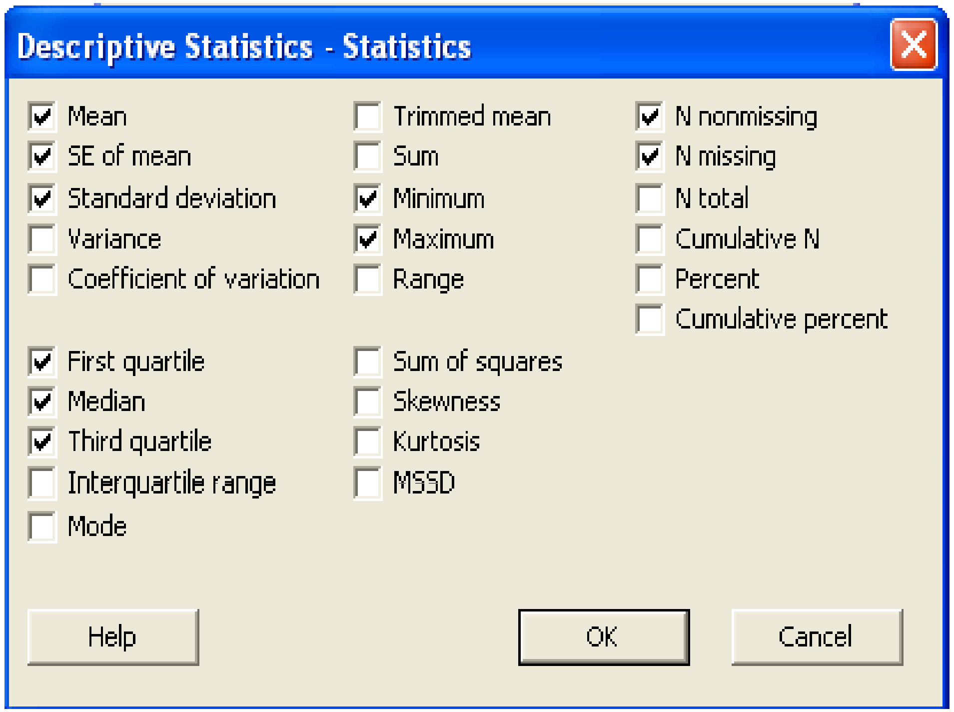
MINITAB VERSION 15

Minitab Statistical Tools


Descriptive Statistics - Statistics
Mean
SE of mean
Standard deviation
Variance
Coefficient of variation
First quartile
Median
Third quartile
Interquartile range
Mode
Trimmed mean
Sum
Minimum
Maximum
Range
Sum of squares
Skewness
Kurtosis
MSSD
N nonmissing
N missing
N total
Cumulative N
Percent
Cumulative percent
Help
OK
Cancel

Αριθμητικές Περιγραφικές Τεχνικές …

Μέτρα Κεντρικής Θέσης

- Μέση Τιμή (Mean), Διάμεσος (Median), Κορυφή (Mode)

Μέτρα Μεταβλητότητας

- Εύρος (Range), Τυπική Απόκλιση (Standard Deviation), Διακύμανση (Variance), Συντελεστής Μεταβλητότητας (Coefficient of Variation)

Μέτρα Σχετικής Τοποθεσίας

- Ποσοστημόρια (Percentile), Τεταρτημόρια (Quartile)

Μέτρα Γραμμικής Σχέσης

- Συνδιακύμανση (Covariance), Συσχέτιση (Correlation), Ευθεία Ελάχιστων Τετραγώνων (Least Square Line)

Μέτρα Κεντρικής Θέσης …

- Η ***αριθμητική μέση τιμή***, *μέσος όρος,* η απλά *μέση τιμή,* είναι το πιο δημοφιλή και χρήσιμο μέτρο κεντρικής θέσης.

- Υπολογίζεται απλά προσθέτοντας όλες τις παρατηρήσεις και διαιρώντας με τον συνολικό αριθμό των παρατηρήσεων:

Minitab Statistical Software

Συμβολισμός …

- Όταν αναφερόμαστε στον αριθμό των παρατηρήσεων ενός ***πληθυσμού***, χρησιμοποιούμε το κεφαλαίο γράμμα **N**.

- Όταν αναφερόμαστε στον αριθμό των παρατηρήσεων ενός ***δείγματος***, χρησιμοποιούμε το μικρό γράμμα **n**.

