ΚΕΦΑΛΑΙΟ 5ο

Ειδικές Συνεχείς Κατανομές Πιθανότητας

Εισαγωγή
Η κανονική κατανομή
Η τυποποιημένη κανονική κατανομή
Εύρεση των τιμών X από τις τιμές Z
Δειγματική κατανομή – Η κατανομή του μέσου $\overline{X}$
Τυπικό σφάλμα του μέσου $\overline{X}$
Κεντρικό Οριακό Θεώρημα (Κ.Ο.Θ.)
Η κατανομή $X^2_{(\nu)}$
Η κατανομή Student – t (W. Gosset 1908)
Η δειγματική κατανομή του μέσου $\overline{X}$ για δείγματα από κανονικό πληθυσμό με άγνωστη διακύμανση σ^2
Ιδιότητες της $t_{(\nu)}$ – κατανομής
Πίνακας της $t_{(\nu)}$ – κατανομής
Κανονική προσέγγιση στη Διωνυμική κατανομή

ΚΕΦΑΛΑΙΟ 5ο
ΕΙΔΙΚΕΣ ΣΥΝΕΧΕΙΣ ΚΑΤΑΝΟΜΕΣ ΠΙΘΑΝΟΤΗΤΩΝ

5.1 Εισαγωγή

Στο κεφάλαιο 3 μελετήσαμε αναλυτικά μεταξύ άλλων, τις κατανομές πιθανότητας συνεχών τυχαίων μεταβλητών και δώσαμε τρόπους υπολογισμού των πιθανοτήτων της μορφής $P[X < \alpha]$, $P[X > \alpha]$ και $P[\alpha < X < \beta]$.

Στο κεφάλαιο αυτό θα ασχοληθούμε με ειδικές, συνεχείς κατανομές πιθανοτήτων, αρχίζοντας με την κατανομή που κατέχει την πρώτη θέση στη θεωρία κατανομών, την κανονική κατανομή.

5.2 Η Κανονική Κατανομή

Η κανονική κατανομή ανακαλύφθηκε το 1720 από το 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.

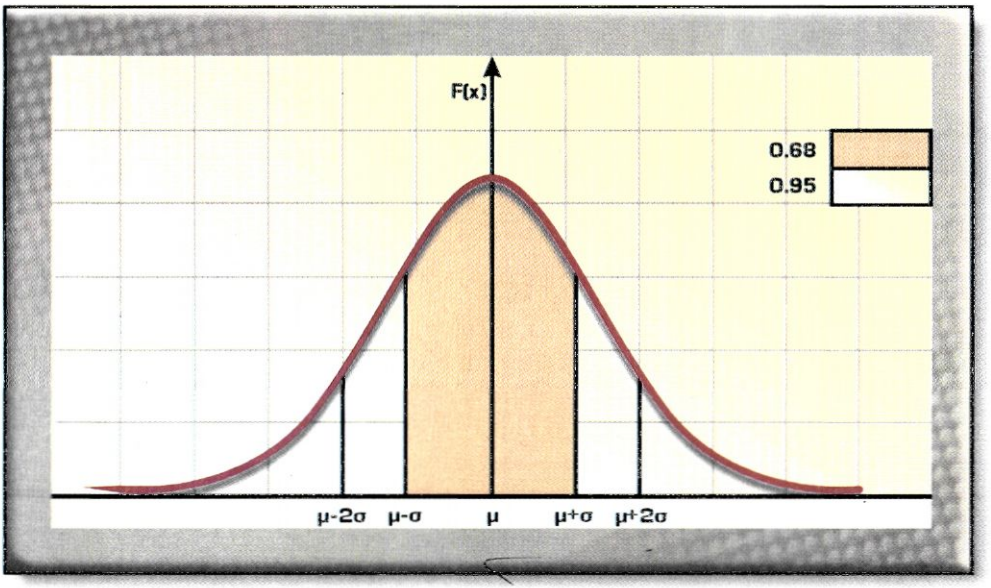
Η συνάρτηση πυκνότητας πιθανότητας της κανονικής κατανομής έχει τη μορφή :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad \text{με } -\infty < x < \infty, \text{ και} \qquad \textbf{(5.1)}$$

$$-\infty < \mu < \infty, \quad 0 < \sigma < \infty$$

στην οποία περιέχονται τρεις από τους πιο γνωστούς αριθμούς στην ιστορία των μαθηματικών, συγκεκριμένα οι: $\sqrt{2}$, π και e. Οι ποσότητες μ και σ είναι παράμετροι της κατανομής, αποτελούν δε το μέσο και την τυπική απόκλιση αντιστοίχως.

