Ε Λ Λ Η Ν Ι Κ Η   Δ Η Μ Ο Κ Ρ Α Τ Ι Α

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ
UNIVERSITY OF PATRAS

Καθηγητής Ι. Μητρόπουλος
ΤΜΗΜΑ ΔΙΟΙΚΗΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ & ΤΕΧΝΟΛΟΓΙΑΣ
Τηλ.: +030 2610 369213, email: imitro@upatras.gr
Διεύθυνση: Μεγάλου Αλεξάνδρου 1, 263 34 ΠΑΤΡΑ

Ποσοτικές Μέθοδοι στην Οικονομία και Διοίκηση (Ι)

ΕΝΟΤΗΤΑ 4η: Αριθμητικές Μέθοδοι Περιγραφικής Στατιστικής

Αριθμητικές μέθοδοι της Περιγραφικής Στατιστικής…

Δείκτες Κεντρικής Θέσης

Αριθμητικός Μέσος, Διάμεσος, Επικρατούσα Τιμή

Δείκτες Μεταβλητότητας

Εύρος Τιμών, Τυπική Απόκλιση, Διασπορά,
Συντελεστής Μεταβλητότητας

Δείκτες Σχετικής Θέσης

Εκατοστημόρια, Τεταρτημόρια

Δείκτες Γραμμικής Συσχέτισης

Συμμεταβλητότητα, Συσχέτιση, Ευθεία Ελαχίστων Τετραγώνων

Δείκτες Κεντρικής Θέσης …

Ο ***αριθμητικός μέσος***, ή αλλιώς *μέσος όρος*, ή πιο σύντομα απλά *μέσος*, είναι το πιο γνωστό και πιο χρήσιμο μέτρο κεντρικής θέσης.

Υπολογίζεται αθροίζοντας όλες τις τιμές των δεδομένων και διαιρώντας δια το πλήθος τους:

$$\text{Μέσος} = \frac{\text{Άθροισμα δεδομένων}}{\text{Πλήθος δεδομένων}}$$

Συμβολισμός…

Όταν αναφερόμαστε στο πλήθος των δεδομένων για έναν ***πληθυσμό***, χρησιμοποιούμε το κεφαλαίο **N**

Όταν αναφερόμαστε στο πλήθος των δεδομένων για ένα ***δείγμα***, χρησιμοποιούμε το μικρό **n**

Ο αριθμητικός μέσος για έναν ***πληθυσμό*** συμβολίζεται με το ελληνικό γράμμα: μ

Ο αριθμητικός μέσος για έναν ***δείγμα*** συμβολίζεται ως: $\bar{x}$

Αριθμητικός Μέσος …

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

Μέσος Πληθυσμού

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Μέσος Δείγματος

Ο Αριθμητικός Μέσος …

… είναι κατάλληλος για την περιγραφή μετρήσιμων δεδομένων, π.χ. ύψος ατόμων, βαθμολογίες φοιτητών, κλπ.

… επηρεάζεται σημαντικά από τις ακραίες τιμές.

Για παράδειγμα με το που μετακομίζει ένας δισεκατομμυριούχος σε μια γειτονιά, το μέσο οικογενειακό εισόδημα αυξάνει κατά πολύ σε σχέση με την προηγούμενη τιμή του!

Δείκτες Κεντρικής Θέσης ...

Η ***διάμεσος*** υπολογίζεται διατάσσοντας όλα τα δεδομένα. Η τιμή του δεδομένου που βρίσκεται στη *μέση* είναι η διάμεσος.

Δεδομένα: {0, 7, 12, 5, 14, 8, 0, 9, 22} N=9 (περιττός)

Διάταξη σε αύξουσα σειρά, εύρεση της διαμέσου:

0 0 5 7 **8** 9 12 14 22

Παράδειγμα 4.1: Δεδομένα: {0, 7, 12, 5, 14, 8, 0, 9, 22, 33}

N=10 (άρτιος)

Διάταξη σε αύξουσα σειρά, ως διάμεσος θεωρείται ο μέσος όρος των 8 & 9:

0 0 5 7 **8 9** 12 14 22 33

διάμεσος = (8+9)÷2 = **8.5**

Η διάμεσος ενός πληθυσμού και ενός δείγματος υπολογίζονται με τον ίδιο τρόπο.

Μέτρα Κεντρικής Τάσης: Εντοπίζοντας την Διάμεσο

- Η θέση της διαμέσου όταν οι τιμές είναι σε αριθμητική σειρά (μικρότερη έως μεγαλύτερη):

$$\text{Θέση διαμέσου} = \frac{n+1}{2}\ \text{θέση στην αρθμητικήσειρά}$$

- Εάν ο αριθμός των τιμών είναι **μονός**, η διάμεσος είναι ο μεσαίος αριθμός

- Εάν ο αριθμός των τιμών είναι **ζυγός**, η διάμεσος είναι ο μέσος όρος των δύο μεσαίων αριθμών

Σημειώστε ότι $\frac{n+1}{2}$ δεν είναι η τιμή της διάμεσου, μόνο η θέση της διαμέσου στα ταξινομημένα δεδομένα

Δείκτες Κεντρικής Θέσης …

Η ***επικρατούσα τιμή*** ενός συνόλου δεδομένων είναι η τιμή που εμφανίζεται με τη μεγαλύτερη *συχνότητα*.

Ένα σύνολο δεδομένων μπορεί να έχει μία επικρατούσα τιμή (ή επικρατούσα κλάση), ή δύο, ή περισσότερες.

Η επικρατούσα τιμή χρησιμοποιείται για όλους τους τύπους δεδομένων, και κυρίως για ονομαστικά δεδομένα.

Για μεγάλα σύνολα δεδομένων η *επικρατούσα κλάση* είναι καταλληλότερη από μια μόνο επικρατούσα τιμή.

Η επικρατούσα τιμή ενός πληθυσμού και ενός δείγματος υπολογίζονται με τον ίδιο τρόπο.

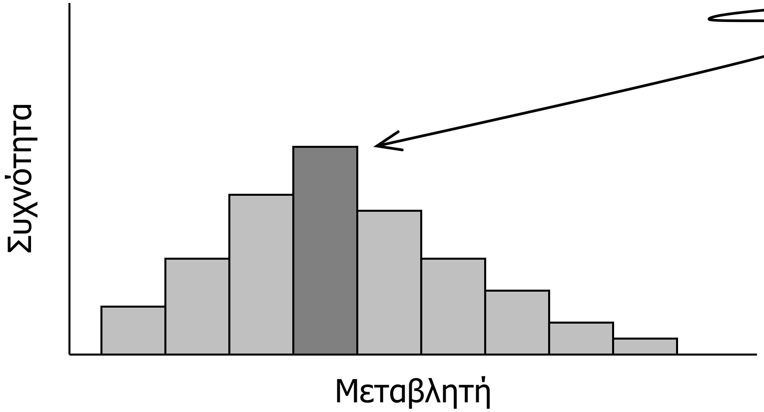
Επικρατούσα τιμή ...

Δεδομένα {**0**, 7, 12, 5, 14, 8, **0**, 9, 22, 33} N=10

Ποια παρατήρηση εμφανίζεται πιο συχνά;

Η επικρατούσα τιμή για τα δεδομένα αυτά είναι **0**. Γιατί είναι μέτρο "κεντρικής" θέσης;

Επικρατούσα κλάση



=MODE(εύρος) στο Excel…

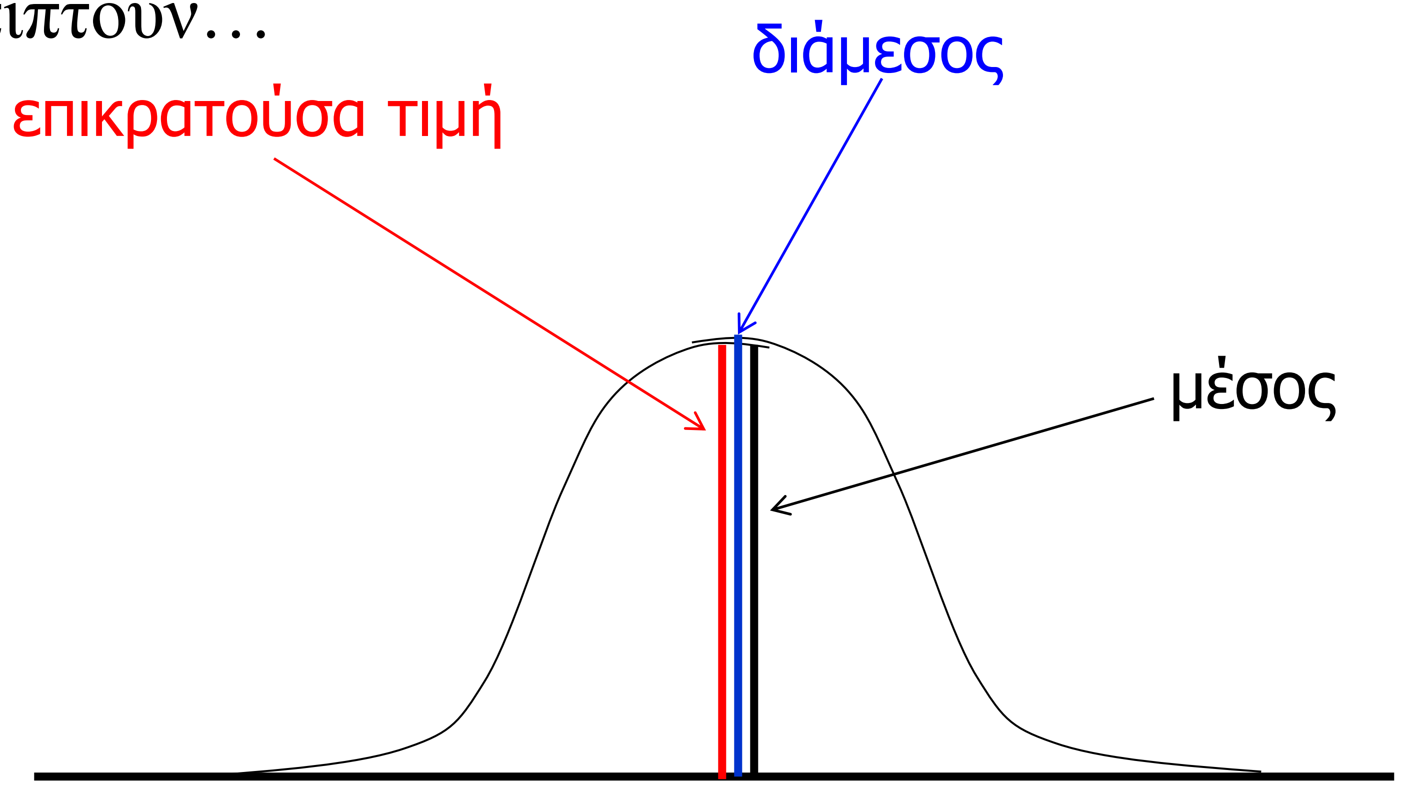
Σημείωση: εάν δουλεύουμε στο Excel και για τα δεδομένα μας υπάρχουν δύο ή περισσότερες επικρατούσες τιμές, το Excel υπολογίζει μόνο τη μικρότερη.

Θα πρέπει να χρησιμοποιηθούν και άλλες τεχνικές (όπως το ιστόγραμμα) για να καθοριστεί εάν υπάρχουν δύο ή περισσότερες επικρατούσες τιμές.

Μέσος, Διάμεσος, Επικρατούσα τιμή …

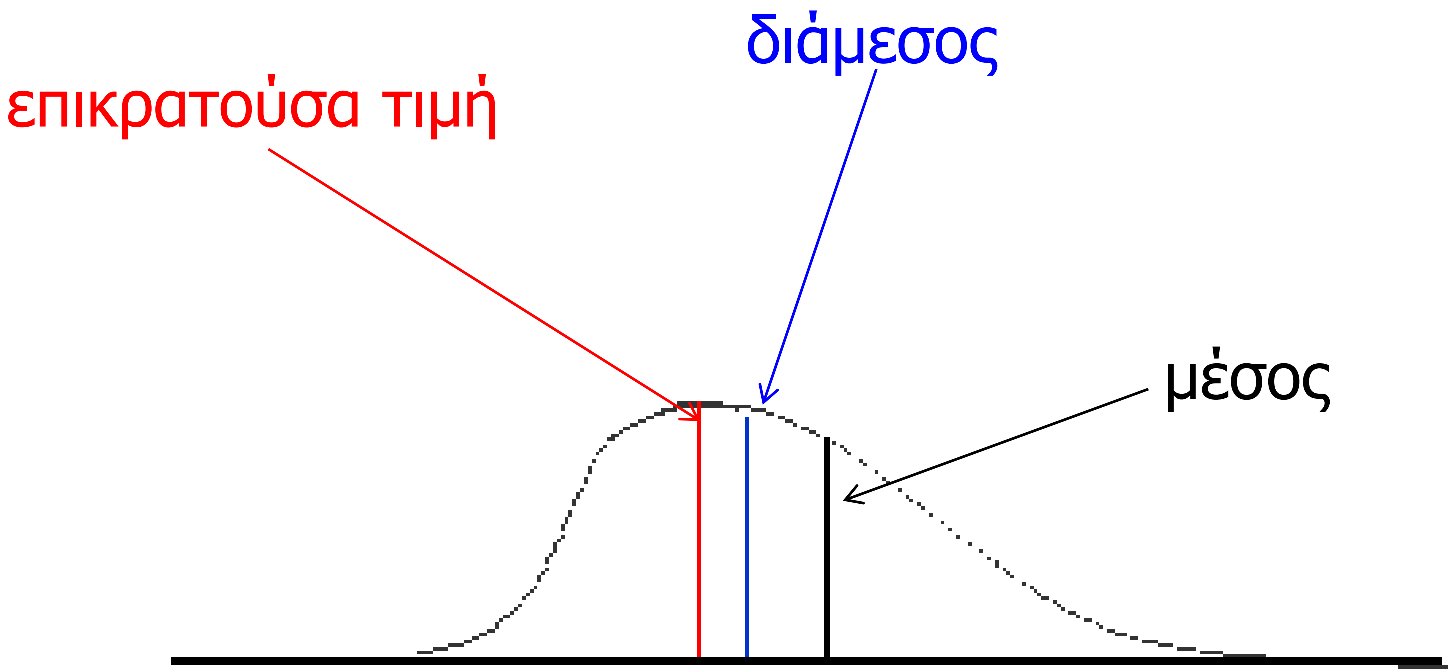
Εάν μια κατανομή είναι συμμετρική,

ο μέσος, η διάμεσος και η επικρατούσα τιμή μπορεί να συμπίπτουν…



Μέσος, Διάμεσος, Επικρατούσα τιμή …

Εάν μια κατανομή δεν είναι συμμετρική, αλλά ασύμμετρη δεξιά - αριστερά, τα τρία μέτρα μπορεί να διαφέρουν. Π.χ.:



Μέσος, Διάμεσος, Επικρατούσα τιμή: Ποιο είναι το καλύτερο;

Έχοντας τρία μέτρα στη διάθεσή μας, ποιο θα πρέπει να χρησιμοποιούμε;

Ο μέσος είναι γενικά η πρώτη επιλογή. Ωστόσο, σε πολλές περιπτώσεις η διάμεσος είναι καλύτερη.

