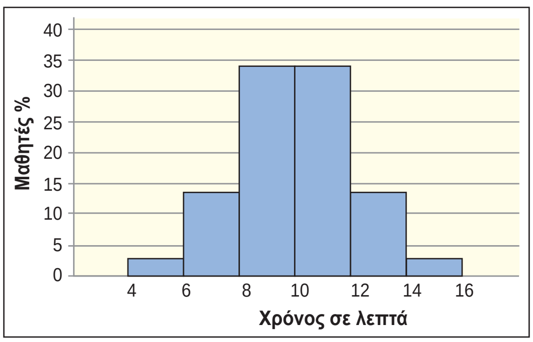
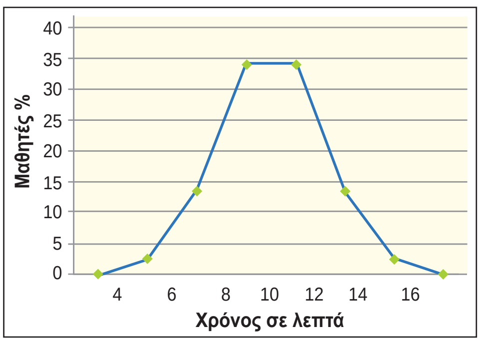
ΕΝΟΤΗΤΑ 2.4 : ΚΑΝΟΝΙΚΗ ΚΑΤΑΝΟΜΗ ΚΑΙ ΕΦΑΡΜΟΓΕΣ

Διερεύνηση

Το παρακάτω ιστόγραμμα και το αντίστοιχο πολύγωνο σχετικών συχνοτήτων παρουσιάζουν τον χρόνο, σε λεπτά, που χρειάζονται οι μαθητές για να πάνε από το σπίτι τους στο σχολείο. Η μέση τιμή είναι 10 λεπτά και η τυπική απόκλιση 2 λεπτά.



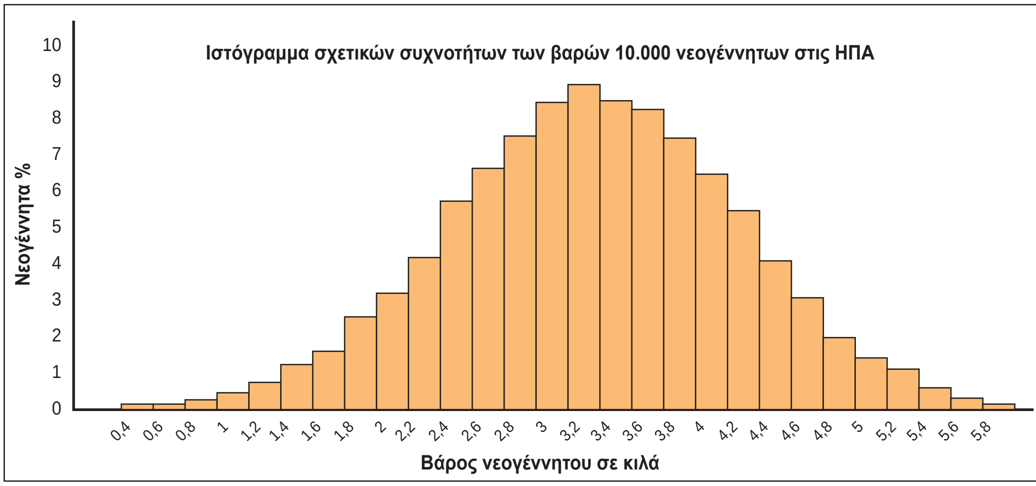


1) Τι έχετε να παρατηρήσετε για τη συμμετρία της κατανομής;
2) Τι ποσοστό μαθητών χρειάζονται το πολύ 10 λεπτά και πόσοι τουλάχιστον 10 λεπτά για να πάνε στο σχολείο;
3) Τι ποσοστό μαθητών χρειάζονται από 8 εως 12 λεπτά για να πάνε στο σχολείο;
4) Τι ποσοστό μαθητών χρειάζονται τουλάχιστον 8 λεπτά για να πάνε στο σχολείο;

