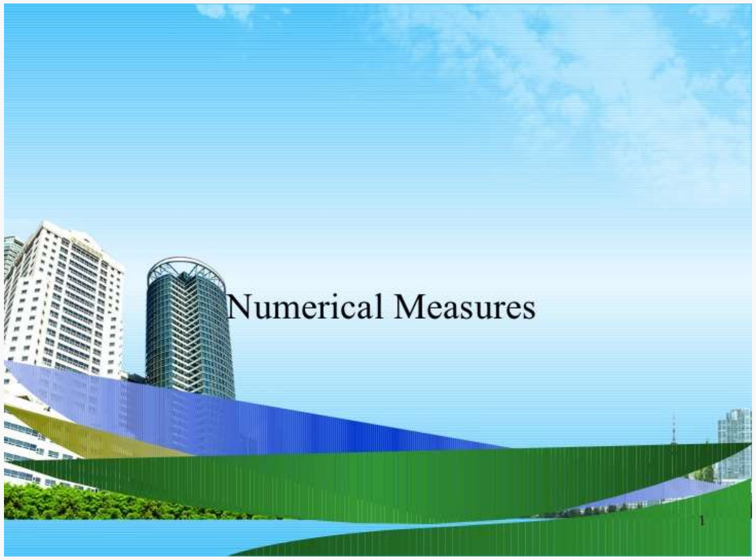
Διάλεξη 2

Περιγράφοντας τα δεδομένα:

Αριθμητικά Μέτρα



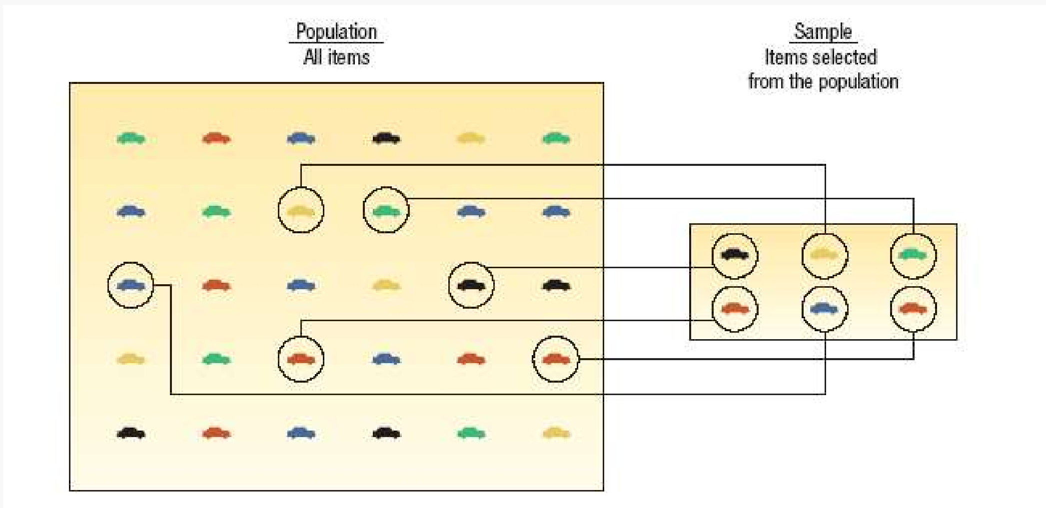
ΣΤΟΧΟΙ

- Υπολογίστε τον αριθμητικό μέσο, σταθμισμένο μέσο, διάμεσο, επικρατούσα τιμή και τον γεωμετρικό μέσο.
- Εξηγείστε τα χαρακτηριστικά, τις χρήσεις, τα πλεονεκτήματα και μειονεκτήματα κάθε μέτρου θέσης.
- Εντοπίστε τη θέση του μέσου, της διαμέσου και της επικρατούσας τιμής και για συμμετρικές και λοξές κατανομές.
- Υπολογίστε και ερμηνεύστε το εύρος, τη μέση απόκλιση, τη διακύμανση και την τυπική απόκλιση.
- Κατανοήστε τα χαρακτηριστικά, τις χρήσεις, τα πλεονεκτήματα και τα μειονεκτήματα κάθε μέτρου διασποράς.
- Κατανοήστε το θεώρημα Chebyshev's και τον Εμπειρικό Κανόνα καθώς σχετίζονται με ένα σύνολο παρατηρήσεων.