Η επικρατούσα τιμή σπάνια είναι το πιο κατάλληλο μέτρο κεντρικής θέσης.

Ένα πλεονέκτημα της διαμέσου είναι ότι δεν επηρεάζεται από τις ακραίες τιμές τόσο όσο ο μέσος.

Μέσος, Διάμεσος, Επικρατούσα τιμή: Ποιο είναι το καλύτερο;

Θεωρούμε τα δεδομένα του προηγούμενου παραδείγματος.

Ο μέσος ήταν 11.0 και η διάμεσος ήταν 8.5.

Ας υποθέσουμε τώρα ότι αντί για 33 είχαμε ως δεδομένο το 133 . Ο μέσος γίνεται

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{0+7+12+5+133+14+8+0+22}{10} = \frac{210}{10} = 21.0$$

Μέσος, Διάμεσος, Επικρατούσα τιμή: Ποιο είναι το καλύτερο;

Την τιμή αυτή υπερβαίνουν μόνο δύο από τις δέκα παρατηρήσεις του δείγματος, κάνοντας αυτή τη στατιστική μέτρηση να μην αντιπροσωπεύει πια *κεντρική* θέση.

Η διάμεσος παραμένει ίδια. Όταν υπάρχει σχετικά μικρό πλήθος ακραίων τιμών (είτε πολύ μικρές είτε πολύ μεγάλες, αλλά όχι και τα δύο), η διάμεσος συνήθως αποτελεί ένα καλύτερο μέτρο του κέντρου των δεδομένων.

Μέσος, Διάμεσος, & Επικρατούσες τιμές για Διατακτικά & Ονομαστικά Δεδομένα

Για διατακτικά και ονομαστικά δεδομένα ο υπολογισμός του μέσου δεν είναι έγκυρος.

Η διάμεσος είναι κατάλληλη για διατακτικά δεδομένα.

Για ονομαστικά δεδομένα, ο υπολογισμός της επικρατούσας τιμής χρησιμοποιείται για τον καθορισμό της υψηλότερης συχνότητας αλλά όχι ως δείκτης “κεντρικής θέσης”.

Δείκτες Κεντρικής Θέσης • Συνοψίζοντας…

Υπολογίζουμε το Μέσο για

- Να περιγράψουμε την κεντρική θέση ενός μόνο συνόλου ποσοτικών δεδομένων

Υπολογίζουμε τη Διάμεσο για

- Να περιγράψουμε την κεντρική θέση ενός μόνο συνόλου ποσοτικών ή διατακτικών δεδομένων

Υπολογίζουμε την Επικρατούσα τιμή για

- Να περιγράψουμε ένα μόνο σύνολο ονομαστικών δεδομένων

Μέτρα Κεντρικής Τάσης: Παράδειγμα επανάληψης

Τιμές σπιτιών:

$2,000,000
$ 500,000
$ 300,000
$ 100,000
$ 100,000

Σύνολο $ **3,000,000**

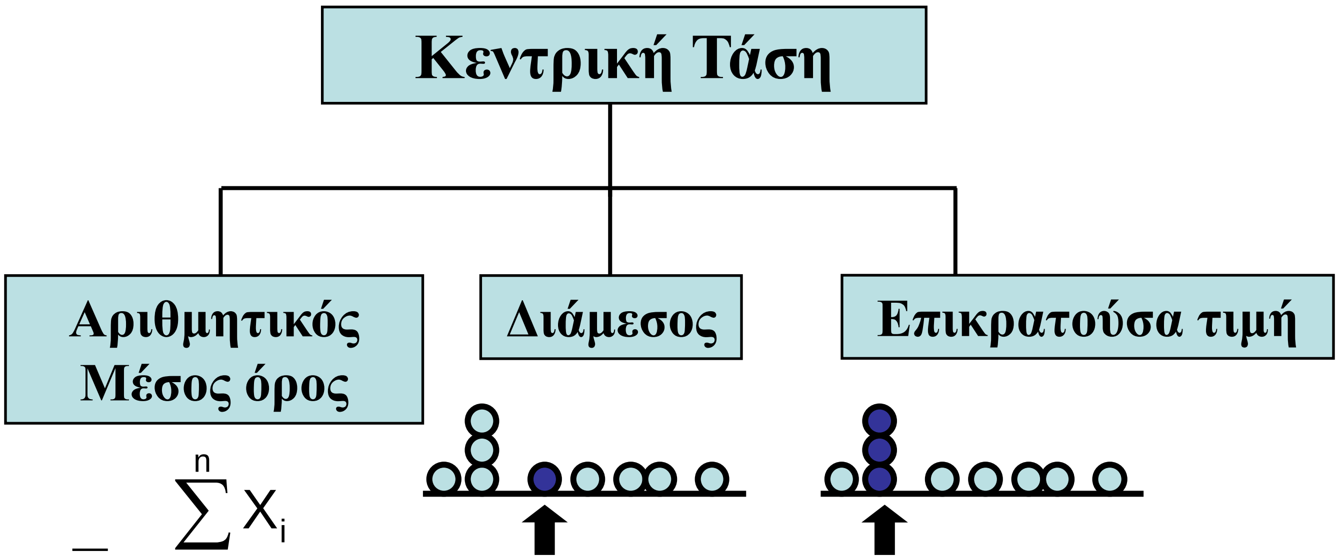
- **Μέσος όρος:** ($3,000,000/5)
  = **$600,000**
- **Διάμεσος:** μέση τιμή ταξινομημένων δεδομένων
  = **$300,000**
- **Επικρατούσα τιμή :** η πιό συχνή τιμή = **$100,000**

Μέτρα Κεντρικής Τάσης:

Ποια μέτρα πρέπει να επιλέξετε;

- Ο **μέσος όρος** χρησιμοποιείται γενικά, εκτός εάν υπάρχουν ακραίες τιμές.
- Η **διάμεσος** χρησιμοποιείται συχνά, αφού η διάμεσος δεν είναι ευαίσθητη σε ακραίες τιμές. Για παράδειγμα, οι διάμεσοι των τιμών κατοικίας μπορεί να αναφέρονται σε μια περιοχή. Είναι λιγότερο ευαίσθητες στις ακραίες τιμές.
- Σε ορισμένες περιπτώσεις είναι λογικό να αναφέρουμε τόσο τον **μέσο όρο** όσο και την **διάμεσο**.

Μέτρα Κεντρικής Τάσης: Περίληψη



$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Μεσσαία τιμή στην σειρά διάταξης

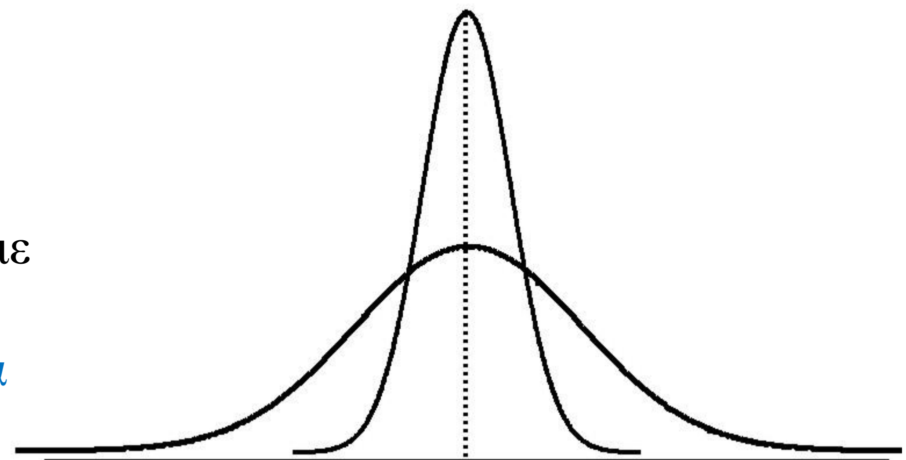
Η πιο συχνά παρατηρούμενη τιμή

Μέτρα Μεταβλητότητας

Μεταβλητότητα

- **Εύρος**
- **Διασπορά**
- **Τυπική Απόκλιση**
- **Συντελεστής μεταβλητότητας**

- Τα μέτρα μεταβλητότητας δίνουν πληροφορίες σχετικά με την εξάπλωση ή τη μεταβλητότητα ή τη διασπορά των τιμών των δεδομένων.



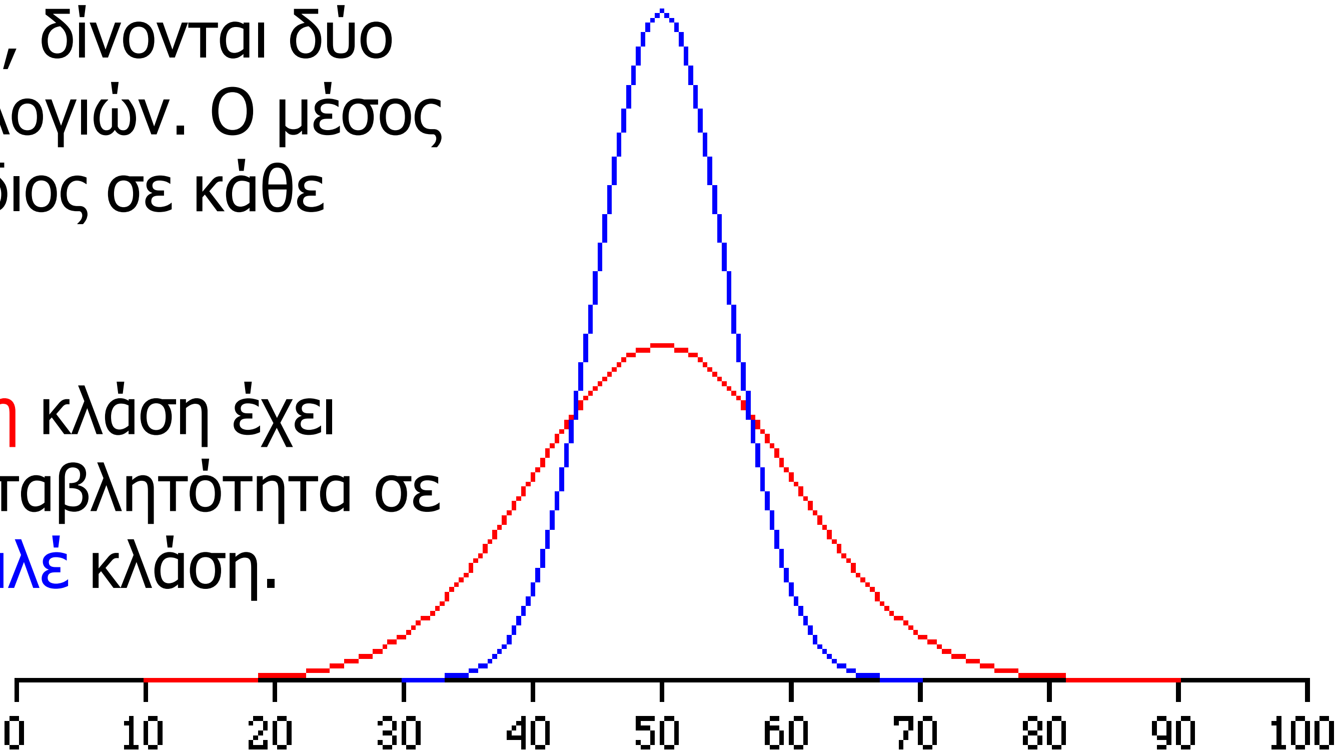
Ίδιο κέντρο, Διαφορετική μεταβλητότητα

Δείκτες Μεταβλητότητας …

Οι δείκτες κεντρικής θέσης δεν καταφέρνουν να περιγράψουν πλήρως τα χαρακτηριστικά μιας κατανομής, όπως το πόσο διαχέονται οι παρατηρήσεις γύρω από τη μέση τιμή.

Για παράδειγμα, δίνονται δύο σύνολα βαθμολογιών. Ο μέσος (=50) είναι ο ίδιος σε κάθε περίπτωση …

Αλλά, η κόκκινη κλάση έχει μεγαλύτερη μεταβλητότητα σε σχέση με τη μπλέ κλάση.



Εύρος Τιμών ...

Το ***εύρος*** είναι ο απλούστερος δείκτης μεταβλητότητας, και υπολογίζεται ως:

Εύρος = Μέγιστη τιμή – Ελάχιστη τιμή

Π.χ.

Δεδομένα: {4, 4, 4, 4, 50} Εύρος = 46

Δεδομένα : {4, 8, 15, 24, 39, 50} Εύρος = 46

Το εύρος είναι το ίδιο και στις δύο περιπτώσεις, αλλά τα σύνολα δεδομένων έχουν πολύ διαφορετική κατανομή …

Εύρος Τιμών ...

Το σημαντικότερο πλεονέκτημά του είναι η ευκολία υπολογισμού του.

Το σημαντικότερο μειονέκτημά του είναι η αδυναμία του να δώσει πληροφορίες για τη διασπορά των παρατηρήσεων μεταξύ των δύο ακραίων τιμών.

Χρειαζόμαστε επομένως ένα μέτρο μεταβλητότητας το οποίο θα ενσωματώνει **όλα τα δεδομένα** και όχι μόνο δύο παρατηρήσεις. Συνεπώς …

Διασπορά ...

Η διασπορά και η τυπική απόκλιση είναι αναμφισβήτητα οι πιο σημαντικοί στατιστικοί δείκτες. Χρησιμοποιούνται για τη μέτρηση της μεταβλητότητας, και παίζουν σημαντικό ρόλο σχεδόν σε όλες τις μεθόδους επαγωγικής στατιστικής.

Η διασπορά του πληθυσμού συμβολίζεται σ^2

Η διασπορά του δείγματος συμβολίζεται s^2

Διασπορά ...

μέσος πληθυσμού

Η διασπορά του **πληθυσμού** είναι: $$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$$

μέγεθος πληθυσμού

μέσος δείγματος

Η διασπορά ενός **δείγματος** είναι: $$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

Σημείωση: Ο παρονομαστής είναι το μέγεθος του δείγματος (n) μείον 1 !

Διασπορά ...

Όπως βλέπετε, πρέπει να υπολογίσετε το μέσο του δείγματος (x- παύλα) προκειμένου να υπολογίσετε τη διασπορά του δείγματος.