Βασικές μαθηματικές έννοιες - Ιδέες - Διεργασίες

Κανονική κατανομή

Στο παρακάτω ιστόγραμμα σχετικών συχνοτήτων επί τοις εκατό (%) παριστάνεται η κατανομή 10.000 νεογέννητων στις ΗΠΑ, ως προς το βάρος τους (σε κιλά). Τα νεογέννητα κατανέμονται σε κλάσεις ίσου πλάτους. Από αυτό το ιστόγραμμα αντλούμε αρκετές πληροφορίες, όπως π.χ. ότι η κλάση με τα περισσότερα νεογέννητα είναι από 3,2 έως 3,4 κιλά, η οποία περιέχει γύρω στο 9% του δείγματος, δηλαδή 900 νεογέννητα περίπου.



Ακολουθεί το αντίστοιχο πολύγωνο σχετικών συχνοτήτων (%) της μεταβλητής «βάρος νεογέννητου στις ΗΠΑ», για το ίδιο δείγμα των 10.000. Παρατηρούμε ότι η μορφή του είναι σχεδόν συμμετρική ως προς μία κατακόρυφη ευθεία, η οποία είναι κάθετη στον άξονα των βαρών, κοντά στα 3,2 κιλά. Από τη μορφή του πολυγώνου συχνοτήτων καταλαβαίνουμε ότι:

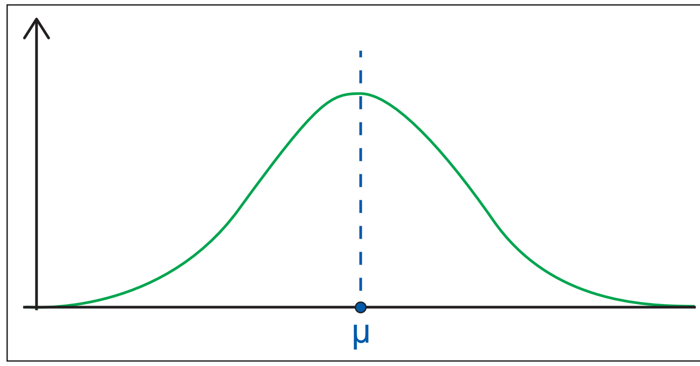
- σχεδόν το 50% των 10.000 νεογέννητων έχει βάρος μικρότερο των 3,2 κιλών και σχεδόν το 50% έχει βάρος μεγαλύτερο των 3,2 κιλών,
- η κατανομή των βαρών «γύρω από τα 3,2 κιλά» γίνεται με συμμετρικό τρόπο,
- καθώς κατευθυνόμαστε από τις ακραίες τιμές βάρους προς τα 3,2 κιλά (δηλαδή προς το «κέντρο» της κατανομής), τα νεογέννητα που κατανέμονται στις κλάσεις είναι περισσότερα.

Στο ιστόγραμμα σχετικών συχνοτήτων επί τοις εκατό (%) που ακολουθεί, παριστάνεται η κατανομή 10.000 ενηλίκων, ως προς τη μέτρηση της διαστολικής αρτηριακής τους πίεσης σε mmHg (χιλιοστά στήλης υδραργύρου), σε κατάσταση ηρεμίας. Κι εδώ, η κατανομή γίνεται σε κλάσεις ίσου πλάτους 5 mmHg. Τα ιστογράμματα των δύο παραδειγμάτων (βάρους νεογέννητων και αρτηριακής πίεσης ενηλίκων) έχουν παρόμοια μορφή.

Η ομοιότητα των δυο προηγούμενων κατανομών φαίνεται, επίσης, και στα πολύγωνα σχετικών συχνοτήτων (%). Στην περίπτωση της μεταβλητής «διαστολική αρτηριακή πίεση ενηλίκων στην κεντρική Ευρώπη», στο δείγμα εμφανίζεται παρόμοια συμμετρία ως προς την ευθεία που είναι κάθετη στον άξονα των τιμών πίεσης, στα 85 mmHg περίπου. Και εδώ, μπορούμε να κάνουμε παρόμοιες παρατηρήσεις με εκείνες του παραδείγματος με τα βάρη νεογέννητων.

