left(\sum_{i=1}^N X_i\right)^2 / N}{N-1}$$

Από πίνακα κατανομής συχνοτήτων:
$$S^2 = \frac{\sum_{k=1}^K (X_k^2 \cdot f_k) - \left(\sum_{k=1}^K (X_k \cdot f_k)\right)^2 / N}{N-1}$$

Από πίνακα ομαδοποιημένης κατανομής συχνοτήτων: Στη θέση των τιμών X_k στον παραπάνω τύπο τοποθετούνται τα κέντρα των διαστημάτων.

Για την τυπική απόκλιση S ισχύει:
$$S = \sqrt{S^2}$$



Υπολογισμός τυπικής απόκλισης στην περίπτωση ομαδοποιημένης κατανομής για την υποκλίμακα WPPSI

$$S^2 = \frac{\sum_{k=1}^K (X_k^2 \cdot f_k) - \left(\sum_{k=1}^K (X_k \cdot f_k) \right)^2 / N}{N - 1}$$

$$\bar{X} = \frac{\sum_{k=1}^K (X_k \cdot f_k)}{N} = \frac{729}{60} = 12,15$$

$$S^2 = \frac{10431 - 729^2/60}{60 - 1} = \frac{1573,65}{59} = 26,67$$

$$S = \sqrt{S^2} = \sqrt{26,67} = 5,16$$

Διάστημα τιμών	Κέντρο X_k	Συχν/τα f_k	$X_k f_k$	X_k^2	$X_k^2 f_k$
2 - 4	3	1	3	9	9
5 - 7	6	9	54	36	324
8 - 10	9	19	171	81	1539
11 - 13	12	11	132	144	1584
14 - 16	15	6	90	225	1350
17 - 19	18	7	126	324	2268
20 - 22	21	5	105	441	2205
23 - 25	24	2	48	576	1152
Σύνολα		60	729		10431

$N = \sum_{k=1}^K f_k$ $\sum_{k=1}^K (X_k \cdot f_k)$ $\sum_{k=1}^K (X_k^2 \cdot f_k)$



Συντελεστής μεταβλητότητας

- Για να συγκριθούν δυο χαρακτηριστικά που μετρώνται με διαφορετικές κλίμακες ή με διαφορετικές μονάδες μέτρησης (π.χ. βάρος (kgr) με ύψος (cm) σε ένα πληθυσμό) δεν μπορεί να χρησιμοποιηθεί κανένα από τα προηγούμενα μέτρα μεταβλητότητας που παρουσιάστηκαν.
- Μια ικανοποιητική προσέγγιση είναι ο **συντελεστής μεταβλητότητας**:

$$CV = \frac{S}{\bar{X}} \cdot 100$$



Οι παράγοντες που επιδρούν στη μεταβλητότητα των τιμών και καθορίζουν την επιλογή του μέτρου διασποράς

- Ακραίες τιμές και έντονη ασυμμετρία
Επιδρούν στο εύρος, διακύμανση και τυπική απόκλιση
- Το μέγεθος του δείγματος
Επιδρά στο εύρος
- Σταθερότητα της δειγματοληψίας
Όταν σχηματίζονται μερικά δείγματα από τον ίδιο πληθυσμό τότε αναμένεται μια ομοιότητα μεταξύ αυτών των δειγμάτων. Στο εύρος δεν υπάρχει αυτή η σταθερότητα.
- Κατανομή ανοικτών ορίων.
Όταν η μεγαλύτερη ή η μικρότερη τιμή μιας κατανομής δεν είναι σαφώς ορισμένες τότε έχουμε μια κατανομή ανοικτών ορίων. Μόνο το ενδοτεταρτημοριακό εύρος υπολογίζεται.



Τέλος Ενότητας

Περιγραφική Στατιστική

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στο πλαίσιο του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Αθηνών**» έχει χρηματοδοτήσει μόνο την 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.



Σημειώματα

Σημείωμα Ιστορικού Εκδόσεων Έργου

Το παρόν έργο αποτελεί την έκδοση 1.0.