- Η μέση τιμή του ***πληθυσμού*** συμβολίζεται με το ελληνικό γράμμα ***μ***:

- Η μέση τιμή του ***δείγματος*** συμβολίζεται με: $\bar{x}$

Συμβολισμός …

	Πληθυσμός	Δείγμα
Μέγεθος	**N**	**n**
Μέση Τιμή	μ	$\bar{x}$

Αριθμητική Μέση Τιμή …

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

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

Συμβολισμός …

	Πληθυσμός	Δείγμα
Μέγεθος	N	n
Μέση Τιμή	$\mu = \dfrac{\sum_{i=1}^{N} x_i}{N}$	$\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$

Αριθμητική Μέση Τιμή …

- …είναι κατάλληλη για να περιγράψουμε δεδομένα μετρήσεων, π.χ. ύψη ανθρώπων, βαθμοί από εξετάσεις, κλπ.

- …επηρεάζεται σοβαρά από «ακραίες τιμές». Π.χ. εφόσον ένας εκατομμυριούχος μετακομίζει σε μία γειτονιά, ο μέσο οικογενειακό εισόδημα αυξάνει πολύ από πριν και δίνει λανθασμένη εντύπωση.

Μέτρα Κεντρικής Θέσης …

- Η ***διάμεσος*** υπολογίζεται βάζοντας όλες τις παρατηρήσεις στην σειρά. Η *μεσαία* παρατήρηση είναι η διάμεσος.

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

  Τα ταξινομούμε από το μικρότερο ως προς το μεγαλύτερο, και βρίσκουμε την κεντρική τιμή:

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

Δεδομένα: {0, 7, 12, 5, 14, 8, 0, 9, 22, 33} N=10 (ζυγός)

Τα ταξινομούμε από το μικρότερο ως προς το μεγαλύτερο, και υπάρχουν δύο κεντρικές τιμές (8 και 9) από τις οποίες παίρνουμε τον μέσο όρο

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

Διάμεσος = (8+9)÷2 = **8.5**

Οι διάμεσοι του δείγματος και του πληθυσμού υπολογίζονται κατά τον ίδιο τρόπο.

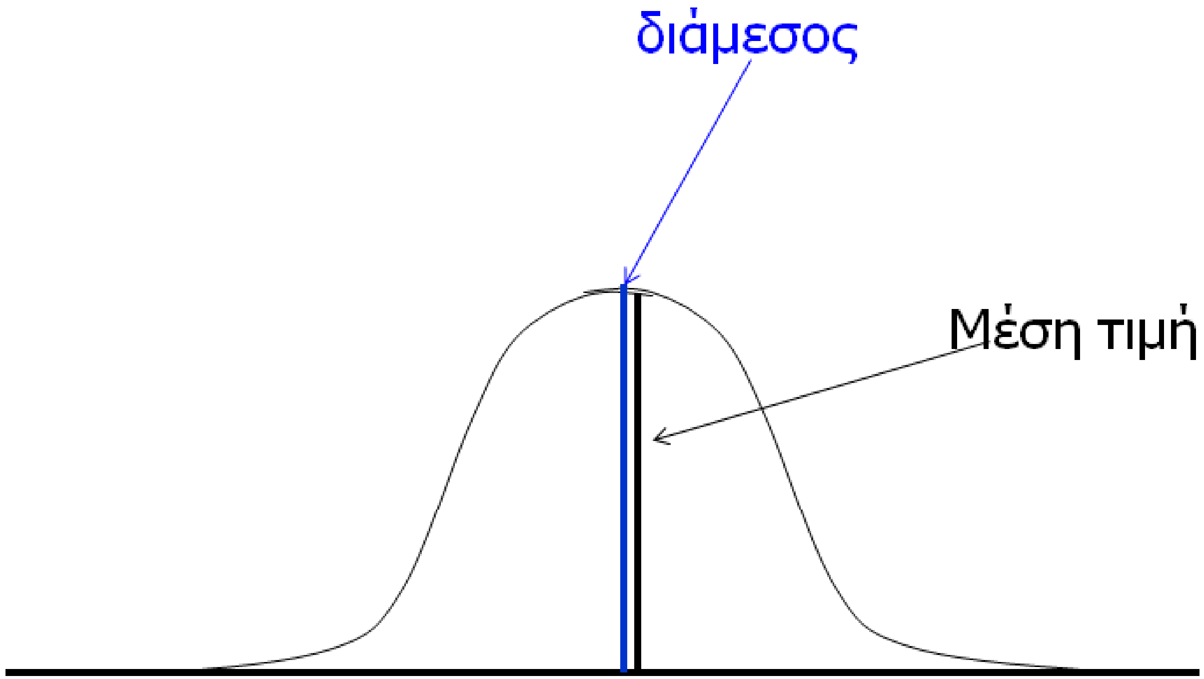
Μέτρα Κεντρικής Θέσης …

- Η ***κορυφή*** ενός συνόλου παρατηρήσεων είναι η τιμή που εμφανίζεται *πιο συχνά*.

