ΆΣΚΗΣΗ 1

Έστω ότι είστε μέλος μιας επιτροπής που καλείται να αξιολογήσει σύμφωνα με την τελική τους επίδοση τρεις μαθητές που αποφοίτησαν από τρία σχολεία με διαφορετικό σύστημα βαθμολόγησης. Δίνεται η τελική επίδοση κάθε μαθητή, και η μέση τιμή και η τυπική απόκλιση της επίδοσης των μαθητών του σχολείου όπου φοιτούσε.

Μαθητής	Τελική Επίδοση Μαθητή	Μέση Επίδοση Σχολείου	Τυπική απόκλιση Επίδοσης Σχολείου
Γιώργος	2,7	3,2	0,8
Πέτρος	87	75	20
Αννούλα	8,6	8	0,4

α) Σε ποιό σχολείο ήταν πιο διεσπαρμένη η κατανομή βαθμολογίας;

Έστω Α, Β, Γ τα τρία σχολεία. Για να συγκριθεί η διασπορά των τριών κατανομών βαθμολογίας θα υπολογιστεί ο συντελεστής μεταβλητότητας (Coefficient of Variation-CV) για τις τρεις κατανομές.

$$CV_{\mathrm{A}} = \frac{\sigma_{\mathrm{A}}}{\overline{\mathrm{X}}_{\mathrm{A}}} = \frac{0{,}8}{3{,}2} = 0{,}25,\quad CV_{\mathrm{B}} = \frac{\sigma_{\mathrm{B}}}{\overline{\mathrm{X}}_{\mathrm{B}}} = \frac{20}{75} = 0{,}26,\quad CV_{\Gamma} = \frac{\sigma_{\Gamma}}{\overline{\mathrm{X}}_{\Gamma}} = \frac{0{,}4}{8} = 0{,}05$$

Συμπερασματικά, πιο διεσπαρμένη είναι η κατανομή βαθμολογίας του σχολείου Β.

β) Ποιός μαθητής είχε την καλύτερη επίδοση λαμβάνοντας υπόψη την μέση επίδοση στο σχολείο όπου φοιτούσε; Εξηγήστε την απάντησή σας αριθμητικά.

Για να βρούμε τον μαθητή με την καλύτερη επίδοση πρέπει πρώτα να μετατρέψουμε τους αρχικούς βαθμούς σε τυπικούς βαθμούς. Οι τυπικοί (z) βαθμοί για κάθε μαθητή υπολογίζονται ως εξής:

$z_{\text{Γιώργος}} = \dfrac{\mathrm{X}_{\text{Γιώργος}} - \overline{\mathrm{X}}_{\mathrm{A}}}{\sigma_{\mathrm{A}}} = \dfrac{2{,}7-3{,}2}{0{,}8} = -\dfrac{0{,}5}{0{,}8} = -0{,}62$, ο βαθμός του Γιώργου είναι 0,62 τυπικές αποκλίσεις μικρότερος από το μέσο βαθμό του σχολείου του.

$z_{\text{Πέτρος}} = \dfrac{\mathrm{X}_{\text{Πέτρος}} - \overline{\mathrm{X}}_{\mathrm{B}}}{\sigma_{\mathrm{B}}} = \dfrac{87-75}{20} = \dfrac{12}{20} = 0{,}6$, ο βαθμός του Πέτρου είναι 0,6 τυπικές αποκλίσεις μεγαλύτερος από το μέσο βαθμό του σχολείου του.

$z_{\text{Αννούλα}} = \dfrac{\mathrm{X}_{\text{Αννούλα}} - \overline{\mathrm{X}}_{\Gamma}}{\sigma_{\Gamma}} = \dfrac{8{,}6-8}{0{,}4} = \dfrac{0{,}6}{0{,}4} = 1{,}5$, ο βαθμός της Αννούλας είναι 1,5 τυπικές αποκλίσεις μεγαλύτερος από το μέσο βαθμό του σχολείου της.

Άρα η **Αννούλα** είχε την καλύτερη επίδοση.