Παράμετρος vs. Στατιστική

ΠΑΡΑΜΕΤΡΟΣ Ένα μετρήσιμο χαρακτηριστικό ενός πληθυσμού

ΣΤΑΤΙΣΤΙΚΗ Ένα μετρήσιμο χαρακτηριστικό ενός δείγματος



Μέσος όρος Πληθυσμού

Σε μη ομαδοποιημένα δεδομένα, ο μέσος του πληθυσμού είναι το άθροισμα όλων των τιμών του πληθυσμού διά του πλήθους των τιμών του πληθυσμού. Ο μέσος του δείγματος είναι το άθροισμα όλως των τιμών του δείγματος διά του πλήθους των παρατηρήσεων του δείγματος.

POPULATION MEAN $\mu = \frac{\Sigma X}{N}$ **[3–1]**

Όπου μ ο πληθυσμιακός μέσος, X οι τιμές των στοιχείων του πληθυσμού, Σ η πράξη της άθροισης και N το μέγεθος του πληθυσμού.

SAMPLE MEAN $\bar{X} = \frac{\Sigma X}{n}$ **[3–2]**

Όπου $\bar{X}$ ο δειγματικός μέσος, X οι τιμές των στοιχείων του δείγματος, Σ η πράξη της άθροισης και n το μέγεθος του δείγματος.

ΠΑΡΑΔΕΙΓΜΑ:

SunCom is studying the number of minutes used monthly by clients in a particular cell phone rate plan. A random sample of 12 clients showed the following number of minutes used last month.

90	77	94	89	119	112
91	110	92	100	113	83

What is the arithmetic mean number of minutes used?

$$\text{Sample mean} = \frac{\text{Sum of all values in the sample}}{\text{Number of values in the sample}}$$

$$\bar{X} = \frac{\Sigma X}{n} = \frac{90 + 77 + \cdots + 83}{12} = \frac{1170}{12} = 97.5$$

Η Διάμεσος

ΔΙΑΜΕΣΟΣ: Το κεντρικό σημείο των τιμών αφού αυτές έχουν διαταχθεί από τη μικρότερη προς τη μεγαλύτερη.

ΙΔΙΟΤΗΤΕΣ ΤΗΣ ΔΙΑΜΕΣΟΥ

1. Υπάρχει μία μοναδική διάμεσος για κάθε σύνολο δεδομένων.
2. Δεν επηρεάζεται από τις υπερβολικά μεγάλες ή μικρές τιμές και επομένως αποτελεί ένα πολύτιμο μέτρο της κεντρικής τάσης, όταν προκύψουν τέτοιες τιμές.

ΠΑΡΑΔΕΙΓΜΑΤΑ:

Οι ηλικίες για ένα δείγμα πέντε φοιτητών είναι:

21, 25, 19, 20, 22

Διάταξη των δεδομένων σε αύξουσα σειρά:

19, 20, 21, 22, 25.

Άρα, η διάμεσος είναι 21.

Το ύψος τεσσάρων καλαθοσφαιριστών σε ίντσες, είναι:

76, 73, 80, 75

Διάταξη των δεδομένων σε αύξουσα σειρά:

73, 75, 76, 80.