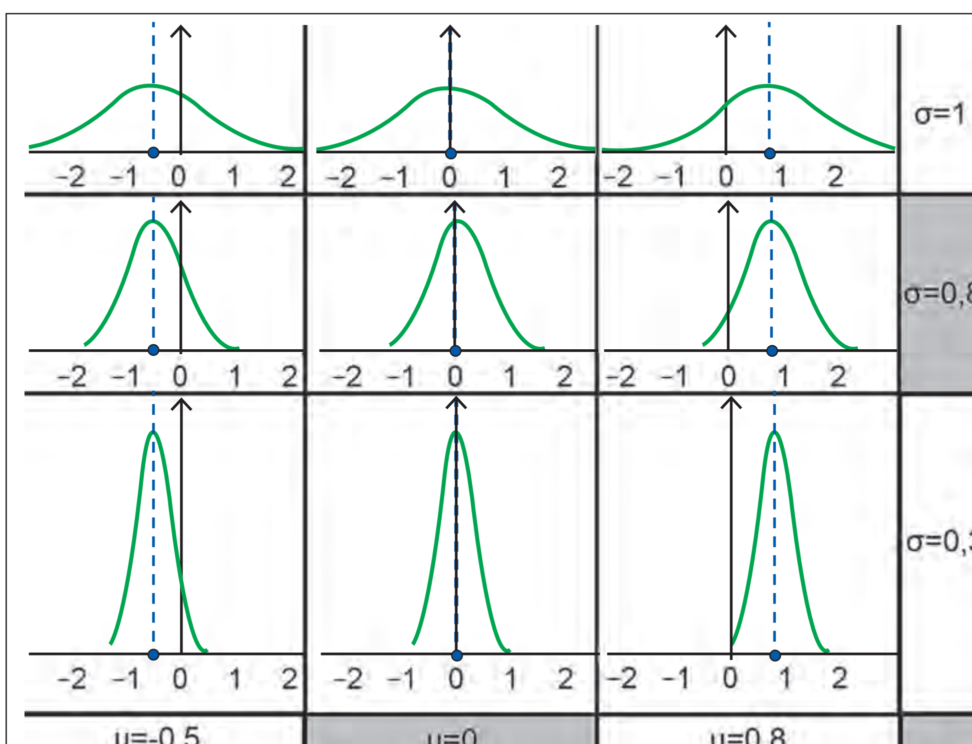
Όμοια μορφή έχουν και τα πολύγωνα σχετικών συχνοτήτων των παραπάνω παραδειγμάτων, σχεδιασμένα σε κατάλληλη κλίμακα (καθώς το 100% αντιστοιχεί στο 1).

Στη φύση συχνά συναντούμε κατανομές συχνοτήτων μεγάλων δειγμάτων, των οποίων τα ιστογράμματα και τα πολύγωνα σχετικών συχνοτήτων έχουν παρόμοια μορφή και ιδιότητες με τα δύο προηγούμενα παραδείγματα. Σε τέτοιες περιπτώσεις, όταν το πλήθος είναι αρκετά μεγάλο και το πλήθος των κλάσεων κατάλληλα μεγάλο, το πολύγωνο σχετικών συχνοτήτων τείνει να μοιάσει με τη γραφική παράσταση της συνάρτησης $y = \frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}}$, για κάποιες τιμές των παραμέτρων μ και σ (η σ είναι θετική). Αυτή τη γραφική παράσταση την ονομάζουμε γκαουσιανή καμπύλη και φαίνεται στο παρακάτω σχήμα.