ΆΣΚΗΣΗ 2

Η τάξη ενός δημοτικού σχολείου έτρεξε 1km σε 11 λεπτά κατά μέσο όρο με τυπική απόκλιση 3 λεπτά. Ο Κωστάκης, ένας μαθητής της τάξης, έτρεξε 1km σε 8 λεπτά. Η τάξη ενός γυμνασίου έτρεξε 1km σε 9 λεπτά κατά μέσο όρο, με τυπική απόκλιση 2 λεπτών. Ο Γιωργάκης, ένας μαθητής της τάξης, έτρεξε 1Km σε 8,5 λεπτά. Μια τάξη λυκείου έτρεξε 1km σε 7 λεπτά κατά μέσο όρο με τυπική απόκλιση 4 λεπτά. Η Αλέκα, μιας μαθήτρια της τάξης έτρεξε 1Km σε 8 λεπτά.

a. Γιατί ο Γιωργάκης θεωρείται ταχύτερος από την Αλέκα, παρόλο που η Αλέκα έτρεξε το χιλιόμετρο πιο γρήγορα απ' αυτόν;

Παρατηρούμε ότι ο Γιωργάκης έτρεξε 0,5 λεπτό ταχύτερα από τον μέσο όρο της τάξης του (9 λεπτά), ενώ η Αλέκα έτρεξε 1 λεπτό αργότερα από τον μέσο όρο της τάξης της (7 λεπτά). Κατά συνέπεια, ο Γιωργάκης μπορεί να θεωρηθεί ταχύτερος από την Αλέκα υπό την προϋπόθεση ότι θα ληφθούν υπόψη οι επιδόσεις της τάξης όπου ανήκει ο καθένας.

b. Ποιός από τους τρεις είναι ο ταχύτερος μαθητής αν ληφθεί υπόψη η επίδοση της τάξης. Αιτιολογήστε την απάντησή σας.

Έστω Α, Β, Γ οι τρεις τάξεις. Για να βρούμε τον ταχύτερο μαθητή πρέπει πρώτα να μετατρέψουμε τους αρχικούς χρόνους σε τυπικούς. Οι τυπικοί (z) χρόνοι για κάθε μαθητή υπολογίζονται ως εξής:

$z_{\text{Κωστάκης}} = \frac{X_{\text{Κωστάκης}} - \overline{X}_{A}}{\sigma_{A}} = \frac{8-11}{3} = -\frac{3}{3} = -1$, ο βαθμός του Κωστάκη είναι 1 τυπική απόκλιση μικρότερος από το μέσο χρόνο της τάξης του.

$z_{\text{Γιωργάκης}} = \frac{X_{\text{Γιωργάκης}} - \overline{X}_{B}}{\sigma_{B}} = \frac{8{,}5-9}{2} = -\frac{0{,}5}{2} = -0{,}25$, ο βαθμός του Γιωργάκη είναι 0,25 τυπικές αποκλίσεις μικρότερος από το μέσο χρόνο της τάξης του.

$z_{\text{Αλέκα}} = \frac{X_{\text{Αλέκα}} - \overline{X}_{\Gamma}}{\sigma_{\Gamma}} = \frac{8-7}{4} = \frac{1}{4} = 0{,}25$, ο βαθμός της Αλέκας είναι 0,25 τυπικές αποκλίσεις μεγαλύτερος από το μέσο χρόνο της τάξης της.

Άρα ο Κωστάκης είναι ο ταχύτερος αν ληφθεί υπόψη η επίδοση της τάξης.

Άσκηση 3

Εννέα ποντικάκια τρέχουν μέσα σε ένα λαβύρινθο. Ο χρόνος σε λεπτά που έκανε το κάθε ποντικάκι για να βγει από το λαβύρινθο δίνεται στον παρακάτω πίνακα:

Π1	Π2	Π3	Π4	Π5	Π6	Π7	Π8	Π9
1	2,5	3	1,5	2	1,25	1	0,9	30

Ποιο από τα τρία μέτρα κεντρικής τάσης είναι το καταλληλότερο για την περίπτωση αυτή; Δικαιολογήστε την απάντησή σας.

Απάντηση

Υπολογίζουμε τον μέσο όρο, τη διάμεσο και την επικρατούσα τιμή της κατανομής.

$$\bar{X}=\frac{1+2,5+3+1,5+2+1,25+1+0,9+30}{9}=4,79$$

Τοποθετούμε τις παρατηρήσεις σε αύξουσα διάταξη.