Άρα, η διάμεσος είναι 75.5

Η Επικρατούσα Τιμή

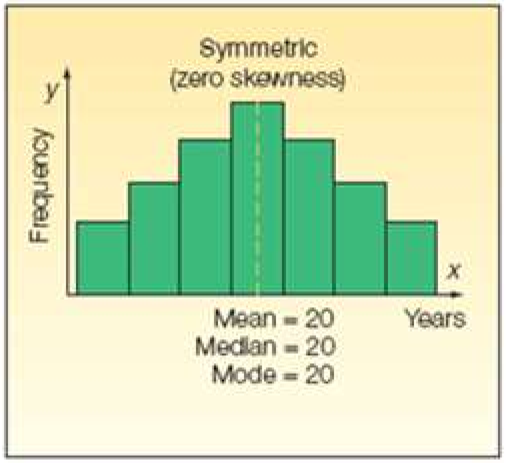
ΕΠΙΚΡΑΤΟΥΣΑ ΤΙΜΗ: Είναι η τιμή μεταξύ των παρατηρήσεων η οποία εμφανίζεται με την μεγαλύτερη συχνότητα.

The annual salaries of quality-control managers in selected states are shown below. What is the modal annual salary?

State	Salary	State	Salary	State	Salary
Arizona	$35,000	Illinois	$58,000	Ohio	$50,000
California	49,100	Louisiana	60,000	Tennessee	60,000
Colorado	60,000	Maryland	60,000	Texas	71,400
Florida	60,000	Massachusetts	40,000	West Virginia	60,000
Idaho	40,000	New Jersey	65,000	Wyoming	55,000

A perusal of the salaries reveals that the annual salary of $60,000 appears more often (six times) than any other salary. The mode is, therefore, $60,000.

Οι σχετικές θέσεις Μέσου, Διαμέσου και Επικρατούσας Τιμής

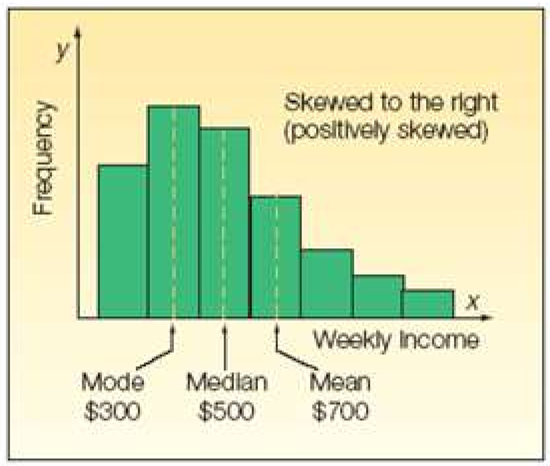


zero skewness

mode = median = mean

0 λοξότητα

Επ. τιμή= Διάμεσος=Μέσος



positive skewness

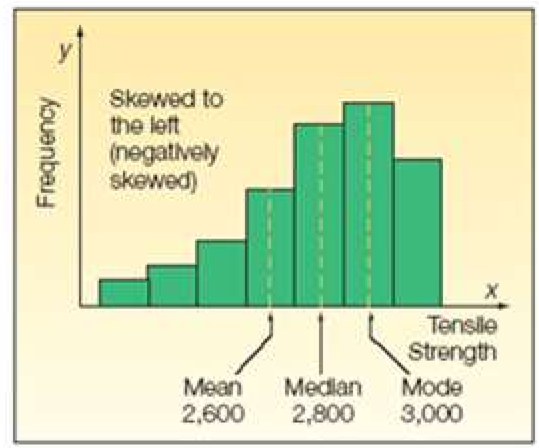
mode < median < mean

θετική λοξότητα

Επ. τιμή< Διάμεσος<Μέσος

αρνητική λοξότητα

Επ. τιμή> Διάμεσος>Μέσος



negative skewness

mode > median > mean

Ο Γεωμετρικός Μέσος

$$\text{GEOMETRIC MEAN} \qquad GM = \sqrt[n]{(X_1)(X_2)\cdots(X_n)} \qquad [3\text{–}4]$$