Η ευθεία x = μ είναι ο άξονας συμμετρίας της γραφικής παράστασης της συνάρτησης. Συνεπώς, μιλώντας κάπως πρακτικά, η τιμή του σ καθορίζει πόσο «απλωμένη» ή «μαζεμένη» είναι η καμπύλη γύρω από το μ και ποιο είναι το μέγιστο «ύψος» της. Η τιμή του μ επηρεάζει τη «θέση» της καμπύλης καθώς καθορίζει τη θέση του άξονα συμμετρίας της. Όταν το σ μεγαλώνει, το σχήμα της μοιάζει να είναι πιο «απλωμένο» πάνω από τον οριζόντιο άξονα και το μέγιστο «ύψος» της καμπύλης μικραίνει. Τα αντίστροφα συμβαίνουν όταν το σ μικραίνει. Ο ρόλος των τιμών των παραμέτρων μ και σ στη μορφή της γραφικής παράστασης φαίνεται και στα παρακάτω παραδείγματα.

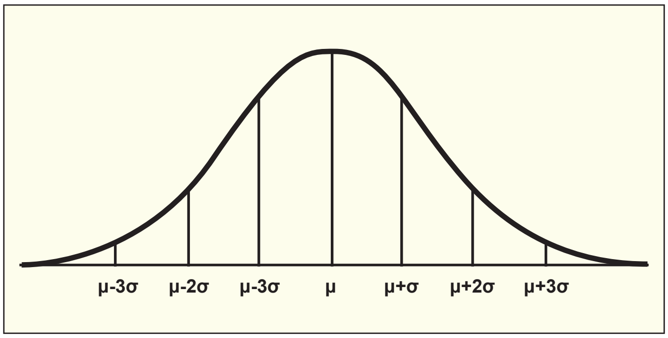
$$\boxed{y = \frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}}}$$



Η κανονική κατανομή είναι το μαθηματικό μοντέλο που χρησιμοποιούμε για να περιγράψουμε πώς κατανέμονται σε έναν ιδεατό, άπειρο πληθυσμό οι τιμές μεταβλητών, όπως εκείνες στα δύο παραδείγματά μας. Για την κανονική κατανομή, το μ εκφράζει το κέντρο συμμετρίας της και το σ είναι ένας δείκτης διασποράς της.

Έστω μία μεταβλητή που μοντελοποιείται από την κανονική κατανομή με μ και σ. Αν επιλέξουμε τυχαία ένα στοιχείο/άτομο από τον άπειρο πληθυσμό και κοιτάξουμε την τιμή της μεταβλητής για αυτό το στοιχείο/άτομο, τότε η πιθανότητα αυτή η τιμή να βρίσκεται σε ένα διάστημα (α,β) ισούται με το εμβαδόν μεταξύ της γκαουσιανής καμπύλης και του άξονα x΄x ανάμεσα στα α και β. Αποδεικνύεται ότι:

- η πιθανότητα του ενδεχομένου «η τιμή της μεταβλητής είναι μεγαλύτερη από μ» ισούται με 0,5 και, λόγω συμμετρίας, η πιθανότητα του ενδεχομένου «η τιμή της μεταβλητής είναι μικρότερη από μ» ισούται, επίσης με 0,5.
- η πιθανότητα του ενδεχομένου «η τιμή της μεταβλητής είναι στο διάστημα (μ - σ, μ + σ) » ισούται κατά προσέγγιση με 0,68 ή 68%.
- η πιθανότητα του ενδεχομένου «η τιμή της μεταβλητής είναι στο διάστημα (μ - 2σ , μ + 2σ) » ισούται κατά προσέγγιση με 0,95 ή 95%
- η πιθανότητα του ενδεχομένου «η τιμή της μεταβλητής είναι στο διάστημα (μ - 3σ , μ + 3σ)» ισούται κατά προσέγγιση με 0,997 ή 99,7%.