Εναλλακτικά, υπάρχει τύπος που υπολογίζει τη διασπορά του δείγματος απευθείας από τα δεδομένα χωρίς το ενδιάμεσο βήμα υπολογισμού του μέσου :

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right]$$

Εφαρμογή ...

Παράδειγμα 4.7. Το ακόλουθο δείγμα αποτελείται από τον αριθμό αιτήσεων για απασχόληση που έκαναν έξι φοιτητές: 17, 15, 23, 7, 9, 13.

Να βρεθεί ο μέσος και η διασπορά του.

Τι ακριβώς θέλουμε να υπολογίσουμε ;

Το ακόλουθο **δείγμα** αποτελείται από τον αριθμό αιτήσεων που έκαναν έξι φοιτητές : 17, 15, 23, 7, 9, 13.

Βρείτε **μέσο** και **διασπορά**.

$\bar{x}$ s^2

… αντί για μ ή σ^2

Μέσος Δείγματος & Διασπορά …

Μέσος Δείγματος

$$\bar{x} = \frac{\sum_{i=1}^{6} x_i}{6} = \frac{17+15+23+7+9+13}{6} = \frac{84}{6} = 14$$

Διασπορά Δείγματος

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{1}{6-1}\left[(17-14)^2 + (15-14)^2 + \ldots (13-14)^2\right] = 33.2$$

Διασπορά Δείγματος (δεύτερος τύπος)

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right] = \frac{1}{6-1}\left[\left(17^2 + 15^2 + \ldots + 13^2\right) - \frac{(17+15+\ldots+13)^2}{6}\right] = 33.2$$

Μέσος Δείγματος & Διασπορά (2)

Παράδειγμα 4.1

Statistics

SPSS

x

N	Valid	10
	Missing	0
Mean		11,0000
Median		8,5000
Mode		,00
Std. Deviation		10,12148
Variance		102,444
Range		33,00
Minimum		,00
Maximum		33,00
Sum		110,00
Percentiles	25	3,7500
	50	8,5000
	75	16,0000

	x	x-μ	(x-μ)*(x-μ)	x2
	0	-11	121	0
	0	-11	121	0
	5	-6	36	25
	7	-4	16	49
	8	-3	9	64
	9	-2	4	81
	12	1	1	144
	14	3	9	196
	22	11	121	484
	33	22	484	1089
Σύνολο =	110	0	922	2132

Μέσος Δείγματος $\bar{X} =$ **11**

Διασπορά Πληθυσμού $\sigma^2 =$ **92,2**

Διασπορά Δείγματος $s^2 =$ **102,444**

Τυπική Απόκλιση ...

Η τυπική απόκλιση είναι απλά η τετραγωνική ρίζα της διασποράς, άρα :

Τυπική απόκλιση πληθυσμού: $\sigma = \sqrt{\sigma^2}$

Τυπική απόκλιση δείγματος: $s = \sqrt{s^2}$

Τυπική Απόκλιση Δείγματος: Παράδειγμα υπολογισμού

Δείγμα Δεδομένων(X_i) :

10	**12**	**14**	**15**	**17**	**18**	**18**	**24**

n = 8 Μέσος όρος = $\overline{\chi}$ = 16

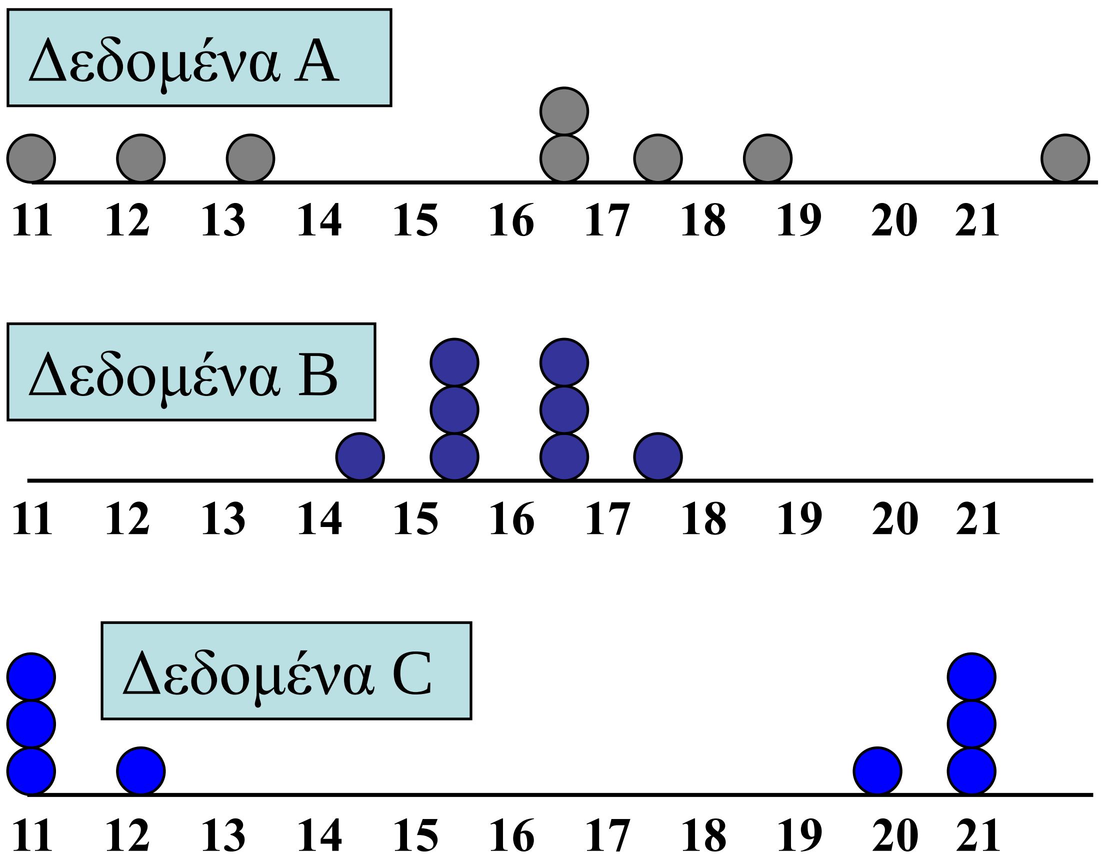
$$S = \sqrt{\frac{(10-\overline{X})^2 + (12-\overline{X})^2 + (14-\overline{X})^2 + \cdots + (24-\overline{X})^2}{n-1}}$$

$$= \sqrt{\frac{(10-16)^2 + (12-16)^2 + (14-16)^2 + \cdots + (24-16)^2}{8-1}}$$

$$= \sqrt{\frac{130}{7}} = \boxed{4.3095}$$

Ένα μέτρο της «μέσης» διασποράς γύρω από τον μέσο όρο

Σύγκριση Τυπικών Αποκλίσεων (1)

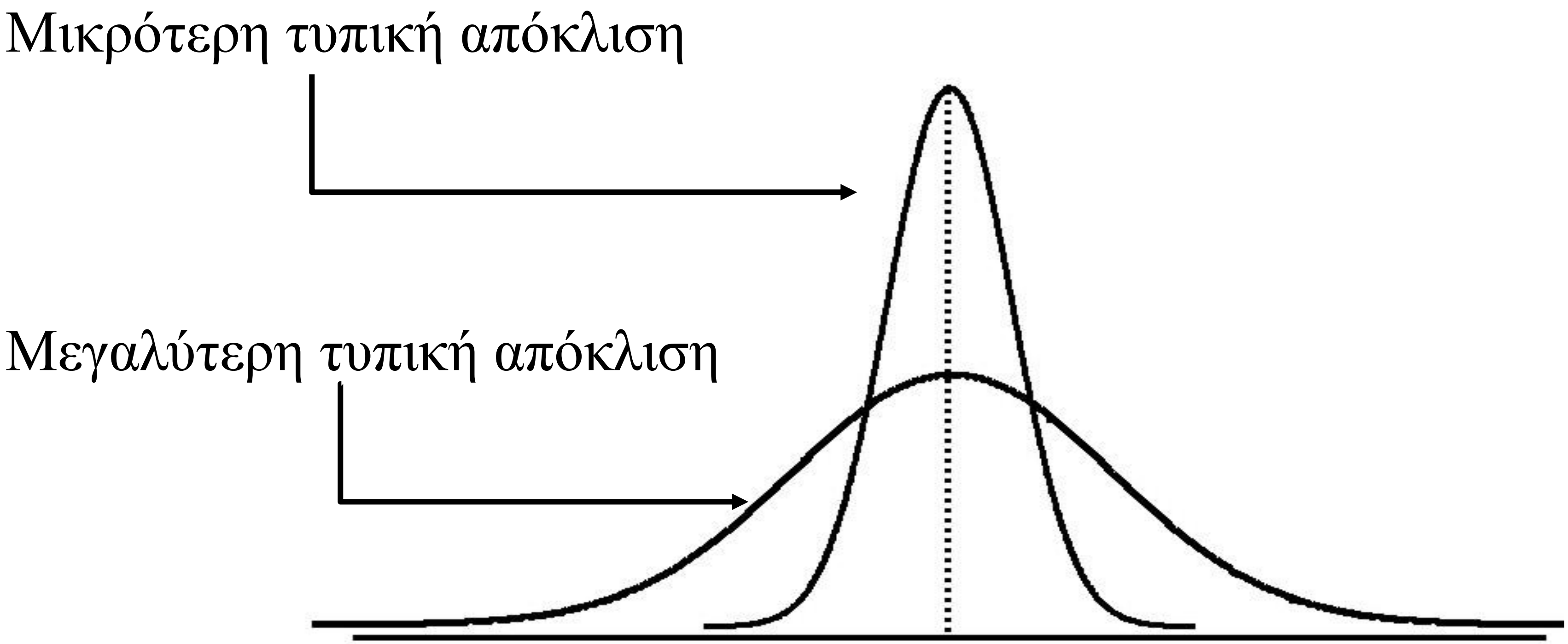


M.O.= 15.5
S = 3.338

M.O. = 15.5
S = 0.926

M.O. = 15.5
S = 4.567

Σύγκριση Τυπικών Αποκλίσεων (2)



Μέτρα Διασποράς:
Συνοπτικά χαρακτηριστικά

- Όσο περισσότερο εκτείνονται τα δεδομένα, τόσο μεγαλύτερο είναι το εύρος, η διακύμανση και η τυπική απόκλιση.
- Όσο περισσότερο συγκεντρώνονται τα δεδομένα, τόσο μικρότερο είναι το εύρος, η διασπορά και η τυπική απόκλιση.
- Εάν οι τιμές είναι όλες οι ίδιες (χωρίς μεταβολή), όλα αυτά τα μέτρα θα είναι μηδέν.
- Κανένα από αυτά τα μέτρα δεν είναι ποτέ αρνητικό.

Τυπική Απόκλιση …

Δείτε το Π.χ. 4.8 [Xm04-08] όπου ένας κατασκευαστής μπαστουνιών γκολφ σχεδίασε ένα νέο μπαστούνι και θέλει να καθορίσει εάν είναι πιο αξιόπιστο (δηλ. με μικρότερη μεταβλητότητα) από το παλαιότερο.

Χρησιμοποιώντας Data > Data Analysis > Descriptive Statistics στο Excel, δημιουργούμε τους πίνακες …

Current 7-iron	
Mean	150.55
Standard Error	0.67
Median	151
Mode	150
Standard Deviation	5.79
Sample Variance	33.55
Kurtosis	0.13
Skewness	-0.43
Range	28
Minimum	134
Maximum	162
Sum	11291
Count	75

New 7-iron	
Mean	150.15
Standard Error	0.36
Median	150
Mode	149
Standard Deviation	3.09
Sample Variance	9.56
Kurtosis	-0.89
Skewness	0.18
Range	12
Minimum	144
Maximum	156
Sum	11261
Count	75

Με το νέο μπαστούνι προκύπτουν πιο αξιόπιστα χτυπήματα.

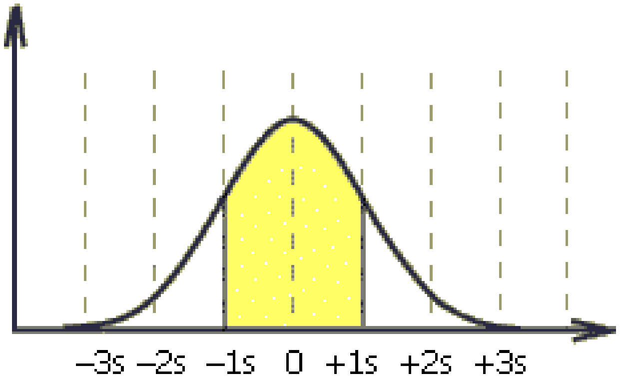
Ερμηνεία της Τυπικής Απόκλισης ...

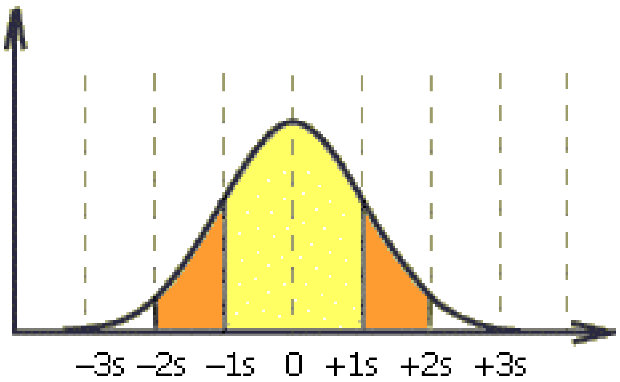
Η τυπική απόκλιση μπορεί να χρησιμοποιηθεί για τη σύγκριση της μεταβλητότητας διαφόρων κατανομών και για την εξαγωγή συμπερασμάτων σχετικά με το γενικό σχήμα της κατανομής. Εάν το ιστόγραμμα έχει **σχήμα καμπάνας**, ωμπορούμε να χρησιμοποιούμε τον ακόλουθο ***Εμπειρικό Κανόνα*** :

1) Περίπου 68% των δεδομένων βρίσκονται σε απόσταση μιας τυπικής απόκλισης από τον μέσο.
2) Περίπου 95% των δεδομένων βρίσκονται σε απόσταση δύο τυπικών αποκλίσεων από τον μέσο.
3) Περίπου 99.7% των δεδομένων βρίσκονται σε απόσταση τριών τυπικών αποκλίσεων από τον μέσο.

