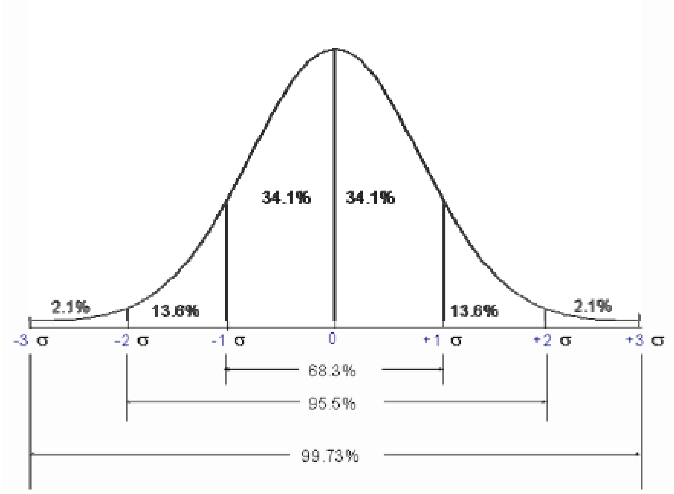
<u>*Η Κανονική Κατανομή*</u>

Η ***κανονική κατανομή (normal distribution)*** θεωρείται η σπουδαιότερη κατανομή της *Θεωρίας Πιθανοτήτων* και της *Στατιστικής*. Οι λόγοι που εξηγούν την εξέχουσα θέση της, είναι βασικά δύο:

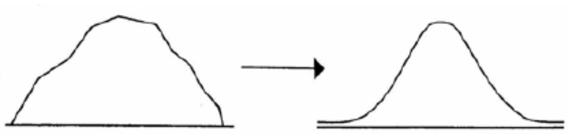
i) *Πολλές τυχαίες μεταβλητές περιγράφονται ικανοποιητικά από την κανονική κατανομή ή περιγράφονται από κατανομές που μπορούν να προσεγγισθούν από την κανονική κατανομή.*

ii) *Οι ιδιότητες της κανονικής κατανομής αξιοποιούνται στη Στατιστική Συμπερασμασματολογία. Ουσιαστικά, η κανονική κατανομή, αποτελεί το θεμέλιο της Στατιστικής Συμπερασμασματολογίας.*

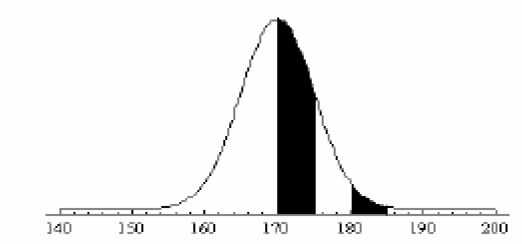


Ιδιότητες της κανονικής καμπύλης

Στην *κανονική καμπύλη* έχουμε ήδη αναφερθεί. Όπως όλες οι καμπύλες συχνοτήτων, προκύπτει ως προσέγγιση του *πολυγώνου συχνοτήτων* των τιμών μιας συνεχούς μεταβλητής. Αυξάνοντας, δηλαδή, το μέγεθος του δείγματος και κατασκευάζοντας το ιστόγραμμα με ολοένα και μικρότερου πλάτους κλάσεις ($c \to 0$), το αντίστοιχο *πολύγωνο* προσεγγίζει μια ομαλή-λεία καμπύλη.

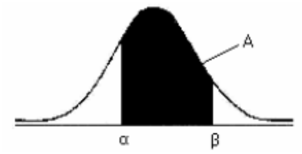


Η *κανονική καμπύλη* έχει κωδωνοειδή μορφή, είναι συμμετρική και οι «ουρές» της πλησιάζουν τον οριζόντιο άξονα ομαλά (ασυμπτωτικά). Η *μέση τιμή* και η *διάμεσος* ταυτίζονται. Επίσης, η *κορυφή* ταυτίζεται με τη *μέση τιμή* και τη *διάμεσο*. Έτσι, η περιοχή που παρουσιάζει τη μεγαλύτερη <u>*πυκνότητα*</u>, βρίσκεται και αυτή στο μέσο της κατανομής. Δηλαδή, όταν οι τιμές μιας μεταβλητής είναι κανονικά κατανεμημένες, τότε γύρω από τη *μέση τιμή* τους υπάρχουν σχετικά πολλές τιμές ενώ μακριά από τη *μέση τιμή* βρίσκονται σχετικά λίγες τιμές. Για παράδειγμα, αν το ύψος των ελλήνων, ηλικίας 18 έως 25 ετών, είναι κανονικά κατανεμημένο, με *μέση τιμή* 170 cm και *τυπική απόκλιση* 5 cm, τότε μεταξύ 170 cm και 175 cm βρίσκονται περισσότερα άτομα από όσα βρίσκονται μεταξύ 180 cm και 185 cm. Επίσης, πολύ λίγα άτομα έχουν ύψος μεγαλύτερο από 185 cm ή μικρότερο από 155 cm.