- Ένα σύνολο δεδομένων ενδέχεται να έχει μία κορυφή ή δύο, ή περισσότερες κορυφές.

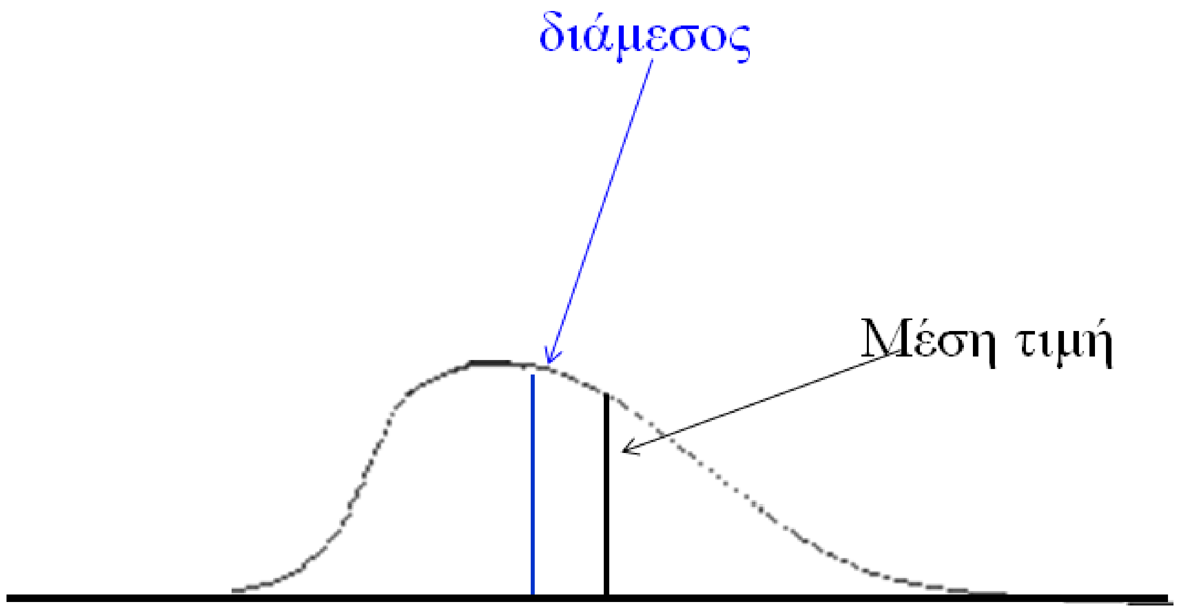
Μέση τιμή, Διάμεσος …

- Εάν η κατανομή είναι συμμετρική, η μέση τιμή, και η διάμεσος συμπίπτουν …



Μέση τιμή, Διάμεσος, …

- Εάν η κατανομή είναι ασύμμετρη, ας πούμε λοξή προς τα αριστερά ή προς τα δεξιά, τα τρία μέτρα θα διαφέρουν, π.χ.