Ο Εμπειρικός Κανόνας…

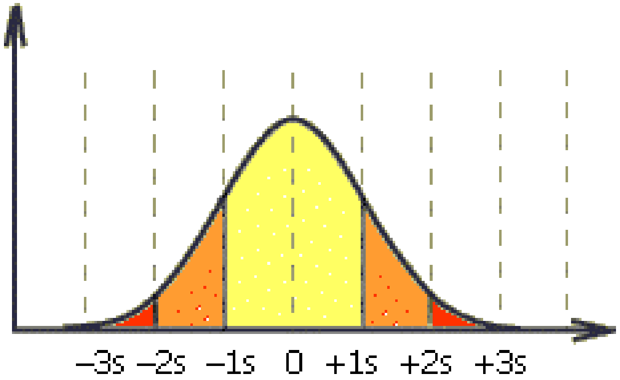
Περίπου 68% των δεδομένων βρίσκονται σε απόσταση **μιας** τυπικής απόκλισης από τον μέσο.





Περίπου 95% των δεδομένων βρίσκονται σε απόσταση **δύο** τυπικών αποκλίσεων από τον μέσο.

Περίπου 99.7% των δεδομένων βρίσκονται σε απόσταση **τριών** τυπικών αποκλίσεων από τον μέσο.



Θεώρημα Chebysheff …

Μια γενικότερη ερμηνεία της τυπικής απόκλισης προκύπτει από το ***Θεώρημα Chebysheff***, το οποίο εφαρμόζεται σε κάθε τύπο ιστογράμματος (όχι μόνο σε αυτά με σχήμα καμπάνας).

Ο λόγος των τιμών των δεδομένων που βρίσκονται σε απόσταση μικρότερη ή ίση από **k** τυπικές αποκλίσεις από τον μέσο είναι *τουλάχιστον:*

$$1-\frac{1}{k^2} \quad for \quad k>1$$

Για k=2, το θεώρημα δείχνει ότι *τουλάχιστον τα* 3/4 των τιμών βρίσκονται σε ακτίνα 2 τυπικών αποκλίσεων από τον μέσο. Αυτό είναι ένα "κάτω φράγμα" σε σχέση με την προσέγγιση του Εμπειρικού Κανόνα (95%).

Ερμηνεία της Τυπικής Απόκλισης

Υποθέτουμε ότι ο μέσος και η τυπική απόκλιση των βαθμών του προηγούμενου εξαμήνου είναι 70 και 5, αντίστοιχα. Εάν το ιστόγραμμα έχει σχήμα καμπάνας τότε γνωρίζουμε ότι:

1. περίπου το 68% των βαθμών είναι μεταξύ 65 και 75,
2. περίπου το 95% των βαθμών είναι μεταξύ 60 και 80, και
3. περίπου το 99.7% των βαθμών είναι μεταξύ 55 και 85.

Εάν το ιστόγραμμα δεν έχει σχήμα καμπάνας τότε μπορούμε να πούμε ότι τουλάχιστον το 75% των βαθμών είναι μεταξύ 60 και 80, και τουλάχιστον το 88.9% των βαθμών είναι μεταξύ 55 και 85. (Μπορούμε να χρησιμοποιήσουμε και άλλες τιμές για το k.)

Συντελεστής Μεταβλητότητας …

Ο ***συντελεστής μεταβλητότητας*** ενός συνόλου δεδομένων είναι το πηλίκο της τυπικής απόκλισης δια τον μέσο τους, δηλαδή:

Συντελεστής μεταβλητότητας πληθυσμού = CV = $\dfrac{\sigma}{\mu}$

Συντελεστής μεταβλητότητας δείγματος = cv = $\dfrac{s}{\bar{x}}$

Συντελεστής Μεταβλητότητας …

Ο συντελεστής αυτός παρέχει ένα ***αναλογικό*** μέτρο διακύμανσης, δηλαδή:

- Μετρά την σχετική μεταβλητότητα
- Πάντα σε ποσοστό (%)
- Εμφανίζει τη μεταβλητότητα σε σχέση με τον μέσο όρο
- Μπορεί να χρησιμοποιηθεί για τη σύγκριση της μεταβλητότητας δύο ή περισσοτέρων συνόλων δεδομένων που μετρήθηκαν σε διαφορετικές μονάδες

Σύγκριση Συντελεστών Μεταβλητότητας (1)

- Απόθεμα Α:
  - –Περσινή μέση τιμή = \$50
  - –Τυπική απόκλιση = \$5

$$CV_A = \left(\frac{S}{\bar{X}}\right) \cdot 100\% = \frac{\$5}{\$50} \cdot 100\% = 10\%$$

- Απόθεμα Β:
  - –Περσινή μέση τιμή = \$100
  - –Τυπική απόκλιση = \$5

$$CV_B = \left(\frac{S}{\bar{X}}\right) \cdot 100\% = \frac{\$5}{\$100} \cdot 100\% = 5\%$$

Και τα δύο αποθέματα έχουν την ίδια τυπική απόκλιση, αλλά το απόθεμα Β είναι λιγότερο μεταβλητό σε σχέση με την τιμή του

Σύγκριση συντελεστών μεταβλητότητας (2)

- Απόθεμα A:
  - –Περσινή μέση τιμή = \$50
  - –Τυπική απόκλιση = \$5

$$CV_A = \left(\frac{S}{\overline{X}}\right) \cdot 100\% = \frac{\$5}{\$50} \cdot 100\% = 10\%$$

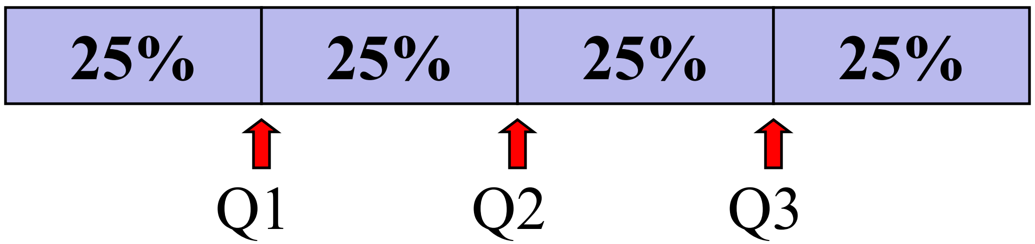
- Απόθεμα C:
  - –Περσινή μέση τιμή = \$8
  - –Τυπική απόκλιση = \$2

$$CV_C = \left(\frac{S}{\overline{X}}\right) \cdot 100\% = \frac{\$2}{\$8} \cdot 100\% = 25\%$$

Το Απόθεμα C έχει πολύ μικρότερη τυπική απόκλιση αλλά πολύ μεγαλύτερο συντελεστή μεταβλητότητας

Τεταρτημόρια Μέτρων

- Τα τεταρτημόρια χωρίζουν τα ταξινομημένα δεδομένα σε 4 τμήματα με ίσο αριθμό τιμών ανά τμήμα



- Το πρώτο τεταρτημόριο, Q_1, είναι η τιμή για την οποία το 25% των παρατηρήσεων είναι μικρότερες και το 75% μεγαλύτερο
- Το Q_2 είναι το ίδιο όπως η διάμεσος (Το 50% των παρατηρήσεων είναι μικρότερο και το 50% είναι μεγαλύτερο)
- Μόνο το 25% των παρατηρήσεων είναι μεγαλύτερο από το τρίτο τεταρτημόριο

Τεταρτημόρια Μέτρων: Εντοπισμός Τεταρτημορίων

Βρείτε ένα τεταρτημόριο καθορίζοντας την τιμή στην κατάλληλη θέση στα ταξινομημένα δεδομένα, όπου

Πρώτη θέση τεταρτημορίου: $Q1 = (n+1)/4$ ταξινομημένη τιμή

Δεύτερη θέση τεταρτημορίου: $Q2 = (n+1)/2$ ταξινομημένη τιμή

Τρίτη θέση τεταρτημορίου: $Q3 = 3(n+1)/4$ ταξινομημένη τιμή

όπου n είναι οι παρατηρούμενες τιμές

Τεταρτημόρια Μέτρων: Κανόνες υπολογισμού

- Κατά τον υπολογισμό της θέσης ταξινόμισης χρησιμοποιήστε τους ακόλουθους κανόνες

  - Αν το αποτέλεσμα είναι ένας ακέραιος αριθμός, τότε είναι η θέση ταξινόμησης που θα χρησιμοποιηθεί

  - Εάν το αποτέλεσμα είναι ένα κλασματικό ήμισυ (π.χ. 2.5, 7.5, 8.5, κ.τ.λ.) τότε βρείτε τον μέσο όρο των δύο αντίστοιχων τιμών δεδομένων.

  - Αν το αποτέλεσμα δεν είναι ένας ακέραιος αριθμός ή ένα κλασματικό ήμισυ, τότε στρογγυλοποιήστε το αποτέλεσμα στον πλησιέστερο ακέραιο για να βρείτε την ταξινομημένη θέση.

Τεταρτημόρια Μέτρων: Εντοπισμός Τεταρτημορίων

Δείγμα δεδομένων σε σειρά διάταξης: 11 12 13 16 16 17 18 21 22

(n = 9)

Το Q_1 είναι στη (9+1)/4 = 2,5 θέση των ταξινομημένων δεδομένων έτσι χρησιμοποιήστε την τιμή στη μέση μεταξύ της 2ης και 3ης τιμής,

έτσι **Q_1 = 12,5**

Τα Q1 και Q3 είναι μέτρα μη κεντρικής τοποθεσίας
Το Q2 = διάμεσος, είναι μέτρο κεντρικής τάσης

Τεταρτημόρια Μέτρων

Υπολογισμός των τεταρτημορίων: Παράδειγμα

Δείγμα δεδομένων σε σειρά διάταξης: 11 12 13 16 16 17 18 21 22

(n = 9)

Το Q_1 είναι στη(9+1)/4 = 2,5 θέση των ταξινομημένων δεδομένων,

έτσι $\mathbf{Q_1 = (12+13)/2 = 12,5}$

Το Q2 είναι στη(9+1)/2 = 5 θέση των ταξινομημένων δεδομένων,

έτσι $\mathbf{Q_2 = median = 16}$

Q_3 is in the 3(9+1)/4 = 7,5 θέση των ταξινομημένων δεδομένων,

έτσι $\mathbf{Q_3 = (18+21)/2 = 19,5}$

Τα Q_1 και Q_3 είναι μέτρα μη κεντρικής τοποθεσίας
Το Q_2 = διάμεσος, είναι μέτρο κεντρικής τάσης

Τεταρτημόρια Μέτρων:
Το Ενδοτεταρτομοριακό Εύρος (IQR)

- Το ενδοτεταρτομορικό εύρος είναι $Q_3 - Q_1$ και υπολογίζει την εξάπλωση στο μέσο του 50% των δεδομένων

- Το ενδοτεταρτομορικό εύρος ονομάζεται επίσης middleadspace επειδή καλύπτει το μεσαίο 50% των δεδομένων

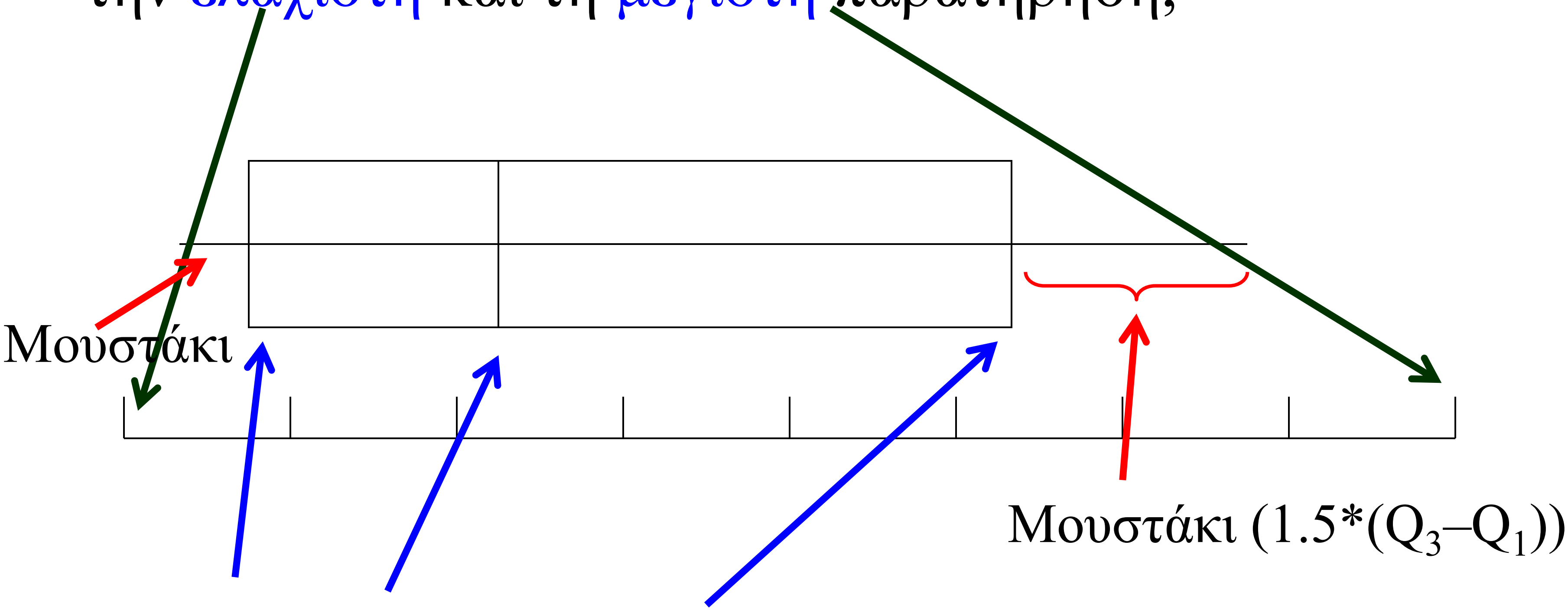
- Το ενδοτεταρτομορικό εύρος είναι ένα μέτρο μεταβλητότητας που δεν επηρεάζεται από υπερβολικές τιμές ή ακραίες τιμές

- Μέτρα όπως το Q1, Q3 και το ενδοτεταρτομορικό εύρος που δεν επηρεάζονται από ακραίες τιμές, ονομάζονται ανθεκτικά μέτρα

Θηκόγραμμα ...