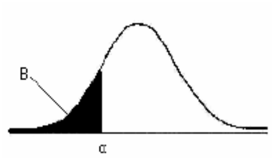


Το εμβαδόν του χωρίου που περικλείεται από την καμπύλη της *συνάρτησης πυκνότητας* και τον άξονα των τιμών της X είναι ίσο με **1** και εκφράζει την πιθανότητα η X να πάρει κάποια τιμή μεταξύ $-\infty$ και $+\infty$. Ανάλογα,

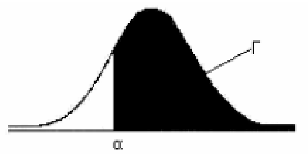
- το εμβαδόν του σκιαγραφημένου χωρίου Α στο επόμενο σχήμα, εκφράζει την πιθανότητα η X να πάρει κάποια τιμή μεταξύ των τιμών α και β, δηλαδή, $A = P(\alpha \leq X \leq \beta)$.



- το εμβαδόν του σκιαγραφημένου χωρίου Β στο επόμενο σχήμα, εκφράζει την πιθανότητα η X να πάρει κάποια τιμή μικρότερη ή ίση του α, δηλαδή, $B = P(X \leq \alpha)$.



- το εμβαδόν του σκιαγραφημένου χωρίου Γ στο επόμενο σχήμα, εκφράζει την πιθανότητα η X να πάρει κάποια τιμή μεγαλύτερη ή ίση του α, δηλαδή, $\Gamma = P(X \geq a)$.



Επισήμανση

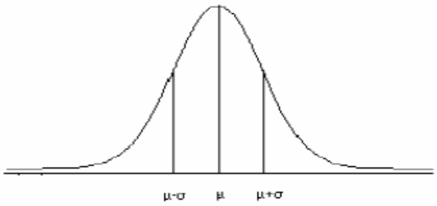
Πρέπει να επισημάνουμε ότι η τιμή $f(x)$ της *συνάρτησης πυκνότητας* για συγκεκριμένη τιμή x της μεταβλητής X, <u>δεν αντιστοιχεί σε πιθανότητα</u>, δηλαδή, <u>δεν ισχύει</u> $f(x) = P(X = x)$. Εξάλλου, στις <u>συνεχείς</u> μεταβλητές, η πιθανότητα $P(X = x)$ είναι μηδέν[3]. Τι εκφράζει επομένως η $f(x)$; Η $f(x)$ εκφράζει <u>*πυκνότητα*</u>,

δηλαδή, όσο μεγαλύτερη είναι η τιμή $f(x)$ τόσο περισσότερο πιθανό είναι να πάρει η μεταβλητή X τιμές κοντά στο x.

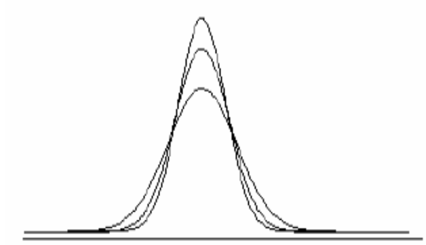
[3] *Αυτός είναι και ο λόγος που στις συνεχείς μεταβλητές έχουμε:* $P(X \leq \alpha) = P(X < \alpha)$, $P(X \geq \alpha) = P(X > \alpha)$ *και* $P(\alpha \leq X \leq \beta) = P(\alpha < X < \beta) = P(\alpha \leq X < \beta) =$
$= P(\alpha < X \leq \beta)$