- Χρήσιμος για την εύρεση της μέσης μεταβολής ποσοστών, δεικτών ή ρυθμών ανάπτυξης στο χρόνο.
- Έχει ευρεία εφαρμογή στις επιχειρήσεις και στα οικονομικά καθώς πολύ συχνά θέλουμε να βρούμε τις ποσοστιαίες αλλαγές στις πωλήσεις, στις αποδοχές ή σε οικονομικά μεγέθη όπως το ΑΕΠ.
- Ο γεωμετρικός μέσος θα είναι πάντοτε μικρότερος ή ίσος από τον αριθμητικό μέσο.
- Ο τύπος για τον Γεωμετρικό Μέσο δίνεται παρακάτω:

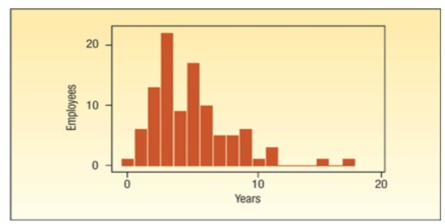
ΠΑΡΑΔΕΙΓΜΑ:

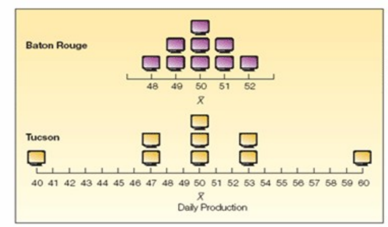
Υποθέστε ότι αυτόν τον χρόνο παίρνετε αύξηση 5% στις αποδοχές σας και 15% αύξηση τον επόμενο χρόνο. Η μέση ετήσια ποσοστιαία αύξηση είναι 9.886, και όχι 10.0. Γιατί όμως; Ας υπολογίσουμε τον Γεωμετρικό Μέσο.

$$GM = \sqrt{(1.05)(1.15)} = 1.09886$$

Μέτρα Διασποράς

- Μέτρα θέσης, όπως η μέση τιμή ή η διάμεσος, περιγραφούν μόνο την κεντρική τάση των δεδομένων. Αυτά τα μέτρα είναι πολύτιμης σημασίας από αυτή τη σκοπιά, αλλά δεν μας δίνουν κανένα στοιχείο σχετικά με τη διασπορά των δεδομένων.
- Για παράδειγμα, αν ο ξεναγός σας λέει ότι ο ποταμός μπροστά σας έχει κατά μέσο όρο βάθος 3 πόδια, θα θέλατε να διασχίσετε το ποτάμι χωρίς επιπρόσθετες πληροφορίες; Μάλλον όχι. Θα θέλατε να γνωρίζετε κάτι παραπάνω σχετικά με τη διακύμανση του βάθους.
- Ένας δεύτερος λόγος για τη μελέτη της διασποράς σε ένα σύνολο δεδομένων είναι για να την συγκρίνουμε σε δύο ή περισσότερες κατανομές.





- ΕΥΡΟΣ

RANGE Range = Largest value − Smallest value **[3–6]**

- ΜΕΣΗ ΑΠΟΚΛΙΣΗ

MEAN DEVIATION $$MD = \frac{\Sigma|X - \bar{X}|}{n}$$ **[3–7]**

- ΠΛΥΘΗΣΜΙΑΚΗ ΔΙΑΚΥΜΑΝΣΗ

POPULATION VARIANCE $$\sigma^2 = \frac{\Sigma(X - \mu)^2}{N}$$ **[3–8]**

- ΠΛΥΘΗΣΜΙΑΚΗ ΤΥΠΙΚΗ ΑΠΟΚΛΙΣΗ

POPULATION STANDARD DEVIATION $$\sigma = \sqrt{\frac{\Sigma(X - \mu)^2}{N}}$$ **[3–9]**

ΠΑΡΑΔΕΙΓΜΑ - Μεση Απόκλιση

MEAN DEVIATION $$MD = \frac{\Sigma|X - \bar{X}|}{n} \quad [3\text{–}7]$$

where:
X is the value of each observation.
$\bar{X}$ is the arithmetic mean of the values.
n is the number of observations in the sample.
$||$ indicates the absolute value.