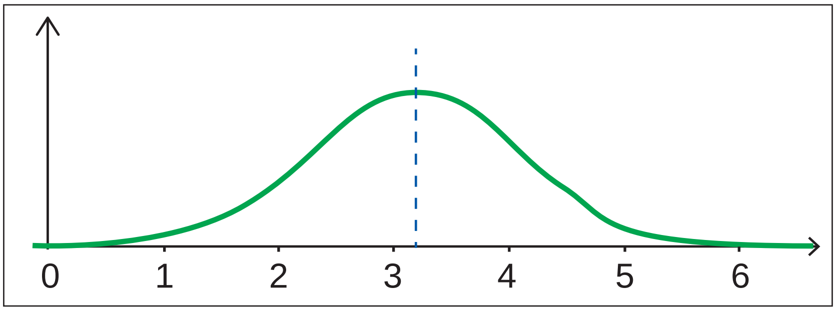
Από τα προηγούμενα και λόγω συμμετρίας της κανονικής κατανομής, υπολογίζουμε τις πιθανότητες και άλλων ενδεχομένων. Π.χ. η πιθανότητα του ενδεχομένου «η τιμή της μεταβλητής είναι στο διάστημα (μ - σ , μ)» ισούται κατά προσέγγιση με 0,68 : 2 = 0,34.

Λόγω στατιστικής ομαλότητας, σε ένα μεγάλο δείγμα (ή σε ολόκληρο τον υπαρκτό πληθυσμό) μπορούμε να εκτιμήσουμε ότι:

- το ποσοστό των στοιχείων/ατόμων για τα οποία η τιμή της μεταβλητής είναι στο διάστημα (μ - σ , μ + σ), είναι περίπου το 68% του πληθυσμού
- το ποσοστό των στοιχείων/ατόμων για τα οποία η τιμή της μεταβλητής είναι στο διάστημα (μ - 2σ , μ + 2σ), είναι περίπου το 95% του πληθυσμού
- το ποσοστό των στοιχείων/ατόμων για τα οποία η τιμή της μεταβλητής είναι στο διάστημα (μ - 3σ , μ + 3σ), είναι περίπου το 99,7% του πληθυσμού.

Λόγω συμμετρίας μπορούμε να κάνουμε και άλλες παρόμοιες εκτιμήσεις.

Επίσης, στην περίπτωση μεγάλου δείγματος (ή ολόκληρου του υπαρκτού πληθυσμού) η δειγματική μέση τιμή της μεταβλητής είναι κοντά στο μ και η δειγματική τυπική απόκλισή της είναι κοντά στο σ. Για παράδειγμα, στην περίπτωση του δείγματος των νεογέννητων, η μεταβλητή «βάρος νεογέννητου στις ΗΠΑ» μοντελοποιείται από την κανονική κατανομή με $\mu = 3{,}18$ και $\sigma = 0{,}91$ (βλέπε το παρακάτω σχήμα). Η μέση τιμή του βάρος των νεογέννητων του δείγματος είναι κοντά στο 3,18 και η τυπική απόκλισή του είναι κοντά στο 0,91. Το ίδιο ισχύει και για τη μέση τιμή και την τυπική απόκλιση του βάρους ολόκληρου του υπαρκτού πληθυσμού των νεογέννητων στις ΗΠΑ.



Εφαρμογές

Εφαρμογή 1

Ο πόρτες που κατασκευάζει μια εταιρία είναι τυποποιημένες με ύψος 183 cm. Θεωρούμε ότι τα ύψη των ενηλίκων στην Ελλάδα ακολουθούν κανονική κατανομή με μέση τιμή 171 cm και τυπική απόκλιση 6 cm:

1) Παίρνουμε ένα πολύ μεγάλο δείγμα ενηλίκων στην Ελλάδα. Να εκτιμήσετε το ποσοστό των ενήλικων του δείγματος που είναι ψηλότεροι από την πόρτα.

2) Να βρείτε ποιο πρέπει να είναι το ύψος της πόρτας, ώστε αν επιλέξουμε τυχαία έναν ενήλικα στην Ελλάδα, η πιθανότητα να είναι ψηλότερος/η από την πόρτα να είναι περίπου 0,15%;

Λύση

1) Τα ύψη των ενήλικων ανθρώπων ακολουθούν κανονική κατανομή με μ=171cm και σ = 6 cm. Ισχύουν:

$$\mu - 2\sigma = 171 - 2\cdot 6 = 159 \text{ και } \mu + 2\sigma = 171+2\cdot 6 = 183$$

Παίρνουμε ένα πολύ μεγάλο δείγμα ενηλίκων στην Ελλάδα. Θέλουμε να εκτιμήσουμε το ποσοστό των ατόμων του δείγματος που είναι ψηλότερα από την πόρτα, δηλαδή από 183 cm.