… είναι μια τεχνική που απεικονίζει **πέντε** δείκτες:

- την ελάχιστη και τη μέγιστη παρατήρηση,

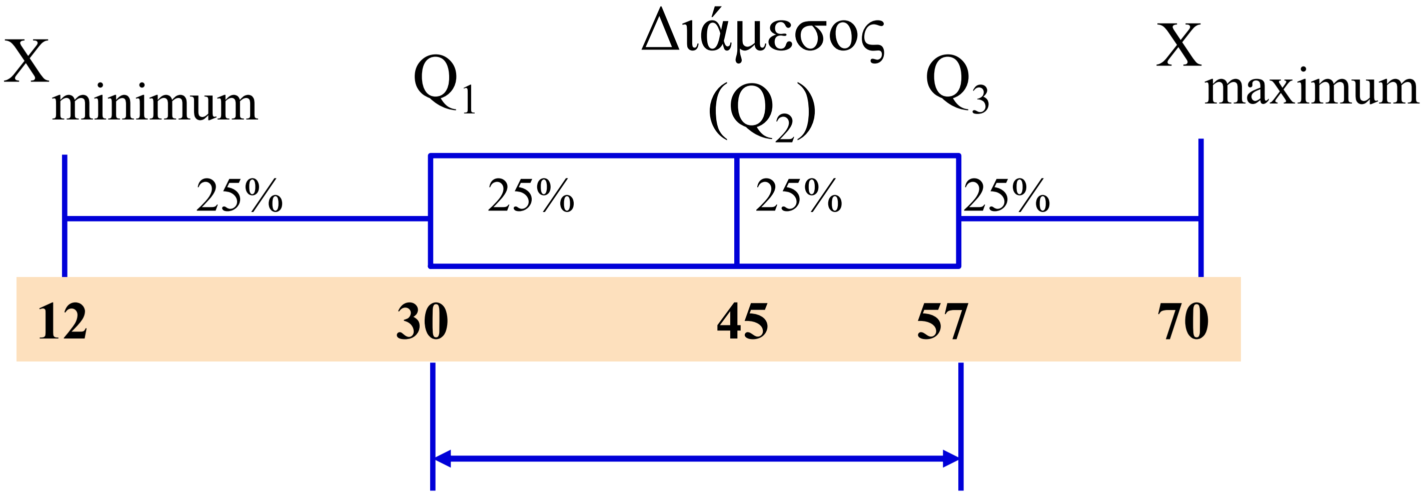


- πρώτο, δεύτερο, και τρίτο τεταρτημόριο.

Οι γραμμές που εκτείνονται αριστερά και δεξιά καλούνται μουστάκια. Οι τιμές που βρίσκονται έξω από τα μουστάκια καλούνται ακραίες τιμές. Τα μουστάκια εκτείνονται μέχρι 1.5 φορά το διατεταρτημοριακό εύρος ή μέχρι την μικρότερη/μεγαλύτερη τιμή που δεν είναι όμως ακραία τιμή.

Υπολογισμός του ενδοτεταρτημοριακού εύρους

Παράδειγμα:



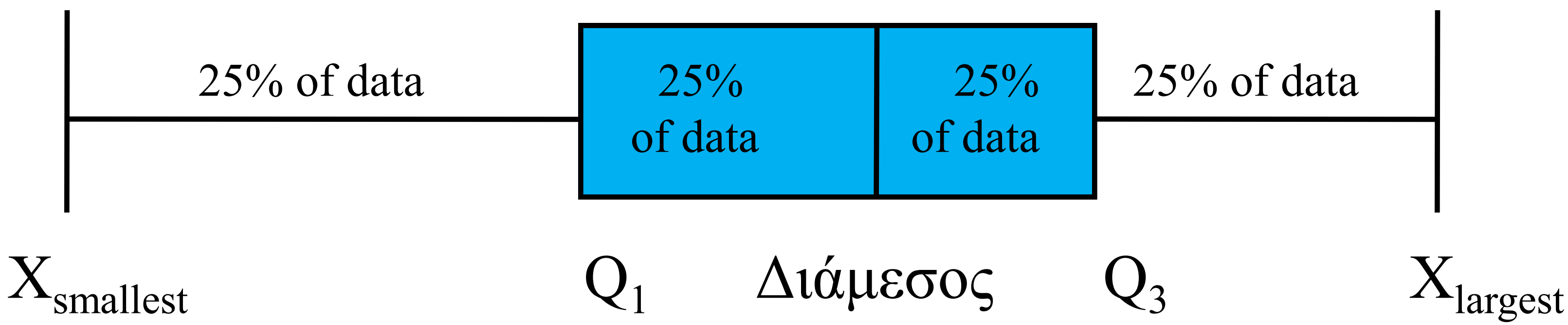
Ενδοτεταρτημοριακό εύρος
$= 57 – 30 = 27$

Περίληψη των πέντε αριθμών και θηκόγραμμα

- Το Θηκόγραμμα: Γραφική απεικόνιση των δεδομένων βάσει της περίληψης πέντε αριθμών:

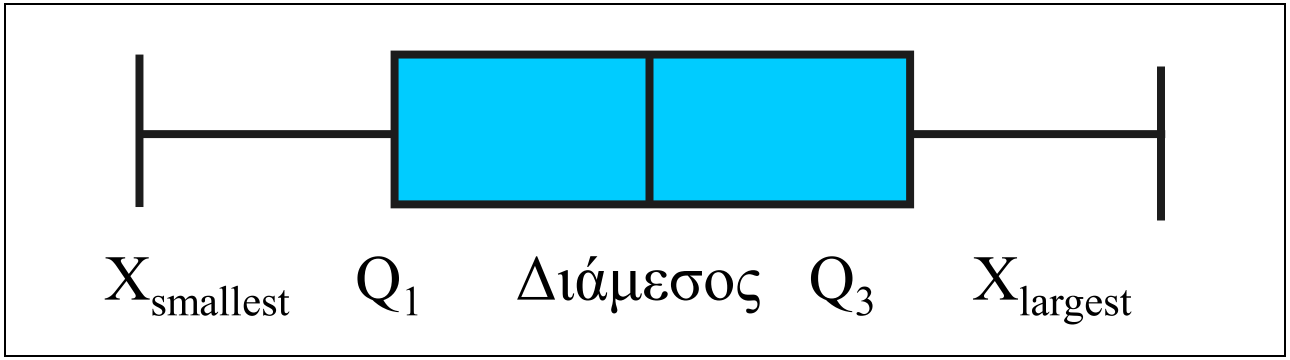
$X_{smallest}$ -- Q_1 -- Διάμεσος -- Q_3 -- $X_{largest}$

Παράδειγμα:



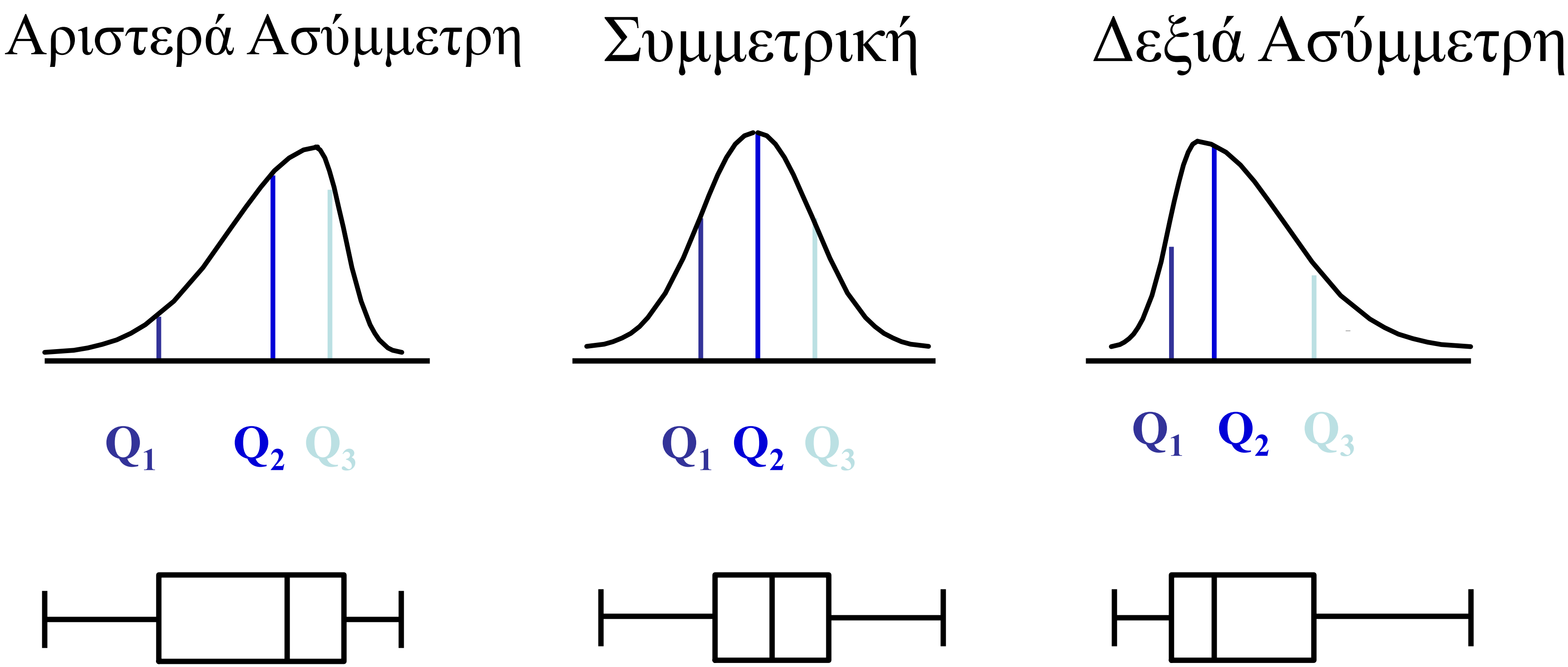
Περίληψη των πέντε αριθμών : Σχήμα των θηκογραμμάτων

- Εάν τα δεδομένα είναι συμμετρικά γύρω από την διάμεσο, τότε το πλαίσιο και η κεντρική γραμμή είναι κεντραρισμένα μεταξύ των τελικών σημείων



- Ένα θηκόγραμμα μπορεί να παρουσιαστεί σε κάθετο ή οριζόντιο προσανατολισμό

Κατανομή σχήματος και το θηκόγραμμα



Παράδειγμα 4.15

Ένας μεγάλος αριθμός εστιατορίων ταχείας εξυπηρέτησης προσφέρει στους οδηγούς και τους επιβάτες οχημάτων το πλεονέκτημα της γρήγορης εξυπηρέτησης. Για να μετρηθεί η ποιότητα της εξυπηρέτησης οργανώθηκε μια έρευνα η οποία κατέγραψε το χρόνο εξυπηρέτησης ενός δείγματος πελατών σε κάθε ένα από πέντε διαφορετικά εστιατόρια.

Να συγκρίνετε τα πέντε σύνολα δεδομένων χρησιμοποιώντας θηκογράμματα και να ερμηνεύσετε τα αποτελέσματα.

Θηκογράμματα …

Τα θηκογράμματα αυτά βασίζονται στα δεδομένα Xm04-15.

Οι χρόνοι εξυπηρέτησης στο εστιατόριο Wendy's

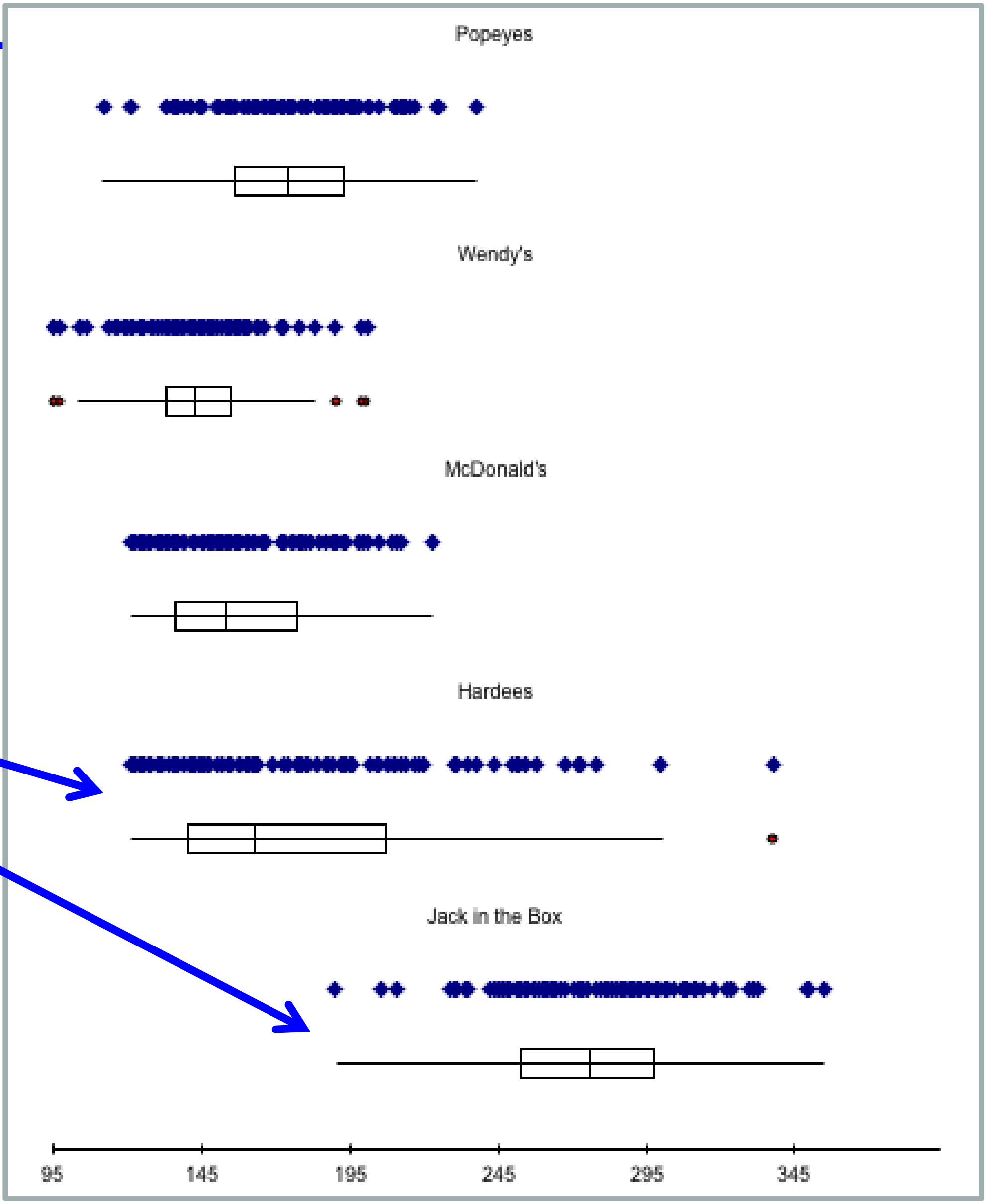
είναι οι μικρότεροι και οι λιγότερο μεταβλητοί

Το εστιατόριο Hardee's

έχει τη μεγαλύτερη διασπορά,

ενώ το εστιατόριο Jack-in-the-Box

έχει τους μεγαλύτερους χρόνους εξυπηρέτησης.



Συνήθη Εκατοστημόρια ...