ΠΑΡΑΔΕΙΓΜΑ:

Ο αριθμός των Cappuccino που πουλήθηκαν στα Starbucks μεταξύ 4 και 7 μ.μ. Για ένα δείγμα 5 ημερών του προηγούμενου έτους είναι 20, 40, 50, 60, και 80. Βρείτε την Μέση Απόκλιση των Cappuccino που πουλήθηκαν.

Βήμα 1: Υπολογίστε τον μέσο όρο

$$\bar{x} = \frac{\sum x}{n} = \frac{20 + 40 + 50 + 60 + 80}{5} = 50$$

Βήμα 2: Αφαιρέστε τον μέσο όρο (50) από κάθε παρατήρηση, μετατρέψτε σε θετικό αρθμό αν το αποτέλεσμα είναι αρνητικό.

Number of Cappuccinos Sold Daily	$(X - \bar{X})$	Absolute Deviation
20	(20 − 50) = −30	30
40	(40 − 50) = −10	10
50	(50 − 50) = 0	0
60	(60 − 50) = 10	10
80	(80 − 50) = 30	30
	Total	80

$$MD = \frac{\Sigma|X - \bar{X}|}{n} = \frac{80}{5} = 16$$

Βήμα 3: Αθροίστε τις απόλυτες τιμές των διαφορών που βρήκατε στο Βήμα 2 και μετά διαιρέστε με τον συνολικό αριθμό των παρατηρήσεων.

Διακύμανση και Τυπική Απόκλιση

ΔΙΑΚΥΜΑΝΣΗ: Ο αριθμητικός μέσος των τετραγωνικών αποκλίσεων από την μέση τιμή.

POPULATION VARIANCE $$\sigma^2 = \frac{\Sigma(X - \mu)^2}{N}$$ **[3–8]**

ΤΥΠΙΚΗ ΑΠΟΚΛΙΣΗ: Η τετραγωνική ρίζα της διακύμανσης.

σ^2 is the population variance (σ is the lowercase Greek letter sigma). It is read as “sigma squared.”
X is the value of an observation in the population.
μ is the arithmetic mean of the population.
N is the number of observations in the population.

- Η Διακύμανση και η Τυπική Απόκλιση δεν παίρνουν αρνητικές τιμές και μηδενίζονται μόνον εάν όλες οι παρατηρήσεις είναι ίδιες μεταξύ τους.
- Για τους πληθυσμούς των οποίων οι τιμές είναι *κοντά στο μέσο*, η Διακύμανση και η Τυπική Απόκλιση θα είναι μικρή.
- Για τους πληθυσμούς των οποίων οι τιμές είναι *ιδιαίτερα διασκορπισμένες από το μέσο*, η Διακύμανση και η Τυπική Απόκλιση θα είναι μεγάλη.
- Η διακύμανση ξεπερνά την αδυναμία του εύρους, κάνοντας χρήση όλων των τιμών του πληθυσμού.

ΠΑΡΑΔΕΙΓΜΑ – Διακύμανση Πληθυσμού και Τυπική Απόκλιση Πληθυσμού

Ο αριθμός των αυτοκινήτων που πουλήθηκαν τους τελευταίους 12 μήνες, αναφέρεται παρακάτω:

Month	January	February	March	April	May	June	July	August	September	October	November	December
Citations	19	17	22	18	28	34	45	39	38	44	34	10

Ποια είναι η Διακύμανση του Πληθυσμού;

Βήμα 1: Βρείτε τον μέσο όρο.

$$\mu = \frac{\sum x}{N} = \frac{19+17+...+34+10}{12} = \frac{348}{12} = 29$$

Βήμα 2: Βρείτε τη διαφορά ανάμεσα σε κάθε παρατήρηση και το μέσο όρο, και τετραγωνίστε αυτή τη διαφορά.