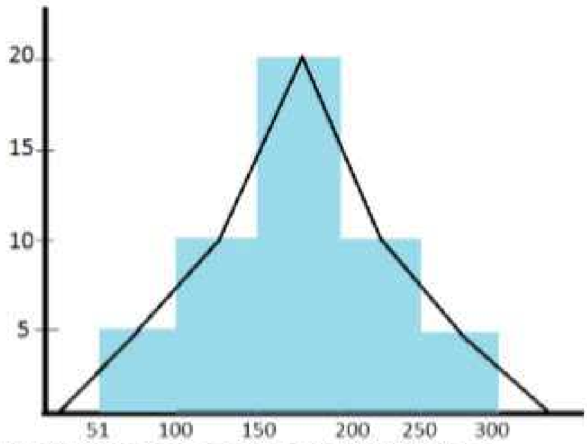
0,9	1	1	1,25	1,5	2	2,5	3	30

Άρα Μ=1,5 και Τ=1.

Διαπιστώνουμε ότι ο μέσος όρος επηρεάζεται από την ακραία τιμή 30 και υπερεκτιμά την κεντρική τάση της κατανομής, κάτι που δεν συμβαίνει στην περίπτωση της διαμέσου και της επικρατούσας τιμής. Λόγω του μικρού αριθμού των παρατηρήσεων, θα προτιμήσουμε τη διάμεσο από την επικρατούσα τιμή διότι η τελευταία θα μπορούσε να επηρεαστεί σημαντικά με την προσθήκη ή την τροποποίηση μιας και μόνο παρατήρησης (π.χ. αν προσθέσουμε την τιμή 3 μπορούμε να αναφέρουμε το 3 ως επικρατούσα τιμή).

Άσκηση Οι ετήσιες πωλήσεις σε εκατ. ευρώ (€) 50 Εμπορικών Επιχειρήσεων ήταν:

Διαστήματα Πωλήσεων	Αριθμός Εμπορικών Επιχειρήσεων (Συχνότητα)	Κέντρο Διαστήματος
51-100	5	75,5
101-150	10	125,5
151-200	20	175,5
201-250	10	225,5
251-300	5	275,5



α) Αφού υπολογίσετε τα κέντρα των διαστημάτων, να κατασκευασετε το ιστογραμμα και το πολύγωνο συχνοτήτων. Τι παρατηρείτε;

β) Να υπολογιστούν η μέση τιμή, η διάμεσος και η επικρατούσα τιμή (να υποθέσετε ότι οι παρατηρήσεις συγκεντρώνονται στο μέσο κάθε διαστήματος).

Παρατηρούμε ότι κατανομή της μεταβλητής είναι κανονική.

β) Από το (α) γνωρίζουμε ότι η μέση τιμή, η διάμεσος και η επικρατούσα τιμή θα ταυτίζονται. Άρα μέση τιμή = διάμεσος = επικρατούσα τιμή = 175,5 εκατομμύρια ευρώ.

Άσκηση 4 Σε πρόσφατη έρευνα ρωτήθηκαν κάποια άτομα σχετικά με το ποιο θεωρο το σημαντικότερο πρόβλημα σήμερα. Οι απαντήσεις στην ερώτηση αυτή διασταυρώθηκ τις διάφορες ηλικιακές κατηγορίες των ατόμων που συμμετείχαν στην έρευνα. Μέρο αποτελεσμάτων δίνεται στον παρακάτω πίνακα:

Πίνακας Συνάφειας "Ηλικιακής Κατηγορίας" με "Σημαντικότερο Πρόβλημα"

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

Α) Να συμπληρωθούν τα στοιχεία που λείπουν (δηλώνονται με ?).

Β) Πόσα άτομα συμμετείχαν στο δείγμα;

Γ) Ποιο είναι το ποσοστό % των ατόμων ηλικίας 35-44 ετών που θεωρούν ως σημαντικότερ πρόβλημα το τζόγο;

Δ) Από τα άτομα που θεωρούν ως σημαντικότερο πρόβλημα το Οικονομικό πόσο τοις εκατ ανήκει στην ηλικιακή κατηγορία 45-54 ετών;

Ε) Ποιο είναι το ποσοστό % των ατόμων ηλικίας 45-64 ετών που θεωρούν ως σημαντικότερ πρόβλημα την υγεία;