Παρατήρηση

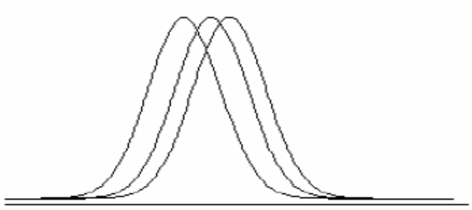
Παρατηρείστε ότι η καμπύλη της συνάρτησης πυκνότητας της κανονικής κατανομής, στη θέση $x=\mu$ *παρουσιάζει μέγιστη τιμή* (ίση με $\frac{1}{\sigma\sqrt{2\pi}}$) *και στις θέσεις* $x=\mu-\sigma$ *και* $x=\mu+\sigma$ *παρουσιάζει σημεία καμπής.*



Είναι φανερό, ότι η *συνάρτηση πυκνότητας* της *κανονικής κατανομής* δεν ορίζει μια συγκεκριμένη κανονική καμπύλη αλλά μια *οικογένεια* κανονικών καμπύλων. Έτσι, για διαφορετικές τιμές των παραμέτρων μ και σ παίρνουμε διαφορετικές κανονικές καμπύλες. Για παράδειγμα, οι *κατανομές,*



είναι όλες *κανονικές κατανομές,* με ίδια *μέση τιμή* και διαφορετικές *τυπικές αποκλίσεις*. Επίσης, οι *κατανομές,*



είναι όλες *κανονικές κατανομές* με ίδιες *τυπικές αποκλίσεις* και διαφορετικές *μέσες τιμές*.

Είναι φανερό, ότι αλλαγή της *μέσης τιμής* προκαλεί μόνο μετατόπιση της *κανονικής καμπύλης* σε μια νέα θέση. Αλλαγή, της *τυπικής απόκλισης*, όμως, προκαλεί αλλαγή στην *κανονική καμπύλη (χωρίς, φυσικά να αλλάζει η κωδωνοειδής μορφή της)*. Για παράδειγμα, όσο μικρότερη είναι η *τυπική απόκλιση*, τόσο ψηλότερη και τόσο πιο στενή είναι η *κανονική καμπύλη*. Δηλαδή, τόσο μικρότερο είναι το διάστημα στο οποίο, πρακτικά, εκτείνεται η κατανομή.

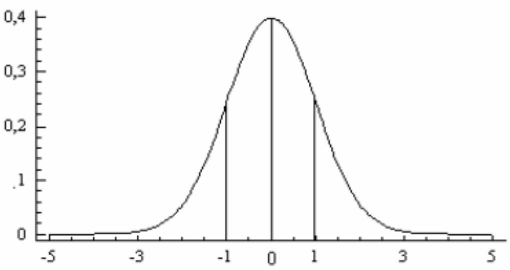
Επισημαίνουμε ότι οι παράμετροι μ και σ <u>χαρακτηρίζουν</u> την *κανονική κατανομή*, δηλαδή, μπορούμε να την προσδιορίσουμε πλήρως αν γνωρίζουμε μόνο τη *μέση τιμή της,* μ και την *τυπική απόκλισή της,* σ. Η *κανονική κατανομή* με *μέση τιμή* μ και *διασπορά* σ^2 (δηλαδή *τυπική απόκλιση* σ) συμβολίζεται με $N(\mu,\sigma^2)$.

Η Τυποποιημένη κανονική Κατανομή

Η *κανονική κατανομή* που έχει *μέση τιμή* **0** και *τυπική απόκλιση* **1** (άρα και *διασπορά* **1**), συμβολίζεται με $N(0,1)$ και ονομάζεται ***τυποποιημένη κανονική κατανομή (standard normal distribution)***. Μια τυχαία μεταβλητή που ακολουθεί την *τυποποιημένη κανονική κατανομή*, έχει επικρατήσει να συμβολίζεται με Z και η συνάρτηση πυκνότητάς της με $\varphi(z)$. Προφανώς είναι:

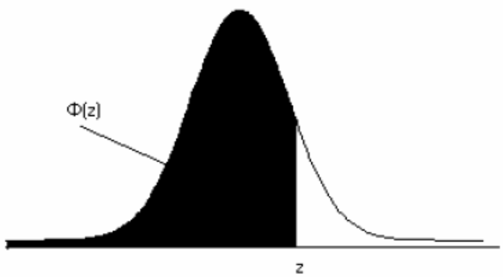
$$\varphi(z) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{z^2}{2}},\ -\infty < z < +\infty.$$