Πρώτο (κάτω) τεταρτημόριο, Q_1,= 25 ^{o} εκατοστημόριο

Δεύτερο (μεσαίο) τεταρτημόριο, Q_2,= 50 ^{o} εκατοστημόριο

Τρίτο τεταρτημόριο, Q_3, = 75 ^{o} εκατοστημόριο

Πρώτο (κάτω) δεκατημόριο D_1 = 10^o εκατοστημόριο

Ένατο (άνω) δεκατημόριο D_9 = 90 ^{o} εκατοστημόριο

Σημείωση: *Εάν ο βαθμός σας τοποθετείται στο 80 ^{o} εκατοστημόριο, αυτό δεν σημαίνει ότι γράψατε 80% στις εξετάσεις – σημαίνει ότι το 80% των φοιτητών έγραψε* ***χαμηλότερα*** *από εσάς. Μιλάμε για τη θέση σας σε σχέση με τους άλλους.*

Θέση Εκατοστημορίου …

Ο παρακάτω τύπος επιτρέπει τον κατά προσέγγιση υπολογισμό της **θέσης** ενός εκατοστημορίου:

$$L_P = (n+1)\frac{P}{100}$$

όπου L_P είναι η **θέση** του P εκατοστημορίου

Θέση Εκατοστημορίου ...

Τα δεδομένα από το Παράδειγμα 4.1:

0 0 5 7 8 9 12 14 22 33

Ποια είναι η θέση του 25^{ου} εκατοστημορίου; Δηλαδή, σε ποιο σημείο το 25% των τιμών είναι κάτω από αυτό και το 75% των τιμών είναι υψηλότερο;

0 0 5 7 8 9 12 14 22 33

$$L_{25} = (10+1)(25/100) = \boxed{2.75}$$

Το 25^{ο} εκατοστημόριο είναι στα τρία τέταρτα της απόστασης ανάμεσα στη δεύτερη (η οποία είναι 0) και την τρίτη (η οποία είναι 5) παρατήρηση. Τρία τέταρτα της απόστασης είναι: (.75)(5 – 0) = 3.75

Επειδή η δεύτερη τιμή είναι 0, το 25^{ο} εκατοστημόριο είναι 0 + 3.75 = **3.75**

Θέση Εκατοστημορίου ...

Τι συμβαίνει με το άνω τεταρτημόριο;

$$L_{75} = (10+1)(75/100) = \boxed{8.25}$$

0 0 5 7 8 9 12 14 22 33

Βρίσκεται στο ένα τέταρτο της απόστασης μεταξύ της όγδοης και της ένατης παρατήρησης, οι οποίες είναι 14 και 22, αντίστοιχα. Ένα τέταρτο της απόστασης είναι: (.25)(22 - 14) = 2, που σημαίνει ότι το 75^{ο} εκατοστημόριο είναι: 14 + 2 = **16**

Θέση Εκατοστημορίου …

Θυμηθείτε …

θέση 2.75

16

0 0 | 5 7 8 9 12 14 | 22 33

3.75

θέση 8.25

Το L_p καθορίζει τη **θέση** στην οποία βρίσκεται η τιμή του εκατοστημορίου στο σύνολο των δεδομένων, και όχι την τιμή του εκατοστημορίου.

Δείκτες Γραμμικής Συσχέτισης …

Παρουσιάζουμε τρεις αριθμητικούς δείκτες γραμμικής συσχέτισης οι οποίοι παρέχουν πληροφορίες για **την ισχύ & την κατεύθυνση** μιας γραμμικής σχέσης μεταξύ δύο μεταβλητών (εάν αυτή υπάρχει).

Είναι:

η ***συνδιασπορά,***

ο ***συντελεστής συσχέτισης,*** και ο

συντελεστής προσδιορισμού.

Συνδιασπορά ...

Μέσος πληθυσμού για την μεταβλητή X, και για την Y

$$\text{Συνδιασπορά πληθυσμού} = \sigma_{xy} = \frac{\sum_{i=1}^{N} (\mathrm{x}_{\mathrm{i}} - \mu_x)(y_i - \mu_y)}{\mathrm{N}}$$

Μέσος δείγματος για την μεταβλητή X, και για την Y

$$\text{Συνδιασπορά δείγματος} = \mathrm{s}_{xy} = \frac{\sum_{i=1}^{n} (\mathrm{x}_{\mathrm{i}} - \overline{x})(y_i - \overline{y})}{\mathrm{n - 1}}$$

Σημείωση: ο διαιρέτης είναι n-1, και όχι n όπως θα περιμένατε.

Συνδιασπορά ...

Όπως υπάρχει ένας “τύπος” για τον υπολογισμό της διασποράς δείγματος ο οποίος δεν χρησιμοποιεί τον μέσο του δείγματος, έτσι υπάρχει και τύπος για τον υπολογισμό της συνδιασποράς του δείγματος χωρίς να απαιτείται ο υπολογισμός των μέσων:

$$s_{xy} = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}\right]$$

Συνδιασπορά …

Θεωρούμε τα ακόλουθα τρία σύνολα δεδομένων

	X	Y	$(X-\bar{X})$	$(Y-\bar{Y})$	$(X-\bar{X})(Y-\bar{Y})$	covariance
	2	13	-3	-7	21	
Set #1	6	20	1	0	0	$S_{xy} = 17.5$
	7	27	2	7	14	
	2	27	-3	7	-21	
Set #2	6	20	1	0	0	$S_{xy} = -17.5$
	7	13	2	-7	-14	
	2	20	-3	0	0	
Set #3	6	27	1	7	7	$S_{xy} = -3.5$
	7	13	2	-7	-14	

For each set: $\bar{X} = 5$ $\bar{Y}=20$

Σε κάθε σύνολο, οι τιμές της X είναι ίδιες, και οι τιμές της Y είναι ίδιες. Διαφοροποιείται μόνο η σειρά των τιμών της Y.

Στο σύνολο #1, καθώς η X αυξάνει, το ίδιο κάνει και η Y. S_{xy} μεγάλη & θετική

Στο σύνολο #2, καθώς η X αυξάνει, η Y μειώνεται. S_{xy} μεγάλη & αρνητική

Στο σύνολο #3, καθώς η X αυξάνει, η Y δεν κινείται με συγκεκριμένο τρόπο.

S_{xy} είναι “μικρή”

Συνδιασπορά … (Γενικά)

Όταν δύο μεταβλητές κινούνται στην ***ίδια κατεύθυνση*** (και οι δύο αυξάνουν ή και οι δύο μειώνονται), η συνδιασπορά θα είναι ένας ***μεγάλος θετικός αριθμός***.

Όταν δύο μεταβλητές κινούνται σε ***αντίθετη κατεύθυνση***, η συνδιασπορά θα είναι ένας ***μεγάλος αρνητικός αριθμός***.

Όταν ***δεν υπάρχει συγκεκριμένο μοτίβο***, η συνδιασπορά είναι ***μικρός αριθμός***.

Συχνά είναι δύσκολο να καθορίσουμε εάν μια συγκεκριμένη συνδιασπορά είναι μεγάλη ή μικρή. Το επόμενο στατιστικό μέτρο λύνει αυτό το πρόβλημα.

Συντελεστής Συσχέτισης ...

Ο συντελεστής συσχέτισης ορίζεται ως το πηλίκο της συνδιασποράς δια τις τυπικές αποκλίσεις των μεταβλητών:

Συντελεστής συσχέτισης πληθυσμού : $\rho = \dfrac{\sigma_{xy}}{\sigma_x \sigma_y}$

ελληνικό γράμμα "ρο"

Συντελεστής συσχέτισης δείγματος : $r = \dfrac{s_{xy}}{s_x s_y}$

Αυτός ο συντελεστής απαντά στο ερώτημα:
Πόσο **ισχυρή** είναι η σχέση μεταξύ των X και Y ;

Συντελεστής Συσχέτισης …

Το πλεονέκτημα του συντελεστή συσχέτισης ως προς τη συνδιασπορά είναι ότι έχει καθορισμένο εύρος από -1 έως +1, επομένως:

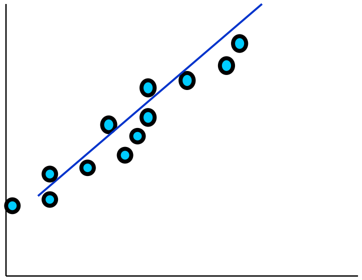
Εάν δύο μεταβλητές είναι ισχυρά θετικά συσχετισμένες, τότε η τιμή του συντελεστή είναι κοντά στο +1 (ισχυρή θετική γραμμική συσχέτιση).

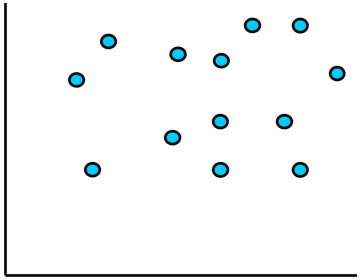
Εάν δύο μεταβλητές είναι ισχυρά αρνητικά συσχετισμένες, τότε η τιμή του συντελεστή είναι κοντά στο -1 (ισχυρή αρνητική γραμμική συσχέτιση).

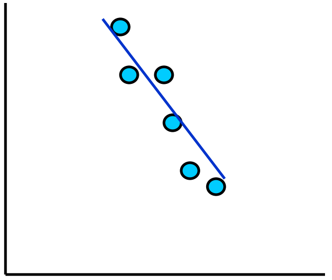
Καμία γραμμική συσχέτιση δεν διαφαίνεται όταν ο συντελεστής είναι κοντά στο 0.

Συντελεστής Συσχέτισης ...

$$\rho \text{ ή } r = \begin{cases} +1 & \text{Ισχυρή θετική γραμμική συσχέτιση} \\ 0 & \text{Καμία γραμμική σχέση} \\ -1 & \text{Ισχυρή αρνητική γραμμική συσχέτιση} \end{cases}$$







Παράδειγμα 4.16

Να υπολογιστεί ο συντελεστής συσχέτισης για τα προηγούμενα τρία σύνολα δεδομένων.

Παράδειγμα 4.16

Επειδή έχουμε ήδη υπολογίσει τις συνδιασπορές, χρειαζόμαστε μόνο τις τυπικές αποκλίσεις των X και Y.

$$\bar{x} = \frac{2+6+7}{3} = 5.0$$

$$\bar{y} = \frac{13+20+27}{3} = 20.0$$

$$s_x^2 = \frac{(2-5)^2 + (6-5)^2 + (7-5)^2}{3-1} = \frac{9+1+4}{2} = 7.0$$

$$s_y^2 = \frac{(13-20)^2 + (20-20)^2 + (27-20)^2}{3-1} = \frac{49+01+49}{2} = 49.0$$

Παράδειγμα 4.16

Οι τυπικές αποκλίσεις είναι

$$s_x = \sqrt{7.0} = 2.65$$

$$s_y = \sqrt{49.0} = 7.00$$

Παράδειγμα 4.16

Οι συντελεστές συσχέτισης είναι

$$\text{Set 1:} \quad r = \frac{s_{xy}}{s_x s_y} = \frac{17.5}{(2.65)(7.0)} = .943$$

$$\text{Set 2:} \quad r = \frac{s_{xy}}{s_x s_y} = \frac{-17.5}{(2.65)(7.0)} = -.943$$

$$\text{Set 3:} \quad r = \frac{s_{xy}}{s_x s_y} = \frac{-3.5}{(2.65)(7.0)} = -.189$$

Μέθοδος Ελαχίστων Τετραγώνων

Ο στόχος του διαγράμματος διασποράς είναι η μέτρηση της ισχύος και της κατεύθυνσης της γραμμικής συσχέτισης.

Και τα δύο μπορούν πιο εύκολα να εξαχθούν σχεδιάζοντας μια ευθεία γραμμή μέσα στα δεδομένα.

Χρειαζόμαστε μια αντικειμενική μέθοδο για τη δημιουργία αυτής της ευθείας.

Μία τέτοια μέθοδος είναι η **μέθοδος ελαχίστων τετραγώνων**.

Μέθοδος Ελαχίστων Τετραγώνων

Θυμίζουμε ότι, η εξίσωση μιας ευθείας με γνωστή κλίση δίνεται από τον τύπο:

$$y = mx + b$$

όπου:

m είναι η κλίση της ευθείας

b είναι το σημείο τομής με τον άξονα y.

Εάν γνωρίζουμε ότι υπάρχει γραμμική σχέση μεταξύ δύο μεταβλητών με γνωστή συνδιασπορά και γνωστό συντελεστή συσχέτισης, μπορούμε να καθορίσουμε τη γραμμική συνάρτηση της σχέσης;

Η Μέθοδος Ελαχίστων Τετραγώνων …

…δημιουργεί μια ευθεία γραμμή ανάμεσα στα σημεία ώστε το άθροισμα των τετραγώνων των αποκλίσεων μεταξύ των σημείων και της ευθείας να ελαχιστοποιείται. Η ευθεία ορίζεται από την εξίσωση:

$$\hat{y} = b_0 + b_1 x$$

b_0 (“b” μηδέν) είναι το σημείο τομής με τον άξονα y,

b_1 είναι η κλίση, και

$\hat{y}$ (“y” καπέλο) είναι η τιμή του y που καθορίζει η ευθεία.

Μέθοδος Ελαχίστων Τετραγώνων

Οι συντελεστές b_0 και b_1 δίνονται από:

$$b_1 = \frac{s_{xy}}{s_x^2}$$

$$b_0 = \bar{y} - b_1\bar{x}$$

$$\hat{y} = b_0 + b_1 x$$

Σταθερό και Μεταβλητό Κόστος

Το σταθερό κόστος είναι το κόστος που πρέπει να πληρωθεί ανεξάρτητα από τον κύκλο δραστηριοτήτων.

Το κόστος αυτή είναι “σταθερό” για μια συγκεκριμένη χρονική περίοδο ή για συγκεκριμένο εύρος παραγωγής.

Το μεταβλητό κόστος είναι αυτό που μεταβάλλεται σε σχέση με το πλήθος των παραγόμενων προϊόντων.

Σταθερό και Μεταβλητό Κόστος

Υπάρχουν ωστόσο και μεικτά έξοδα.

Υπάρχουν αρκετοί τρόποι να διαχωρίσουμε το μεικτό κόστος στη σταθερή και στη μεταβλητή συνιστώσα του. Μια τέτοια μέθοδος είναι η ευθεία ελαχίστων τετραγώνων. Δηλαδή, εκφράζουμε το συνολικό κόστος ως

$$y = b_0 + b_1 x$$

όπου y = συνολικό μεικτό κόστος, b_0 = σταθερό κόστος και b_1 = μεταβλητό κόστος, και x το πλήθος των παραγόμενων μονάδων

Παράδειγμα 4.17

Ένα μικρό μηχανουργείο κατασκευάζει εργαλεία.

Σκέφτεται να επεκτείνει τις δραστηριότητές του και χρειάζεται ανάλυση του κόστους παραγωγής.