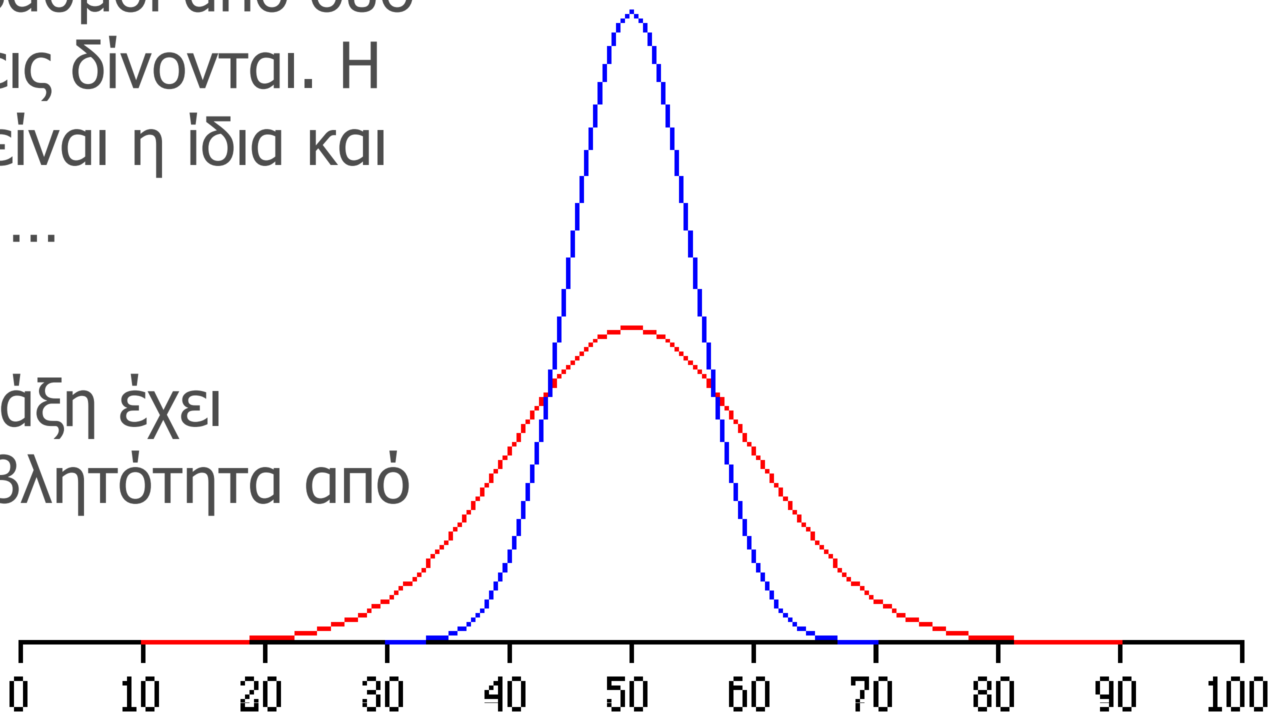


Μέτρα Μεταβλητότητας …

- Τα μέτρα της κεντρικής θέσης δεν μας δίνουν όλα τα χαρακτηριστικά για μία κατανομή. Δηλαδή, πόσο πολύ είναι οι παρατηρήσεις απλωμένες γύρω από κέντρο;

Για παράδειγμα, βαθμοί από δύο διαφορετικές τάξεις δίνονται. Η μέση τιμή (=50) είναι η ίδια και για τις δύο τάξεις …

Αλλά, η κόκκινη τάξη έχει μεγαλύτερη μεταβλητότητα από την μπλε τάξη.



Εύρος …

- Το ***εύρος*** είναι το απλούστερο μέτρο μεταβλητότητας, υπολογίζεται ως:

- Εύρος = Μεγαλύτερη παρατήρηση – Μικρότερη παρατήρηση

- 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 των δεδομένων έχουν πολύ
- διαφορετικές κατανομές …

Εύρος …

- Τα πλεονέκτημα του είναι η ευκολία με την οποία μπορεί να υπολογιστεί.

- Το βασικό μειονέκτημα είναι ότι δεν δίνει καμία πληροφορία για την διασπορά των παρατηρήσεων ανάμεσα στα δύο ακραία σημεία (min και max).

- Επομένως χρειαζόμαστε ένα μέτρο που να ενσωματώνει **όλα τα δεδομένα** και όχι μόνο δύο παρατηρήσεις. Επομένως …

Διακύμανση /Διασπορά…

- Η διασπορά και το παρεμφερή της μέτρο, τυπική απόκλιση είναι από τις πιο σημαντικές στατιστικές ποσότητες. Χρησιμοποιούνται για να μετρήσουν μεταβλητότητα, και επίσης παίζουν ένα κρίσιμο ρόλο σε όλες σχεδόν τις στατιστικές διαδικασίες για συμπερασματολογία (επαγωγή).