Σύμφωνα με τα προηγούμενα, η καμπύλη της *τυποποιημένης κανονικής κατανομής* στη θέση $x = 0$ παρουσιάζει *μέγιστη τιμή* (ίση με $\frac{1}{\sqrt{2\pi}} = 0.4$) και στις θέσεις $x = -1$ και $x = 1$ παρουσιάζει *σημεία καμπής.*



Υπολογισμός πιθανοτήτων

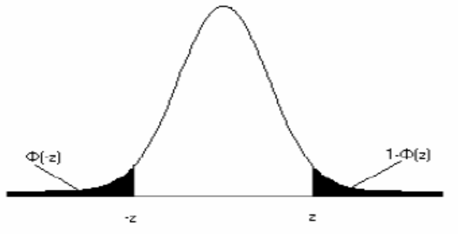
Σύμφωνα με όσα ήδη έχουμε αναφέρει, ο υπολογισμός πιθανοτήτων, ανάγεται στον υπολογισμό εμβαδών επιπέδων χωρίων. Δυστυχώς, καμία από τις γνωστές τεχνικές ολοκλήρωσης δε μας επιτρέπει τον αναλυτικό υπολογισμό του κατάλληλου, κατά περίπτωση, ορισμένου ολοκληρώματος της $f(x)$. Στην πράξη, για να υπολογίσουμε τις πιθανότητες που αφορούν τις τιμές τυχαίας μεταβλητής που ακολουθεί *κανονική κατανομή* $N(\mu,\sigma^2)$, χρησιμοποιούμε τον ***πίνακα της τυποποιημένης κανονικής κατανομής*** $N(0,1)$. Ο *πίνακας της τυποποιημένης κανονικής κατανομής*[4], μας δίνει την πιθανότητα $P(Z \le z)$ για όλα τα z από 0 έως 3,59 με βήμα 0,01. Ας συμβολίσουμε αυτή την πιθανότητα με $\Phi(z)$. Δηλαδή, $\Phi(z) = P(Z \le z)$. Ο πίνακας, επομένως, της *τυποποιημένης κανονικής κατανομής* μας δίνει το εμβαδόν του σκιαγραφημένου χωρίου το οποίο συμβολίζεται με $\Phi(z)$.



Εύκολα μπορεί να αποδειχθεί ότι:

- $\Phi(-z) = 1 - \Phi(z)$

Άρα $P(Z \le -z) = \Phi(-z) = 1 - \Phi(z)$



Σημείωση

Η ιδιότητα αυτή εξηγεί γιατί ο πίνακας της τυποποιημένης κανονικής κατανομής δίνει τις τιμές της $\Phi(z)$ μόνο για μη αρνητικά z.

- $P(\alpha \le Z \le \beta) = \Phi(\beta) - \Phi(\alpha)$
- $P(-\alpha \le Z \le \alpha) = \Phi(\alpha) - \Phi(-\alpha) = 2 \cdot \Phi(\alpha) - 1$
- $P(Z > a) = 1 - P(Z \le \alpha) = 1 - \Phi(\alpha)$.

Είναι φανερό, ότι μπορούμε πλέον, να υπολογίσουμε οποιαδήποτε πιθανότητα για τη Z με βάση μόνο τις τιμές $\Phi(z)$ του πίνακα *της τυποποιημένης κανονικής κατανομής*. Ας δούμε μερικά παραδείγματα:

$$P(Z \le 0) = \Phi(0) = 0.5$$
$$P(Z \le 1{,}37) = \Phi(1.37) = 0.9147$$
$$P(Z > 1.37) = 1 - P(Z \le 1.37) = 1 - \Phi(1.37) = 1 - 0.9147 = 0.0853$$
$$P(Z \le -1.55) = \Phi(-1.55) = 1 - \Phi(1.55) = 1 - 0.9394 = 0.606$$
$$P(-1.55 \le Z \le 2.1) = \Phi(2.1) - \Phi(-1.55) = \Phi(2.1) - [1 - \Phi(1.55)] =$$
$$= \Phi(2.1) - 1 + \Phi(1.55) = 0.9821 - 1 + 0.9394 = 0.9215$$
$$P(-1 \le Z \le 1) = 2 \cdot \Phi(1) - 1 = 2 \cdot 0.8413 - 1 = 0.6826 \cong 68.3\%$$
$$P(-2 \le Z \le 2) = 2 \cdot \Phi(2) - 1 = 2 \cdot 0.9772 - 1 = 0.9544 \cong 95.5\%$$
$$P(-3 \le Z \le 3) = 2 \cdot \Phi(3) - 1 = 2 \cdot 0.9987 - 1 = 0.9974 \cong 99.7\%$$