Month	Citations (X)	$X - \mu$	$(X - \mu)^2$
January	19	−10	100
February	17	−12	144
March	22	−7	49
April	18	−11	121
May	28	−1	1
June	34	5	25
July	45	16	256
August	39	10	100
September	38	9	81
October	44	15	225
November	34	5	25
December	10	−19	361
Total	348	0	1,488

Βήμα 3: Αθροίστε όλες τις τετραγωνισμένες διαφορές που βρήκατε στο βήμα 3

Βήμα 4: Διαιρέστε το άθροισμα των τετραγωνισμένων διαφορών με τον αριθμό των στοιχείων του πληθυσμού.

$$\sigma^2 = \frac{\sum (X-\mu)^2}{N} = \frac{1{,}488}{12} = 124$$

Διακύμανση και Τυπική Απόκλιση Δείγματος

SAMPLE VARIANCE $$s^2 = \frac{\Sigma(X - \overline{X})^2}{n - 1}$$ [3–10]

SAMPLE STANDARD DEVIATION $$s = \sqrt{\frac{\Sigma(X - \overline{X})^2}{n - 1}}$$ [3–11]

Όπου:

s^2 είναι η διακύμανση του δείγματος

X είναι η τιμή της κάθε παρατήρησης στο δείγμα

$\overline{X}$ είναι η μέση τιμή του δείγματος (17)

n είναι ο αριθμός των παρατηρήσεων του δείγματος

ΠΑΡΑΔΕΙΓΜΑ

Το ωρομίσθιο για ένα δείγμα ημιαπασχολούμενων υπαλλήλων είναι: $12, $20, $16, $18, και $19.

Ποια είναι η διακύμανση του δείγματος;

Hourly Wage (X)	$X - \overline{X}$	$(X - \overline{X})^2$
$12	−$5	25
20	3	9
16	−1	1
18	1	1
19	2	4
$85	0	40

$$s^2 = \frac{\Sigma(X - \overline{X})^2}{n - 1} = \frac{40}{5 - 1}$$

$= 10$ in dollars squared

Το θεώρημα του Chebyshev και ο Εμπειρικός Κανόνας

Ας υποθέσουμε ότι ο αριθμητικός μέσος για ένα σύνολο δεδομένων είναι \$51.54, και η τυπική απόκλιση είναι \$7.51. Τουλάχιστον τι ποσοστό παρατηρήσεων ανήκει στο διάστημα $\pm$ 3.5 τυπικές αποκλίσεις από το μέσο όρο?

Θεώρημα Chebysev: Για οποιοδήποτε σύνολο δεδομένων (δείγμα ή πληθυσμό), η αναλογία των παρατηρήσεων οι οποίες θα ανήκουν στο διάστημα $\pm$ *k* τυπικές αποκλίσεις από το μέσο είναι τουλάχιστον $1-(1/k^2)$, όπου *k* οποιαδήποτε σταθερά μεγαλύτερη από 1.

CHEBYSHEV'S THEOREM For any set of observations (sample or population), the proportion of the values that lie within *k* standard deviations of the mean is at least $1 - 1/k^2$, where *k* is any constant greater than 1.

$$1 - \frac{1}{k^2} = 1 - \frac{1}{(3.5)^2} = 1 - \frac{1}{12.25} = 0.92$$

EMPIRICAL RULE For a symmetrical, bell-shaped frequency distribution, approximately 68 percent of the observations will lie within plus and minus one standard deviation of the mean; about 95 percent of the observations will lie within plus and minus two standard deviations of the mean; and practically all (99.7 percent) will lie within plus and minus three standard deviations of the mean.