- Η διασπορά του πληθυσμού συμβολίζεται με $\boldsymbol{\sigma}^2$.
- (μικρό Ελληνικό γράμμα)

- Η διασπορά του δείγματος συμβολίζεται με $\boldsymbol{s}^2$.
- (μικρό “S” στο τετράγωνο)

Minitab® Statistical Software

Συμβολισμός …

	Πληθυσμός	Δείγμα
Μέγεθος	N	n
Μέση Τιμή	μ	$\bar{x}$
Διασπορά	σ^2	s^2

Διασπορά…

- Η διασπορά του **πληθυσμού**
- είναι:

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$$

- Η διασπορά του **δείγματος**
- είναι:

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

<u>Σημειώστε</u> ότι ο παρανομαστής είναι το μέγεθος του δείγματος (n) μείον 1

- Όπως μπορούμε να δούμε, έχουμε να υπολογίσουμε την μέση τιμή (x-παύλα) για να υπολογίσουμε την διασπορά του δείγματος.

- Εναλλακτικά, υπάρχει ένας πιο σύντομος τύπος για να υπολογίσουμε την διασπορά του δείγματος άμεσα από τα δεδομένα χωρίς να έχουμε το ενδιάμεσο βήμα του υπολογισμού της μέσης τιμής. Αυτός ο τύπος δίνεται από:

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

Τυπική Απόκλιση …

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

- Η τυπική απόκλιση του πληθυσμού: $\sigma = \sqrt{\sigma^2}$

- Η τυπική απόκλιση του δείγματος: $s = \sqrt{s^2}$

Συμβολισμός …

	Πληθυσμός	Δείγμα
Μέγεθος	N	n
Μέση Τιμή	μ	$\bar{x}$
Διακύμανση	σ^2	s^2
Τυπική Απόκλιση	$\sigma = \sqrt{\sigma^2}$	$s = \sqrt{s^2}$

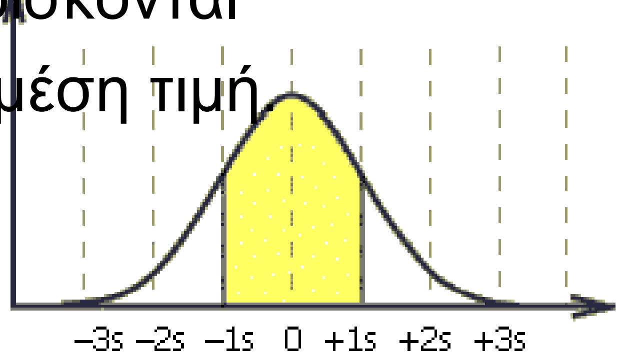
Ερμηνεύοντας την Τυπική Απόκλιση …

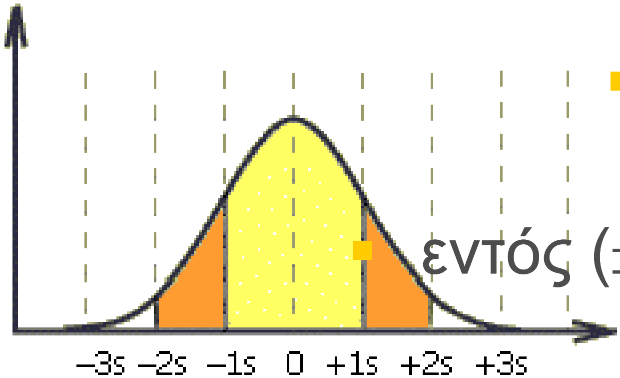
- Η τυπική απόκλιση μπορεί να χρησιμοποιηθεί για την σύγκριση της μεταβλητότητας αρκετών κατανομών και τον χαρακτηρισμό της γενικής μορφής μιας κατανομής. Εάν το ιστόγραμμα έχει **το σχήμα της καμπάνας**, μπορούμε να χρησιμοποιήσουμε τον ***Εμπειρίκο Κανόνα***, ο οποίος λέει:

1) Προσεγγιστικά 68% των παρατηρήσεων βρίσκονται εντός (±) μιας τυπικής απόκλισης από την μέση τιμή.
2) Προσεγγιστικά 95% των παρατηρήσεων βρίσκονται εντός (±) δύο τυπικών αποκλίσεων από την μέση τιμή.
3) Προσεγγιστικά 99.7% των παρατηρήσεων βρίσκονται εντός (±) τρεις τυπικές αποκλίσεις από την μέση τιμή.