Μια πηγή κόστους είναι το ηλεκτρικό, το οποίο απαιτείται για τη λειτουργία των μηχανών και το φωτισμό. (Μερικές εργασίες απαιτούν ιδιαίτερα ισχυρό φωτισμό.)

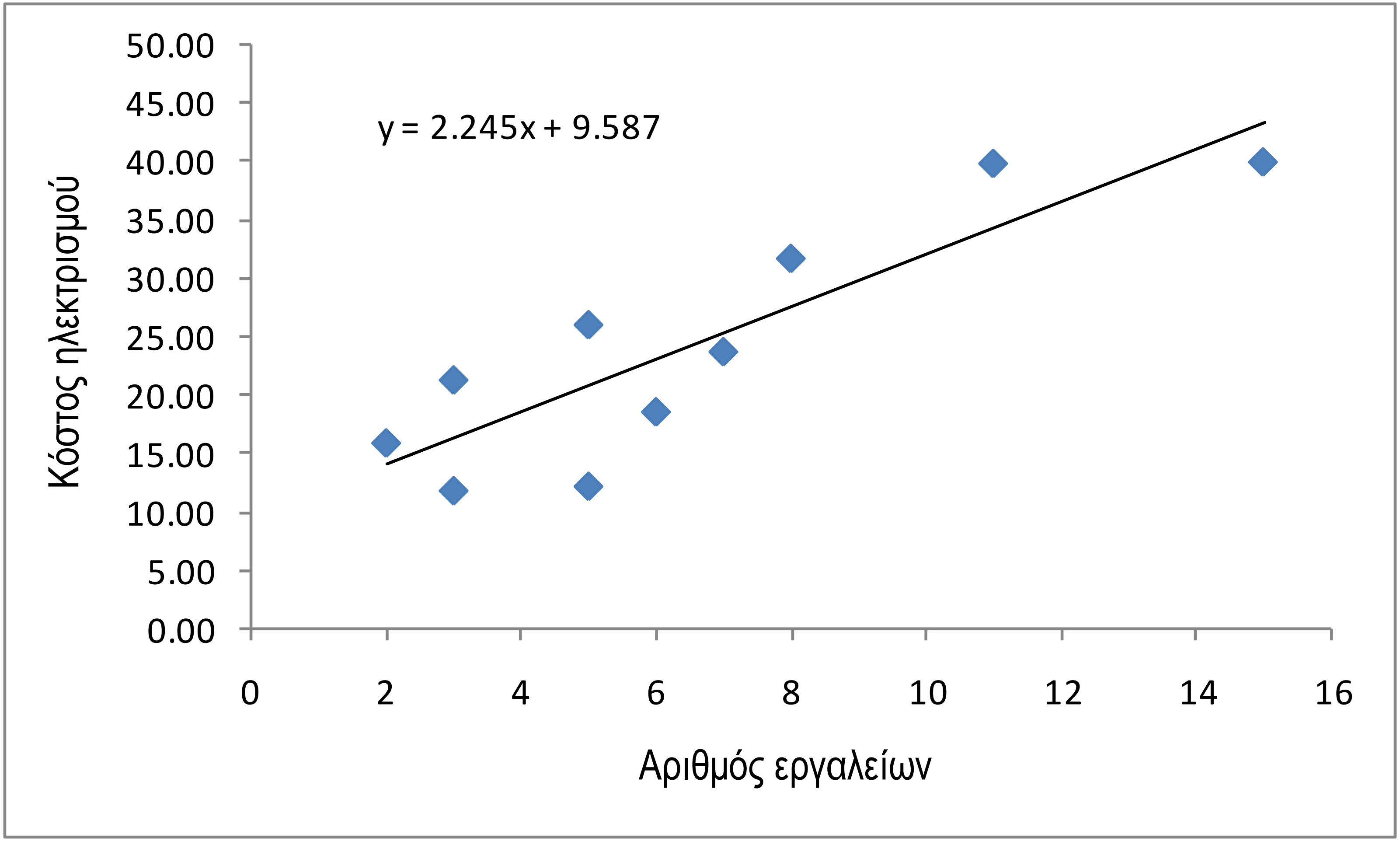
Έχει καταγράψει το καθημερινό κόστος ηλεκτρισμού και τον αριθμό των εργαλείων που παράγει. Να καθορίσετε το σταθερό και το μεταβλητό κόστος ηλεκτρισμού. [Xm04-17]

Παράδειγμα 4.17

Λαμβάνει ένα δείγμα 10 ημερών και καταγράφει για κάθε μέρα τον αριθμό παραγόμενων εργαλείων και το σχετικό κόστος παραγωγής.

Ημέρα	Αριθμός Εργαλείων	Κόστος Ηλεκτρισμού
1	7	23,80
2	3	11,89
3	2	15,98
4	5	26,11
5	8	31,79
6	11	39,93
7	5	12,27
8	15	40,06
9	3	21,38
10	6	18,65

Παράδειγμα 4.17



Παράδειγμα 4.17

$$\hat{y} = 9.587 + 2.245x$$

Η κλίση δείχνει μεταβολή/μονάδα, που σημαίνει ότι είναι η μεταβολή του y (αύξηση) για κάθε 1-μονάδα αύξησης του x .

Η κλίση μετράει την *οριακή* μεταβολή της εξαρτημένης μεταβλητής. Η οριακή μεταβολή αναφέρεται στο αποτέλεσμα της αύξησης της ανεξάρτητης μεταβλητής κατά μία επιπλέον μονάδα.

Στο παράδειγμα αυτό η κλίση είναι 2.25, που σημαίνει ότι για κάθε 1-μονάδα αύξησης στον αριθμό των εργαλείων, η οριακή αύξηση στο κόστος ηλεκτρισμού είναι 2.25. Άρα, το εκτιμώμενο μεταβλητό κόστος είναι $2.25 ανά εργαλείο.

Παράδειγμα 4.17

$$\hat{y} = 9.59 + 2.25x$$

Το σημείο τομής με τον άξονα y είναι 9.59.
Αυτή είναι απλά η τιμή όταν x = 0.

Ωστόσο, όταν x = 0 δεν έχουμε παραγωγή εργαλείων και συνεπώς το εκτιμώμενο σταθερό κόστος ηλεκτρισμού είναι $9.59 ανά ημέρα.

Συντελεστής Προσδιορισμού

Όταν είδαμε τον συντελεστή συσχέτισης σημειώσαμε ότι εκτός από τις τιμές −1, 0, και +1, δεν μπορούμε να ερμηνεύσουμε τη σημασία του.

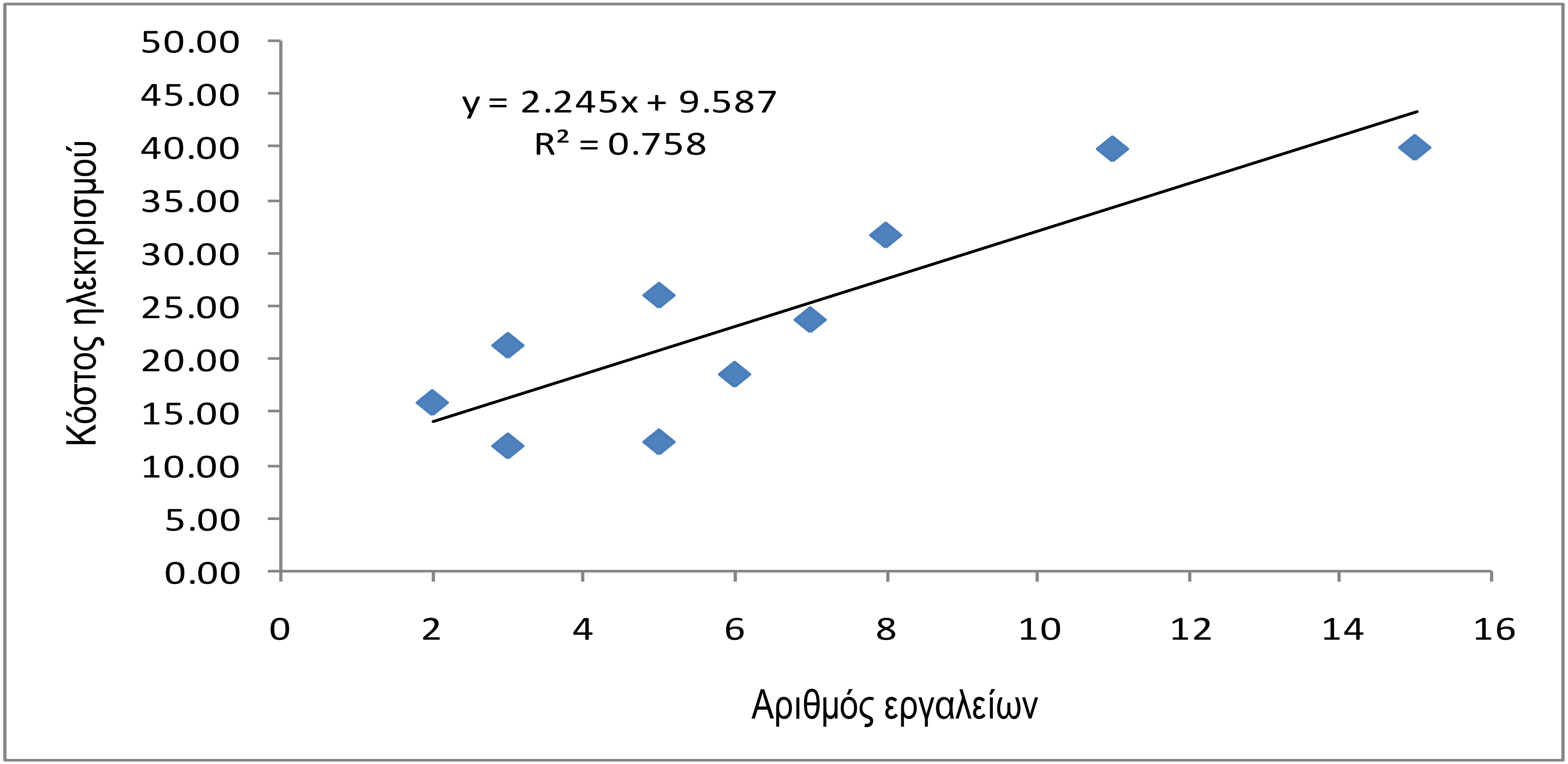
Μπορούμε να κρίνουμε τον συντελεστή συσχέτισης μόνο σε σχέση με το πόσο πλησιάζει στο −1, 0, και +1 .

Ευτυχώς, έχουμε άλλο ένα δείκτη που μπορεί να ερμηνευθεί καλύτερα. Είναι ο ***συντελεστής προσδιορισμού***, ο οποίος υπολογίζεται υψώνοντας στο τετράγωνο τον συντελεστή συσχέτισης. Γι’ αυτό συμβολίζεται R^2 .

Ο συντελεστής προσδιορισμού εκφράζει σε ποιο βαθμό η μεταβλητότητα της εξαρτημένης μεταβλητής εξηγείται από τη μεταβλητότητα της ανεξάρτητης.

Παράδειγμα 4.17

Υπολογίστε τον συντελεστή προσδιορισμού για το Παράδειγμα 4.17



Παράδειγμα 4.17

Ο συντελεστής προσδιορισμού είναι

$$R^2 = .758$$

Αυτό μας λέει ότι το 75.8% του κόστους ηλεκτρισμού εξηγείται από τον αριθμό των εργαλείων που κατασκευάζονται.

Το υπόλοιπο 24.2% οφείλεται σε άλλους λόγους.

Ερμηνεία της Συσχέτισης …

Λόγω της σημασίας της, θυμίζουμε τη σωστή ερμηνεία την ανάλυσης της σχέσης μεταξύ δύο συνεχών μεταβλητών.

Δηλαδή, εάν δύο μεταβλητές είναι γραμμικά συσχετισμένες αυτό δεν σημαίνει ότι η X προκαλεί τη μεταβολή της Y. Μπορεί να σημαίνει ότι μια άλλη μεταβλητή επηρεάζει και την X και την Y ή ότι η Y προκαλεί τη μεταβολή της X.

Θυμηθείτε

“Συσχέτιση δεν είναι αιτιολόγηση”

SPSS output, Παράδειγμα 4.17

Regression

Descriptive Statistics

	Mean	Std. Deviation	N
Ecost	24,1860	10,33103	10
Ntools	6,5000	4,00694	10

Correlations

		Ecost	Ntools
Pearson Correlation	Ecost	1,000	**,871**
	Ntools	,871	1,000
Sig. (1-tailed)	Ecost	.	**,001**
	Ntools	,001	.
N	Ecost	10	10
	Ntools	10	10

Model Summary[b]

Model	**R**	**R Square**	Adjusted R Square	Std. Error of the Estimate
1	**,871[a]**	**,759**	,729	5,38185

a. Predictors: (Constant), Ntools

b. Dependent Variable: Ecost

SPSS output, Παράδειγμα 4.17 (2)