Απάντηση

	Τζόγος	Στεγαστικό	Σπουδές	Οικονομικό	Υγεία	Σύνολα
18-24	**24**	**4**	**23**	**82**	**12**	**145**
25-34	**42**	**2**	**20**	**146**	**13**	**223**
35-44	**54**	**0**	**12**	**136**	**18**	**220**
45-54	**40**	**1**	**6**	**121**	**9**	**177**
55-64	**37**	**1**	**7**	**68**	**15**	**128**
65+	**20**	**5**	**3**	**61**	**13**	**102**
Σύνολα	**217**	**13**	**71**	**614**	**80**	**995**

Β. 995 άτομα

Γ. 54 / 220 = 0,245 ή 24,5%

Δ. 121 / 614 = 0,197 ή 19,7%

Ε. (9+15)/(177+128)= 24/305 = 0,078 ή 7,8%

ΣΤ. 23 / 995 = 0,023 ή 2,3%

Άσκηση 7. Να χαρακτηρίσετε με Σωστό ή Λάθος τις παρακάτω προτάσεις.

1. Αν αυξήσουμε το μέγεθος του δείγματος, το σφάλμα δειγματοληψίας αναμένεται να μειωθεί.	1. Σ,
2. Η σύνθεση ενός δείγματος κατά στρώματα συνήθως δεν αντιπροσωπεύει την σύνθεση του πληθυσμού από τον οποίο προέρχεται.	2. Λ,
3. Ένα πρόβλημα κατά τη δειγματοληψία είναι ότι ο στατιστικός πληθυσμός μπορεί να μην είναι αντιπροσωπευτικός του πληθυσμού-στόχου.	3.Σ,
4. Όταν ο ερευνητής δεν γνωρίζει τον ακριβή αριθμό μελών του πληθυσμού, είναι απαραίτητο να χρησιμοποιήσει μια μέθοδο μη τυχαίας δειγματοληψίας.	4. Σ,
5. Σφάλμα δειγματοληψίας είναι η διαφορά ανάμεσα σε ένα στατιστικό του δείγματος και σε μια παράμετρο του πληθυσμού, η οποία οφείλεται σε τυχαίους παράγοντες και σε ατομικές διαφορές ανάμεσα στα υποκείμενα που αποτελούν κάθε φορά το δείγμα μας.	5.Σ,
6. Κατά την εγκυρότητα ενός ψυχομετρικού τεστ που μετρά το δείκτη ευφυΐας, το ζήτημα που τίθεται δεν είναι αν η συγκεκριμένη τιμή του δείκτη αντανακλά τον πραγματικό δείκτη ευφυΐας των υποκειμένων, αλλά αν το τεστ τείνει να δίνει γενικά μετρήσεις πολύ κοντά στον πραγματικό δείκτη ευφυΐας των υποκειμένων.	6.Λ,
7. Η πιο συχνά χρησιμοποιούμενη μέθοδος στην εκπαιδευτική έρευνα είναι η μέθοδος της απλής τυχαίας δειγματοληψίας.	7.Λ,
8. Η ποσοστιαία δειγματοληψία επιτρέπει στον ερευνητή να βελτιώσει τη σύνθεση ενός δείγματος ευκολίας.	8.Σ,
9. Η αξιοπιστία αφορά στο κατά πόσο μια δοκιμασία αναδεικνύει το πραγματικό μέγεθος του υπό μέτρηση χαρακτηριστικού.	9.Σ,
10. Ένας από τους πιο διαδεδομένους δείκτες αξιοπιστίας είναι ο δείκτης εσωτερικής συνέπειας του Cronbach.	10.Σ,
11. Μια προϋπόθεση της αξιοπιστίας ισοδύναμων ή παράλληλων μορφών είναι οι ισοδύναμες μορφές να μετρούν το ίδιο χαρακτηριστικό.	11.Σ,
12. Το συστηματικό σφάλμα της μέτρησης οφείλεται σε απροσδιόριστα αίτια.	12.Λ,
13. Ο όρος τυχαίο σφάλμα αναφέρεται σε όλα τα λάθη του σχεδιασμού μιας μελέτης από τον ερευνητή.	13.Λ,
14. Ο καλός σχεδιασμός μιας μελέτης μπορεί να προφυλάξει από την εμφάνιση συστηματικού σφάλματος.	14.Σ,
15. Ένας ερευνητής κατά την επιλογή των υποκειμένων της έρευνας βασίστηκε μόνο σε εθελοντές συμμετέχοντες. Αυτό είναι ένα είδος συστηματικού σφάλματος.	15.Σ