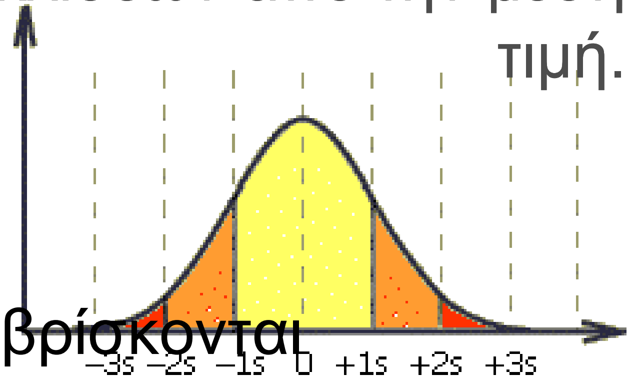
Ο Εμπειρικός Κανόνας …

- Προσεγγιστικά 68% των παρατηρήσεων βρίσκονται
- εντός (±) **μιας** τυπικής απόκλισης από την μέση τιμή.





- Προσεγγιστικά 95% των παρατηρήσεων βρίσκονται
- εντός (±) **δύο** τυπικών αποκλίσεων από την μέση τιμή.



- Προσεγγιστικά 99.7% των παρατηρήσεων βρίσκονται
- εντός (±) **τρεις** τυπικές αποκλίσεις από την μέση τιμή.

Το Θεώρημα του Chebysheff's ...

- Μία πιο γενική ερμηνεία της τυπικής απόκλισης εξάγεται από το θεώρημα του ***Chebysheff's***, το ποιο εφαρμόζεται σε όλες τις κατανομές με οποιαδήποτε μορφή.

- Το ποσοστό των παρατηρήσεων στο δείγμα πέφτει εντός ($\pm$) **k** τυπικών αποκλίσεων από την μέση τιμή είναι *τουλάχιστον*:

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

Συντελεστής Μεταβλητότητας …

- Ο ***συντελεστής μεταβλητότητας*** ενός συνόλου παρατηρήσεων είναι η τυπική απόκλιση των παρατηρήσεων διαιρούμενη με την μέση τιμή, δηλαδή:

- Ο συντελεστής μεταβλητότητας του πληθυσμού = CV = $\frac{\sigma}{\mu}$

- Συντελεστής Μεταβλητότητας του δείγματος = cv = $\frac{s}{\bar{x}}$

Συμβολισμός …

	Πληθυσμός	Δείγμα
Μέγεθος	N	n
Μέση Τιμή	μ	$\overline{x}$
Διακύμανση	σ^2	s^2
Τυπική Απόκλιση	σ	s
Συντελεστής Μεταβλητότητας	CV	cv

Minitab Statistical Software

Συντελεστής Μεταβλητότητας …

- Αυτός ο συντελεστής δείχνει ένα ***αναλογικό*** μέτρο της μεταβλητότητας, π.χ.

- Η τυπική απόκλιση του 10 ενδέχεται να ληφθεί ως μεγάλη όταν η μέση τιμή είναι 100, αλλά μόνο ως μέτρια όταν η μέση τιμή είναι 500.

Μέτρα Μεταβλητότητας …

- Εάν τα δεδομένα είναι συμμετρικά, χωρίς ακραίες τιμές, χρησιμοποιήστε το εύρος και την τυπική απόκλιση.

- Εάν συγκρίνουμε την μεταβλητότητα ανάμεσα σε δύο σύνολα δεδομένων χρησιμοποιήστε τον συντελεστής μεταβλητότητας.

- Τα μέτρα μεταβλητότητας που εισήχθην σε αυτή την ενότητα μπορούν να χρησιμοποιηθούν μόνο για **διαστημικά** δεδομένα.

Μέτρα Σχετικής Τοποθεσίας & Box Plots

- Μέτρα Σχετικής Τοποθεσίας σχεδιάζονται για να προβάλουν πληροφόρηση σχετικά με την ***τοποθεσία*** κάποιων συγκεκριμένων τιμών ***σε σχέση*** με ολόκληρο το σύνολο των δεδομένων.

- ***Ποσοστημόριο:*** το P^o ποσοστημόριο είναι η τιμή από την οποία *P* ποσοστό των τιμών είναι *μικρότερο από* την τιμή αυτή και (100-*P*)% είναι μεγαλύτερο από την τιμή αυτή.

- Υποθέστε ότι το σκορ σας είναι το 60^o ποσοστημόριο στο GMAT τεστ, το οποίο σημαίνει ότι το 60% των άλλων σκορ ήταν *κάτω* από το δικό σας, ενώ το 40% των άλλων σκορ ήταν *κάτω* από το δικό σας.