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	728,856	1	728,856	25,164	,001[a]
	Residual	231,715	8	28,964		
	Total	960,571	9			

a. Predictors: (Constant), Ntools

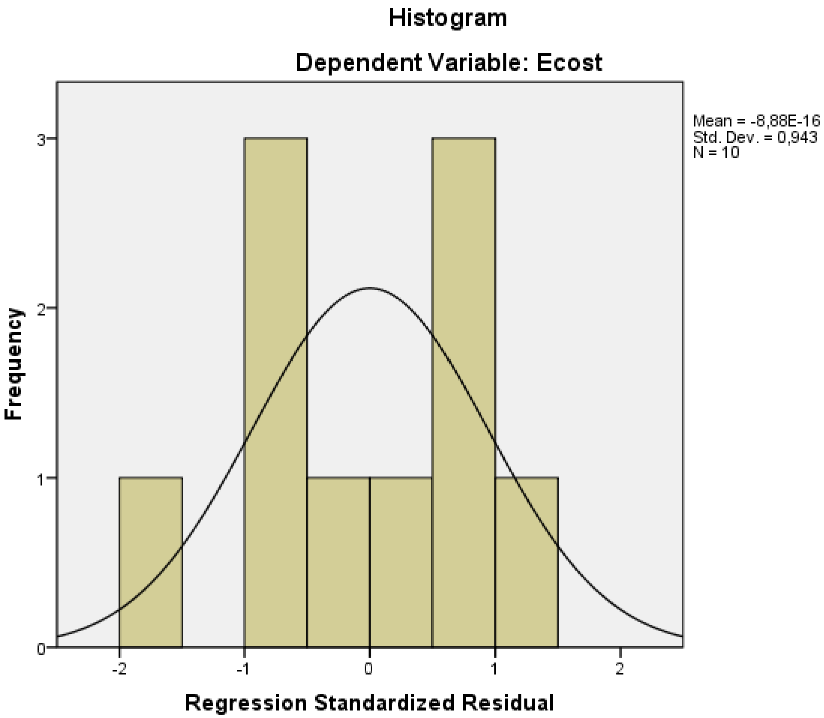
b. Dependent Variable: Ecost

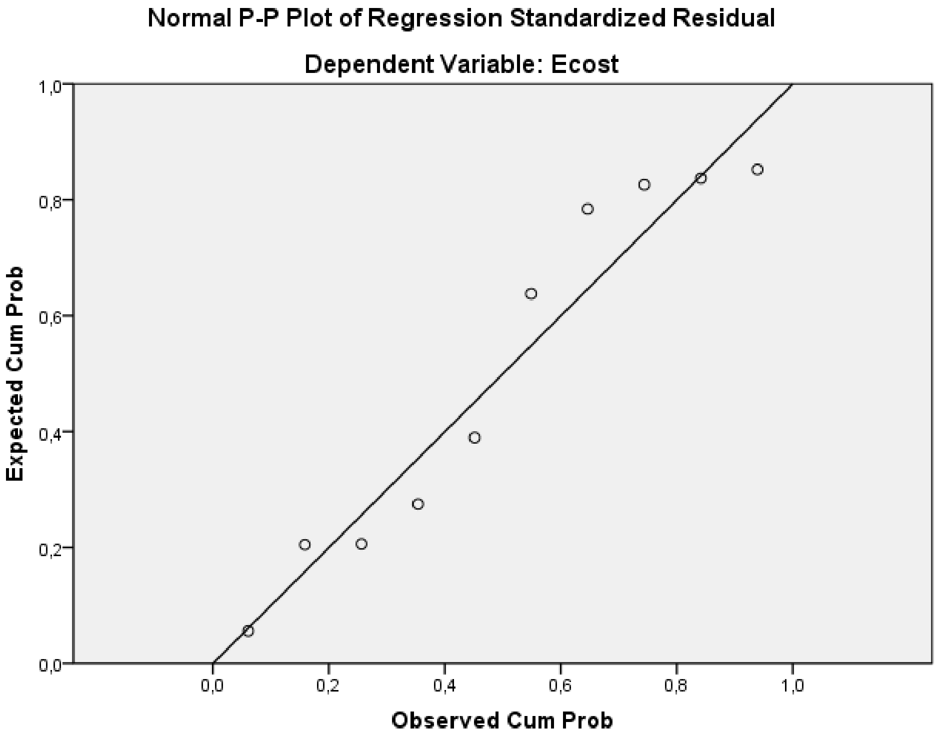
Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	**9,588**	3,371		2,844	**,022**	1,814	17,362
	Ntools	**2,246**	,448	,871	5,016	**,001**	1,213	3,278

a. Dependent Variable: Ecost

SPSS output, Παράδειγμα 4.17 (3)





Στατιστικά μέτρα

	Πληθυσμός	Δείγμα
Μέγεθος	N	n
Μέσος	μ	$\overline{x}$
Διασπορά	σ^2	s^2
Τυπική Απόκλιση	σ	s
Συντελεστής Μεταβλητότητας	CV	cv
Συνδιασπορά	σ_{xy}	S_{xy}
Συντελεστής Συσχέτισης	ρ_{xy}	r

Άσκηση

Ένα μικρό μηχανουργείο κατασκευάζει εργαλεία κατά παραγγελία. Προκειμένου να επεκτείνει τις δραστηριότητες του αναλύει το κόστος παραγωγής, και μια σημαντική πηγή κόστους είναι ο ηλεκτρισμός, τόσο για τη λειτουργία των μηχανών όσο και για τον ειδικό φωτισμό που απαιτείται. Για τον λόγο αυτό έχει καταγράψει τον αριθμό των εργαλείων που κατασκευάστηκαν και το κόστος του ηλεκτρικού ρεύματος ανά ημέρα, για ένα δείγμα 10 ημερών, ως εξής:

Ημέρα	1	2	3	4	5	6	7	8	9	10
Εργαλεία	7	3	2	5	8	11	5	15	3	6
Κόστος	23,80	11,89	15,98	26,11	31,79	39,93	12,27	40,06	21,38	18,65

Να αναλύσετε το κόστος σε σταθερό και μεταβλητό

Τα δεδομένα σε λειτουργική δομή

Ημέρα	Αριθμός Εργαλείων	Κόστος Ηλεκτρισμού
1	7	23,80
2	3	11,89
3	2	15,98
4	5	26,11
5	8	31,79
6	11	39,93
7	5	12,27
8	15	40,06
9	3	21,38
10	6	18,65

Επίλυση:

Επίλυση:

Η ανεξάρτητη μεταβλητή Χ είναι η ποσότητα των εργαλείων ανά ημέρα, και η εξαρτημένη μεταβλητή Υ είναι το ημερήσιο κόστος του ηλεκτρικού ρεύματος.

Οι συντελεστές της ευθείας των ελάχιστων τετραγώνων δίνονται από τους τύπους:

$$b_1 = \frac{s_{xy}}{s_x^2} \text{ και } b_0 = \bar{y} - b_1 \cdot \bar{x}$$

Όπου:

$$S_{xy} = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i \cdot Y_i - \frac{\sum_{i=1}^{n} X_i \cdot \sum_{i=1}^{n} Y_i}{n}\right]$$

$$S_X^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - \frac{(\sum_{i=1}^{n} X_i)^2}{n}\right]$$

$$\bar{X} = \frac{\sum_{i=1}^{10} x_i}{n}$$

$$\bar{Y} = \frac{\sum_{i=1}^{10} y_i}{n}$$

Σύμφωνα με τους παραπάνω τύπους, για τον υπολογισμό των συντελεστών $\mathbf{b_0}$ και $\mathbf{b_1}$ της ευθείας των ελάχιστων τετραγώνων απαιτούνται τα αθροίσματα των ποσοτήτων Χ, Υ, $X \cdot Y$, X^2 και Y^2, που σχηματίζονται στον παρακάτω πίνακα:

Υπολογισμοί

Ημέρα	X	Y	X·Y	X^2	Y^2
1	7	23,80	166,60	49	566,44
2	3	11,89	35,67	9	141,37
3	2	15,98	31,96	4	255,36
4	5	26,11	130,55	25	681,73
5	8	31,79	254,32	64	1010,60
6	11	39,93	439,23	121	1594,40
7	5	12,27	61,35	25	150,55
8	15	40,06	600,90	225	1604,80
9	3	21,38	64,14	9	457,10
10	6	18,65	111,90	36	347,82
Σ	**65**	**241,86**	**1896,62**	**567**	**6810,20**

Εφαρμογή τύπων

Με τη βοήθεια των παραπάνω αποτελεσμάτων βρίσκουμε:

$$\bar{X} = \frac{\sum_{i=1}^{10} x_i}{n} = \frac{65}{10} = 6,5$$

$$\bar{Y} = \frac{\sum_{i=1}^{10} y_i}{n} = \frac{241,86}{10} = 24,19$$

$$s_{xy} = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i \cdot Y_i - \frac{\sum_{i=1}^{n} X_i \cdot \sum_{i=1}^{n} Y_i}{n}\right] = \frac{1}{10-1}\left[1896,62 - \frac{65 \cdot 241,86}{10}\right] = 0,11 \cdot \left[1896,62 - \frac{15720,90}{10}\right]$$

$$= 0,11 \cdot [1896,62 - 1572,09] = 36,06$$

$$S_X^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - \frac{(\sum_{i=1}^{n} X_i)^2}{n}\right] = \frac{1}{10-1}\left[567 - \frac{(65)^2}{10}\right] = \frac{1}{9} \cdot \left[567 - \frac{4225}{10}\right] =$$

$$= \frac{1}{9} \cdot [567 - 422,5] = \frac{1}{9} \cdot 144,5 = 16,06$$

Οι συντελεστές της ευθείας των ελάχιστων τετραγώνων δίνονται από τους τύπους:

$$b_1 = \frac{s_{xy}}{s_x^2} = \frac{36,06}{16,06} = 2,25$$

και:

$$b_0 = \bar{y} - b_1 \cdot \bar{x} = 24,19 - 2,25 \cdot 6,5 = 24,19 - 14,625 = 9,57$$

Συνεπώς η εξίσωση της ευθείας των ελάχιστων τετραγώνων είναι:

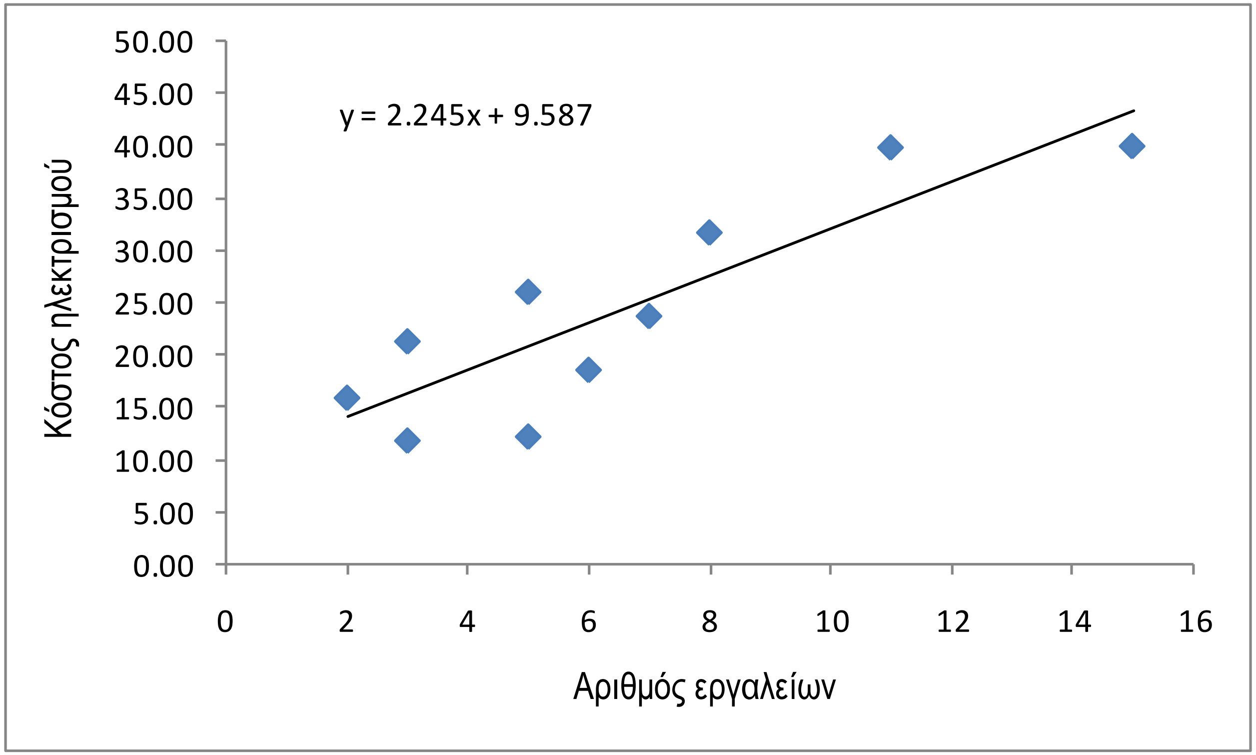
$y = 9,57 + 2,25 \cdot x$

Ερμηνεία των συντελεστών b_0 και b_1:

- Ο συντελεστής διεύθυνσης $b_1 = 2{,}25$ είναι η κλίση *(slope)* της ευθείας. Με απλά λόγια, η κλίση της ευθείας αντιπροσωπεύει την οριακή αύξηση της μεταβλητής Y ανά μονάδα της μεταβλητής X. Αυτό είναι **το μεταβλητό κόστος** της ηλεκτρικής κατανάλωσης, που ερμηνεύεται ως 2,25 ευρώ ανά εργαλείο.
- Ο συντελεστής $b_0 = 9{,}57$ αντιπροσωπεύει το σημείο τομής (*intercept*) της ευθείας με τον άξονα Y. Αυτό είναι **το σταθερό κόστος**, που ερμηνεύεται ως: η ημερήσια ηλεκτρική κατανάλωση αν δεν παραχθεί κανένα εργαλείο είναι 9,57.

Γραφική Απεικόνιση

Παράδειγμα 4.17



Συντελεστής συσχέτισης:

Για τον υπολογισμό του συντελεστή συσχέτισης απαιτούνται η συμμεταβλητότητα και οι τυπικές αποκλίσεις των δυο μεταβλητών.

$$r = \frac{S_{xy}}{S_x \cdot S_y}$$

Παραπάνω υπολογίσαμε ότι:

$S_{xy} = 36{,}06$

$S_x^2 = 16{,}06$,

δηλαδή: $S_{x=} \sqrt{16{,}06} = 4{,}01$

Μένει να υπολογιστεί η τυπική απόκλιση της μεταβλητής Y:

$$S_y^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n}\right] = \frac{1}{10-1}\left[6810{,}20 - \frac{(241{,}86)^2}{10}\right] =$$

$$\frac{1}{9}\cdot\left[6810{,}20 - \frac{58496{,}26}{10}\right] = \frac{1}{9}\cdot\left[6810{,}20 - \frac{58496{,}26}{10}\right] =$$

$$\frac{1}{9}\cdot[6810{,}20 - 5849{,}63] = \frac{1}{9}\cdot[960{,}56] = 106{,}73$$

Δηλαδή: $S_{y=} \sqrt{106{,}73} = 10{,}33$

Άρα ο συντελεστής συσχέτισης των δύο μεταβλητών είναι:

$$r = \frac{S_{xy}}{S_x \cdot S_y} = \frac{36{,}06}{4{,}01 \cdot 10{,}33} = 0{,}8705$$

Ερμηνεία:

Ο συντελεστής συσχέτισης r=0,8705 δείχνει μια ισχυρή θετική γραμμική συσχέτιση μεταξύ του αριθμού των παραγόμενων εργαλείων και της ημερήσιας κατανάλωσης ηλεκτρισμού. Αυτό σημαίνει ότι η ευθεία των ελάχιστων τετραγώνων ταιριάζει καλά στα δεδομένα, συνεπώς η παραπάνω ανάλυση του μεικτού κόστους σε σταθερό και μεταβλητό είναι αρκετά αξιόπιστη.

Συντελεστής προσδιορισμού:

$$R^2 = r^2 = 0{,}8705^2 = 0{,}7578$$

Αυτό σημαίνει ότι το 75,78% της ημερήσιας ηλεκτρικής κατανάλωσης εξηγείται από τον αριθμό των εργαλείων που κατασκευάζονται. Το υπόλοιπο 24,22% οφείλεται σε άλλες μεταβλητές, που προς το παρόν δεν γνωρίζουμε.

Παράδειγμα 4.17

Υπολογίστε τον συντελεστή προσδιορισμού για το Παράδειγμα 4.17