Το ποσοστό των ατόμων του δείγματος που το ύψος τους (σε cm) ανήκει στο διάστημα (159 , 183) εκτιμάται σε 95%.

Λόγω συμμετρίας της κανονικής κατανομής εκτιμούμε:

- Το ποσοστό των ατόμων του δείγματος που το ύψος τους ανήκει στο διάστημα (171 , 183) σε 95 : 2 = 47,5 %.
- Το ποσοστό των ατόμων του δείγματος που είναι ψηλότερα από 171 cm σε 50%.

Επομένως, το ποσοστό των ατόμων του δείγματος που είναι ψηλότερα από 183 cm, άρα και από την πόρτα, εκτιμάται σε 50% - 47,7% = 2,5%

2) Υποθέτουμε ότι η πόρτα έχει ύψος υ cm.

Έστω ένας τυχαίος ενήλικας και το ενδεχόμενο «ο τυχαίος ενήλικας έχει ύψος μεγαλύτερο από υ cm». Θέλουμε να βρούμε την τιμή του υ, ώστε η πιθανότητα του ενδεχομένου να είναι περίπου 0,15%.

Παρατηρούμε ότι 0,15% = 50% - 49,85% .

Επίσης ισχύει ότι $\mu+3\sigma = 171+3\cdot 6 = 189$.

Η πιθανότητα ο τυχαίος ενήλικας να είναι ψηλότερος από 171 cm είναι 50% και η πιθανότητα να έχει ύψος (σε εκατοστά) που ανήκει στο διάστημα $(\mu, \mu+3\sigma)=(171,189)$ είναι περίπου ίση με $\frac{99,7\%}{2} = 49,85\%$, λόγω συμμετρίας της κανονικής κατανομής.

Άρα, η πιθανότητα ο τυχαίος ενήλικας να είναι ψηλότερος από 189 cm είναι περίπου 50% - 49,85% = 0,15%.

Άρα η πόρτα πρέπει να έχει ύψος υ = 189 cm.

Ασκήσεις – Προβλήματα – Δραστηριότητες

1) Το σύνολο των μαθητών/τριών μιας πόλης, ρωτήθηκαν για τον χρόνο που κάνουν να πάνε από το σπίτι στο σχολείο. Το 50% περίπου έδωσε απαντήσεις που ήταν από 12 λεπτά και πάνω, ενώ το 16% περίπου έδωσε απαντήσεις που ήταν από 10 λεπτά και κάτω. Υποθέτουμε ότι η κατανομή του χρόνου της διαδρομής σπίτι-σχολείο των μαθητών είναι κανονική.

1) Να εκτιμήσετε τον μέσο χρόνο διαδρομής σπίτι-σχολείο, των μαθητών και την τυπική απόκλιση του χρόνου διαδρομής τους, σύμφωνα με τις απαντήσεις τους.

2) Αν οι μαθητές/τριες της πόλης είναι 4.000, να εκτιμήσετε πόσοι/ες απάντησαν ότι έχουν χρόνο διαδρομής σπίτι-σχολείο μεταξύ 14 και 16 λεπτών;

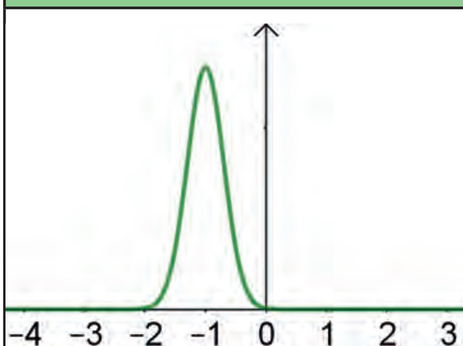
2) Υποθέτουμε ότι το βάρος των μαθητών λυκείου ακολουθεί κανονική κατανομή και παίρνουμε ένα μεγάλο δείγμα μαθητών λυκείου. Το 50% των μαθητών του δείγματος έχουν βάρος το πολύ 65 Kg, ενώ περίπου το 47,5% αυτών έχουν βάρος από 65 Kg έως 75 Kg.