Ποσοστημόριο …

- Έχουμε ειδικά ονόματα για το 25ο, 50ο, και 75ο ποσοστημόριο, χαρακτηριστικά ***τεταρτημόρια.***

- Το πρώτο τεταρτημόριο χαρακτηρίζει Q_1 = 25ο ποσοστημόριο.

- Το δεύτερο τεταρτημόριο, Q_2 = 50ο ποσοστημόριο (το οποίο είναι επίσης η διάμεσος).

- Το τρίτο τεταρτημόριο, Q_3 = 75ο ποσοστημόριο.

- Μπορούμε επίσης να αντιστοιχίσουμε ποσοστημόρια σε πεμπτημόρια (quintiles, fifths) και δεκατημόρια (deciles, tenths).

Minitab Statistical Software

Χρήσιμα Ποσοστημόρια …

- Πρώτο δεκατημόριο = 10ο ποσοστημόριο
- Πρώτο τεταρτημόριο, Q_1, = 25ο ποσοστημόριο
- Διάμεσος, Q_2, = 50ο ποσοστημόριο
- Τρίτο τεταρτημόριο, Q_3, = 75ο ποσοστημόριο
- Ένατο δεκατημόριο = 90ο ποσοστημόριο

- **Σημειώστε:** Εάν ο βαθμό σου σε φέρνει στο 80ο ποσοστημόριο, αυτό δεν σημαίνει ότι απάντησες το 80% των ερωτήσεων σωστά – αυτό σημαίνει ότι το 80% των συμφοιτητών σου είχε σκορ **χαμηλότερο** από το δικό σου. Δείχνει την θέση σου σε σχέση με τους άλλους.

Ενδοτεταρτημοριακό Εύρος …

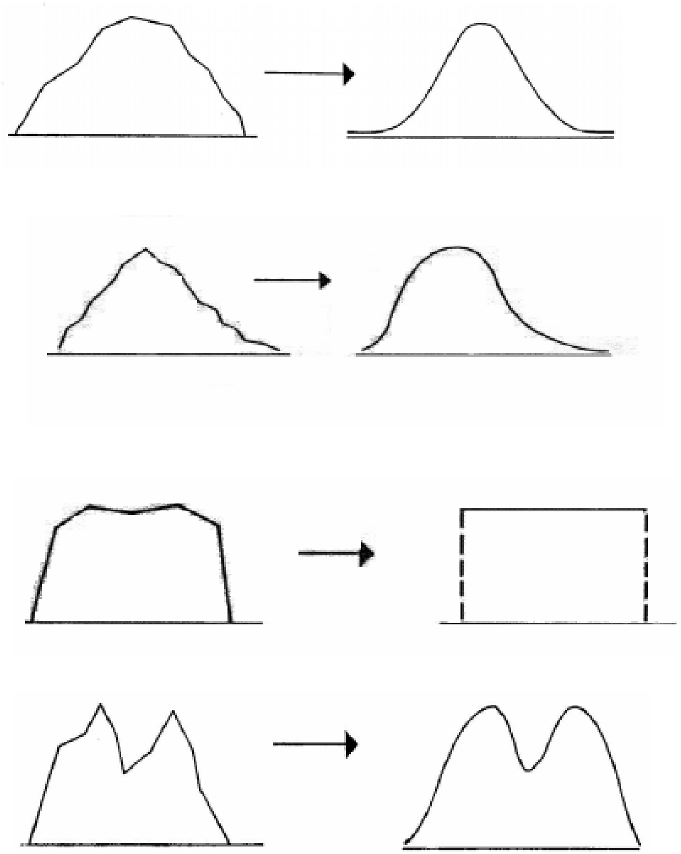
- Τα τεταρτημόρια μπορούν να χρησιμοποιηθούν για να δημιουργήσουν ένα άλλο μέτρο μεταβλητότητας, το ***ενδοτεταρτημοριακό εύρος*** το οποίο ορίζεται ως εξής:

$$\text{Ενδοτεταρτημοριακό Εύρος} = Q_3 - Q_1$$

- Το ενδοτεταρτημοριακό εύρος μετράει το άπλωμα του 50% των μεσαίων παρατηρήσεων.
- Μεγάλες τιμές αυτής της στατιστικής σημαίνουν ότι το 1ο και 3ο τεταρτημόριο απέχουν υποδεικνύοντας υψηλό επίπεδο μεταβλητότητας.