Ασυμμετρία και Κυρτότητα:

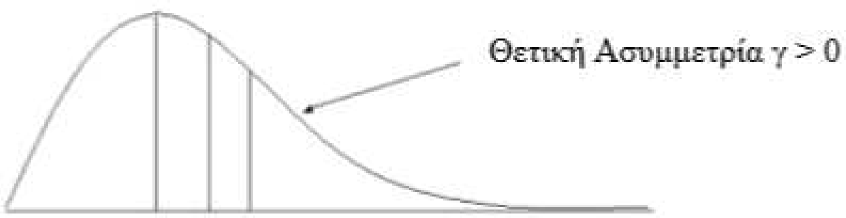
Α) Συντελεστής Ασυμμετρίας:

Η κατανομή ενός πληθυσμού μπορεί να είναι είτε συμμετρική είτε μη συμμετρική. Στην πρώτη περίπτωση η κορυφή, η διάμεσος και η μέση τιμή συμπίπτουν . Στις άλλες περιπτώσεις ένα από τα τμήματα στα οποία χωρίζει την κατανομή η κορυφή περιέχει περισσότερες παρατηρήσεις από το άλλο. Υπάρχουν δύο ειδών ασυμμετρίες, η θετική ασυμμετρία στην οποία οι περισσότερες παρατηρήσεις, καθώς επίσης και η διάμεσος και η μέση τιμή, βρίσκονται δεξιά της κορυφής και στην περίπτωση αυτή μάλιστα ισχύει $M_0 < x_\delta < \bar{x}$, και η αρνητική ασυμμετρία στην οποία οι περισσότερες παρατηρήσεις, όπως η διάμεσος και η μέση τιμή, βρίσκονται αριστερά της κορυφής και στην περίπτωση αυτή μάλιστα ισχύει $\bar{x} < x_\delta < M_0$.

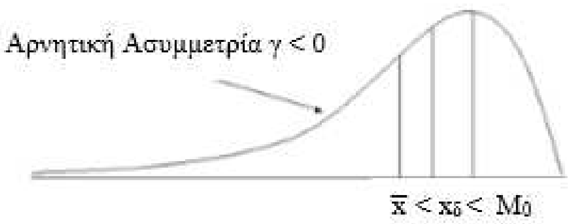
Σαν αριθμητικό μέτρο καθορισμού της ασυμμετρίας το συνηθέστερο είναι ο συντελεστής ασυμμετρίας με βάση τις ροπές ο οποίος ορίζεται ως:

$$\gamma = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{\left\{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}\right\}^3}.$$

Όταν $\gamma > 0$ έχουμε θετική ασυμμετρία, όταν $\gamma < 0$ έχουμε αρνητική ασυμμετρία, ενώ για $\gamma = 0$ έχουμε συμμετρία.



$M_0 < x_\delta < \bar{x}$



Β) Συντελεστής Κυρτότητας:

Μια κατανομή η οποία έχει σχετικά μεγάλη μέγιστη συχνότητα (κορυφή) και επομένως μεγάλη συγκέντρωση τιμών γύρω από το μέσο λέγεται λεπτόκυρτη (leptokurtic), ενώ αν η μέγιστη συχνότητα της είναι σχετικά μικρή λέγεται πλατύκυρτη (platykurtic). Κατανομές που προσεγγίζονται από την κανονική κατανομή λέγονται μεσόκυρτες (mesokurtic).

Ένα μέτρο που εκφράζει το βαθμό κυρτότητας μιας κατανομής είναι ο συντελεστής κύρτωσης του Pearson ο οποίος ορίζεται από τον τύπο:

$$\alpha = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^4}{\left\{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}\right\}^4}.$$

Επειδή για κανονικές κατανομές έχουμε α = 3 συνηθίζεται να μετράμε την κυρτότητα με την διαφορά α -3, η οποία για λεπτόκυρτες κατανομές παίρνει θετικές τιμές (θετική κύρτωση), ενώ για πλατύκυρτες κατανομές γίνεται αρνητική (αρνητική κύρτωση).