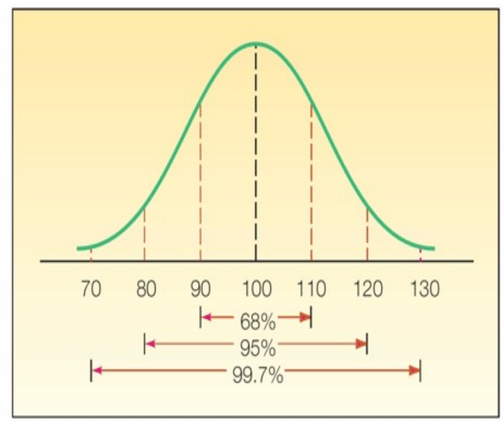


CHART 3–7 A Symmetrical, Bell-Shaped Curve Showing the Relationships between the Standard Deviation and the Observations

Εμπειρικός κανόνας: Για μία συμμετρική, μορφής καμπάνας, κατανομή, περίπου 68% των παρατηρήσεων θα είναι εντός ± 1 τυπικής απόκλισης από το μέσο, 95% αυτών εντός ± 2 τυπικών αποκλίσεων από το μέσο και 99.7% αυτών εντός ± 3 τυπικών αποκλίσεων από το μέσο.

Ο Αριθμητικός Μέσος Ορος και Τυπική Απόκλιση των Ομαδοποιημένων Δεδομένων

ARITHMETIC MEAN OF GROUPED DATA $\bar{X} = \frac{\Sigma fM}{n}$ [3–12]

Όπου $\bar{X}$ ο δειγματικός μέσος, M η κεντρική τιμή του εύρους, f η συχνότητα κάθε κλάσης και n το μέγεθος του δείγματος.

ΠΑΡΑΔΕΙΓΜΑ:

Υπολογίστε την αριθμητική μέση τιμή πώλησης των οχημάτων που αναφέρονται στον παρακάτω πίνακα συχνοτήτων.

Selling Price ($ thousands)	Frequency (f)	Midpoint (M)	fM
15 up to 18	8	$16.5	$ 132.0
18 up to 21	23	19.5	448.5
21 up to 24	17	22.5	382.5
24 up to 27	18	25.5	459.0
27 up to 30	8	28.5	228.0
30 up to 33	4	31.5	126.0
33 up to 36	2	34.5	69.0
Total	80		$1,845.0

Solving for the arithmetic mean using formula (3–12), we get:

$$\bar{X} = \frac{\Sigma fM}{n} = \frac{\$1,845}{80} = \$23.1 \text{ (thousands)}$$

STANDARD DEVIATION, GROUPED DATA $s = \sqrt{\frac{\Sigma f(M - \bar{X})^2}{n - 1}}$ [3–13]

Όπου s η δειγματική τυπική απόκλιση, M η κεντρική τιμή του εύρους, f η συχνότητα κάθε κλάση, n το μέγεθος του δείγματος και $\bar{X}$ ο δειγματικός μέσος.

ΠΑΡΑΔΕΙΓΜΑ:

Υπολογίστε την τυπική απόκλιση της τιμής πώλησης των οχημάτων που αναφέρονται στον παρακάτω πίνακα συχνοτήτων.

Selling Price ($ thousands)	Frequency (f)	Midpoint (M)	$(M - \bar{X})$	$(M - \bar{X})^2$	$f(M - \bar{X})^2$
15 up to 18	8	16.5	−6.6	43.56	348.48
18 up to 21	23	19.5	−3.6	12.96	298.08
21 up to 24	17	22.5	−0.6	0.36	6.12
24 up to 27	18	25.5	2.4	5.76	103.68
27 up to 30	8	28.5	5.4	29.16	233.28
30 up to 33	4	31.5	8.4	70.56	282.24
33 up to 36	2	34.5	11.4	129.96	259.92
	80				1,531.80

$$s = \sqrt{\frac{\Sigma f(M - \bar{X})^2}{n - 1}} = \sqrt{\frac{1531.8}{80 - 1}} = 4.403.$$