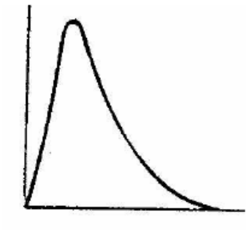
Μέτρα Ασυμμετρίας

- Οι *καμπύλες συχνοτήτων, πέραν της προφανούς χρησιμότητάς τους στο πλαίσιο της Περιγραφικής Στατιστικής. Οι καμπύλες συχνοτήτων μπορεί να έχουν* διάφορες μορφές όπως:

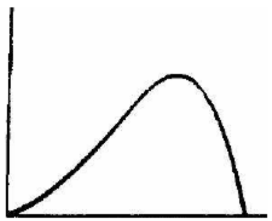


Συντελεστής Λοξότητας

- Όταν μια *καμπύλη συχνοτήτων είναι συμμετρική ως προς τον κατακόρυφο άξονα που* διέρχεται από την κορυφή της κατανομής, όπως η πρώτη από τις παραπάνω, τότε η κατανομή είναι *συμμετρική. Τα δύο άκρα της καμπύλης λέγονται ουρές της κατανομής* και πλησιάζουν ασυμπτωτικά τον άξονα των τιμών. Προφανώς, σε μια *συμμετρική κατανομή, δεξιά και αριστερά του άξονα συμμετρίας βρίσκεται το ίδιο ποσοστό* παρατηρήσεων (50%).
- Όταν η *καμπύλη συχνοτήτων δεν είναι συμμετρική, δηλαδή, όταν δεξιά και αριστερά* του κατακόρυφου άξονα που περνάει από την κορυφή δε βρίσκεται το ίδιο ποσοστό παρατηρήσεων, τότε η κατανομή είναι *ασύμμετρη. Υπάρχουν δύο ειδών ασυμμετρίες:*
- *Θετικές και αρνητικές. Μια καμπύλη συχνοτήτων παρουσιάζει θετική ασυμμετρία* όταν οι περισσότερες παρατηρήσεις βρίσκονται δεξιά της κορυφής, ενώ, παρουσιάζει *αρνητική ασυμμετρία όταν οι περισσότερες παρατηρήσεις βρίσκονται αριστερά της* κορυφής.



Θετική ασυμμετρία



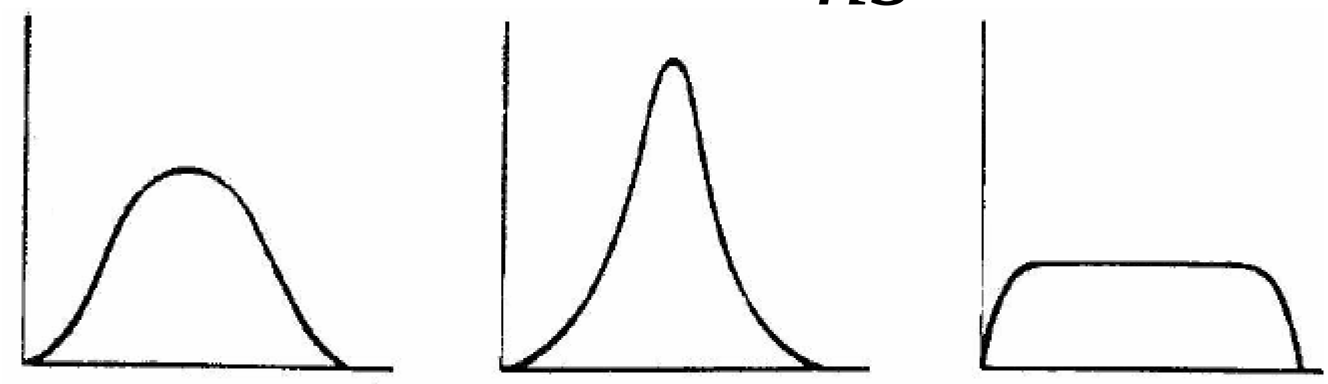
Αρνητική ασυμμετρία

$$skewness = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^3}{ns^3}$$

Συντελεστής Κύρτωσης

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

$$kurtosis = \frac{\sum_{i}^{n} (x_i - \overline{x})^4}{ns^4}$$



Μεσόκυρτη *Λεπτόκυρτη* *Πλατύκυρτη*