1) Να εκτιμήσετε τη μέση τιμή και την τυπική απόκλιση του βάρους των μαθητών του δείγματος.

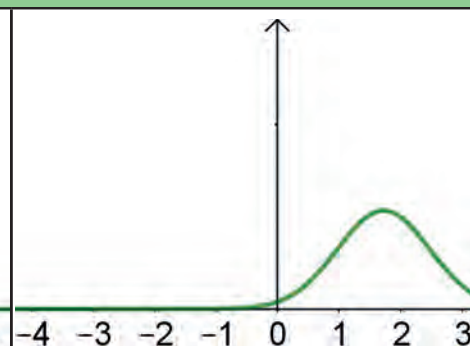
2) Να εκτιμήσετε το ποσοστό των μαθητών του δείγματος που έχουν βάρος από 55 Kg έως 70 Kg.

3) 1) Να αντιστοιχίσετε τις γραφικές παραστάσεις των συνάρτησεων που είναι μοντέλα κανονικών κατανομών και περιέχονται στον παρακάτω πίνακα με τα ζεύγη τιμών των παραμέτρων σ και μ που ακολουθούν. Ο κατακόρυφος άξονας των συστημάτων συντεταγμένων ακολουθεί την ίδια κλίμακα σε όλες τις περιπτώσεις.

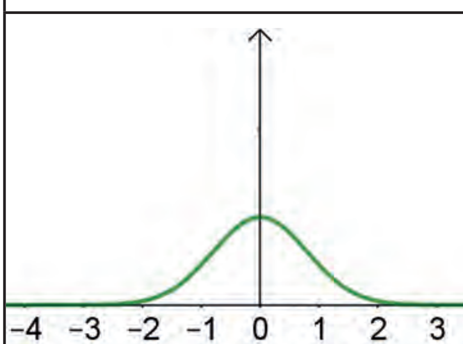
Πίνακας



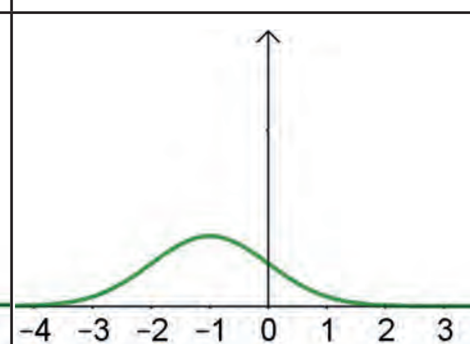
(α)



(β)



(γ)



(δ)

Α. μ = -1, σ = 1 Β. μ = -1, σ = 0,3 Γ. μ >0, σ = 0,75

Δ. μ = 0, σ < 1

2) Να συγκρίνετε την τιμή του σ στο σχήμα (γ) με το 0,3.

Πρόσθετο Υλικό

1) Η κανονική κατανομή ανακαλύφθηκε το 1720 από τον μαθηματικό Abraham de Moivre, στην προσπάθειά του να λύσει προβλήματα παιγνίων τύχης. Εκατόν πενήντα χρόνια αργότερα περί το 1870 ο Adolph Quetelet, Βέλγος μαθηματικός, χρησιμοποιεί την καμπύλη της κανονικής κατανομής ως το ιδεώδες οριακό ιστόγραμμα - πρότυπο, προς το οποίο συγκρίνονται τα ιστογράμματα δεδομένων.

   Με τη χρήση κατάλληλου εκπαιδευτικού λογισμικού να σχεδιάσετε την καμπύλη (δηλ. τη γραφική παράσταση) της συνάρτησης $f(x) = \dfrac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}}$, για διάφορες τιμές των φυσικών αριθμών μ και σ και να κάνετε παρατηρήσεις για τη μορφή της.