Σημείωμα Αναφοράς

Copyright Εθνικών και Καποδιστριακών Πανεπιστημίων Αθηνών 2015, Βασίλης Γιαλαμάς 2015. Βασίλης Γιαλαμάς. «Μεθοδολογία των Επιστημών του Ανθρώπου: Στατιστική. Περιγραφική Στατιστική». Έκδοση: 1.0. Αθήνα 2015. Διαθέσιμο από τη δικτυακή διεύθυνση: <http://opencourses.uoa.gr/courses/ECD102/>.



Σημείωμα Αδειοδότησης

Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά, Μη Εμπορική Χρήση Παρόμοια Διανομή 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».



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

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

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



Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)

μαζί με τους συνοδευόμενους υπερσυνδέσμους.



Σημείωμα Χρήσης Έργων Τρίτων (1/2)

Το Έργο αυτό κάνει χρήση των ακόλουθων έργων:

Εικόνες/Σχήματα/Διαγράμματα/Φωτογραφίες

Εικόνες 1-4, Σελίδες 5-8: Εικόνα με ιστόγραμμα κατανομής συχνοτήτων / Copyright Σ. Πατάκης ΑΕΕΔΕ (Εκδόσεις Πατάκη) και Β. Γιαλαμάς, 2004 / Πηγή: «Στατιστικές Τεχνικές και Εφαρμογές στις Επιστήμες της Αγωγής» Β. Γιαλαμάς, Εκδόσεις Πατάκη

Εικόνα 5, Σελίδα 10: Εικόνα με ιστόγραμμα δεδομένων / Copyrighted

Εικόνα 6, Σελίδα 11: Εικόνα με ιστόγραμμα κατανομής συχνοτήτων οικογενειακού εισοδήματος / Copyright Σ. Πατάκης ΑΕΕΔΕ (Εκδόσεις Πατάκη) και Β. Γιαλαμάς, 2004 / Πηγή: «Στατιστικές Τεχνικές και Εφαρμογές στις Επιστήμες της Αγωγής» Β. Γιαλαμάς, Εκδόσεις Πατάκη



Σημείωμα Χρήσης Έργων Τρίτων (2/2)

Το Έργο αυτό κάνει χρήση των ακόλουθων έργων:

Πίνακες

Πίνακας 1, Σελίδα 12: Πίνακας με τον αριθμό του μαθητή και τον χρόνο που χρειάστηκε / Copyright Σ. Πατάκης ΑΕΕΔΕ (Εκδόσεις Πατάκη) και Β. Γιαλαμάς, 2004 / Πηγή: «Στατιστικές Τεχνικές και Εφαρμογές στις Επιστήμες της Αγωγής» Β. Γιαλαμάς, Εκδόσεις Πατάκη

Πίνακας 2, Σελίδα 13: Πίνακας με κατανομή αριθμού παιδιών / Copyright Σ. Πατάκης ΑΕΕΔΕ (Εκδόσεις Πατάκη) και Β. Γιαλαμάς, 2004 / Πηγή: «Στατιστικές Τεχνικές και Εφαρμογές στις Επιστήμες της Αγωγής» Β. Γιαλαμάς, Εκδόσεις Πατάκη

Πίνακας 3, Σελίδα 17: Πίνακας με κατανομή συχνοτήτων / Copyright Σ. Πατάκης ΑΕΕΔΕ (Εκδόσεις Πατάκη) και Β. Γιαλαμάς, 2004 / Πηγή: «Στατιστικές Τεχνικές και Εφαρμογές στις Επιστήμες της Αγωγής» Β. Γιαλαμάς, Εκδόσεις Πατάκη

Πίνακας 4, Σελίδα 24: Πίνακας συχνοτήτων / Copyright Σ. Πατάκης ΑΕΕΔΕ (Εκδόσεις Πατάκη) και Β. Γιαλαμάς, 2004 / Πηγή: «Στατιστικές Τεχνικές και Εφαρμογές στις Επιστήμες της Αγωγής» Β. Γιαλαμάς, Εκδόσεις Πατάκη