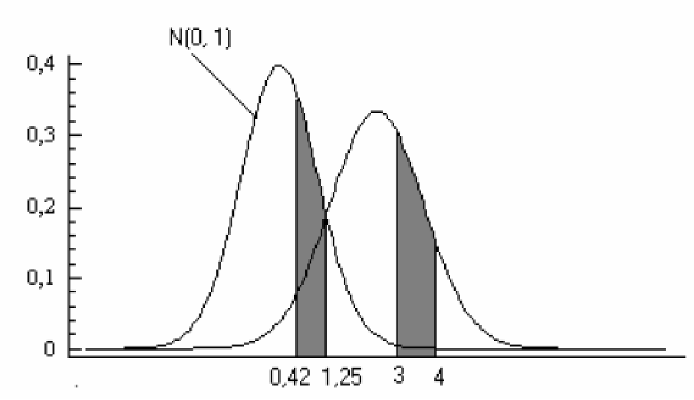
Όπως, ήδη, έχουμε αναφέρει, μέσω του πίνακα *της τυποποιημένης κανονικής κατανομής*, μπορούμε να υπολογίσουμε πιθανότητες για οποιαδήποτε *κανονική κατανομή* $N(\mu, \sigma^2)$. Αυτό μπορεί να γίνει διότι έχει αποδειχθεί ότι:

Αν η τυχαία μεταβλητή X ακολουθεί την κανονική κατανομή $N(\mu, \sigma^2)$ τότε η τυχαία μεταβλητή $Z = \dfrac{X - \mu}{\sigma}$, ακολουθεί την τυποποιημένη κανονική $N(0,1)$.

Έτσι, αν η τυχαία μεταβλητή X, ακολουθεί *κανονική κατανομή* με $\mu = 2.5$ και $\sigma = 1.2$, η πιθανότητα $P(3 \le X \le 4)$ μπορεί να υπολογισθεί ως εξής:

$$P(3 \le X \le 4) = P(\frac{3-2.5}{1,2} \le \frac{X-2.5}{1,2} \le \frac{4-2.5}{1,2}) = P(0.42 \le Z \le 1.25) =$$
$$= \Phi(1.25) - \Phi(0.42) = 0.8944 - 0.6628 = 0.2316$$

Στο παρακάτω σχήμα φαίνεται ο μετασχηματισμός της $N(2.5, 1.2^2)$ στην $N(0,1)$.



Παράδειγμα 1

Έχει παρατηρηθεί ότι ο χρόνος που χρειάζεται ένα ασθενοφόρο για να φθάσει από ένα κέντρο υγείας, στο πλησιέστερο περιφερειακό νοσοκομείο, ακολουθεί κατά προσέγγιση *κανονική κατανομή* με *μέση τιμή* $\mu = 17\,min$ και *τυπική απόκλιση* $\sigma = 3\,min$. Να βρεθεί η πιθανότητα, ο χρόνος που θα χρειασθεί το ασθενοφόρο για να φθάσει στο περιφερειακό νοσοκομείο,

α) να είναι το πολύ 15 *min*

β) να είναι περισσότερο από 22 *min*

γ) να είναι τουλάχιστον 13 *min* και το πολύ 21 *min*

Απάντηση

α) $P(X \le 15) = P(\frac{X-17}{3} \le \frac{15-17}{3}) = P(Z \le -0.67) = \Phi(-0.67) =$

$= 1 - \Phi(0.67) = 1 - 0.7486 = 0.25$

β) $P(X > 22) = P(\frac{X-17}{3} > \frac{22-17}{3}) = P(Z > 1.67) = 1 - P(Z \le 1.67) =$

$= 1 - \Phi(1.67) = 1 - 0.9525 = 0.0475$

γ) $P(13 \le X \le 21) = P(\frac{13-17}{3} \le \frac{X-17}{3} \le \frac{21-17}{3}) = P(-1.33 \le Z \le 1.33) =$

$= 2 \cdot \Phi(1.33) - 1 = 2 \cdot 0.9082 - 1 = 0.8164$