Η μορφή της γραφικής παράστασης της f(x) είναι χαρακτηριστικό της ***κανονικής κατανομής*** και δίνεται στο Σχ. (5.1).



Σχ. (5.1) Κανονική καμπύλη (Καμπύλη της κανονικής κατανομής)

Η κανονική κατανομή έχει τις ακόλουθες ιδιότητες:

- **Η κατανομή είναι συμμετρική περί το μ, που είναι ο μέσος όλων των δυνατών τιμών της τ.μ. Χ.**
- **Η επικρατούσα τιμή, η διάμεσος και ο μέσος ταυτίζονται λόγω συμμετρίας της καμπύλης της κατανομής.**
- **Το εύρος των τιμών της τ.μ. Χ είναι το σύνολο R των πραγματικών αριθμών.**
- **Ο οριζόντιος άξονας είναι ασύμπτωτος της καμπύλης, όταν το** $x \to \pm\infty$.
- **Το συνολικό εμβαδόν που περικλείεται από την καμπύλη και τον άξονα των x είναι μοναδιαίο.**

Χρησιμοποιώντας τις ιδιότητες της κανονικής κατανομής και μία σειρά από τεχνικές, μπορούμε να υπολογίσουμε εμβαδά χωρίων μεταξύ της καμπύλης και δύο ευθειών $x=\alpha$ και $x=\beta$.

Από τη μορφή της κανονικής κατανομής προκύπτει ότι **πρόκειται για κατανομή, το σχήμα της οποίας εξαρτάται από δύο παραμέτρους και συγκεκριμένα από την παράμετρο μ που αποτελεί το μέσο της κατανομής και την παράμετρο σ που είναι η τυπική απόκλιση της κατανομής,** όπως έχουμε ήδη αναφέρει. Θα πρέπει στο σημείο αυτό να τονίσουμε τη σημασία που έχουν οι δύο αυτές παράμε-

τροι στη σημαντικότερη κατανομή της στατιστικής, όπως έχει κριθεί ότι είναι η κανονική κατανομή.

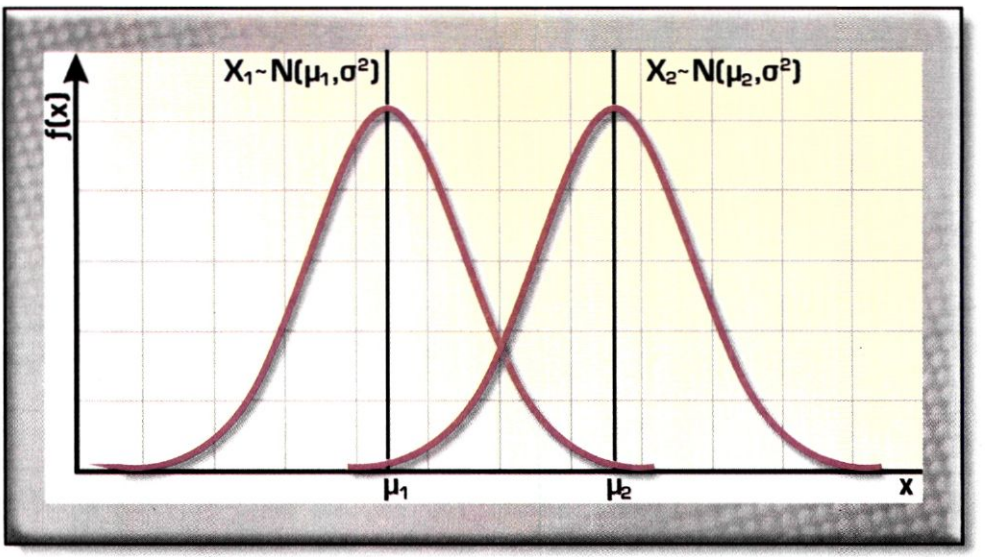
Η σημασία αυτών των παραμέτρων υπογραμμίζεται και από το συμβολισμό **$X \sim N(\mu,\sigma^2)$** που δηλώνει ότι η τ.μ. Χ έχει κανονική κατανομή με μέσο μ και τυπική απόκλιση σ, (ισοδύναμα με διακύμανση σ^2).

Π.χ. αν η τ.μ. Χ έχει κανονική κατανομή με μέσο μ=10 και διακύμανση σ^2=2, γράφουμε: $X \sim N(10,2)$

Αποδεικνύεται ότι αν η τυχαία μεταβλητή Χ ακολουθεί κανονική κατανομή με παραμέτρους μ,σ, τότε

$$\mathbf{E[X] = \mu \text{ και } Var[X] = \sigma^2}$$

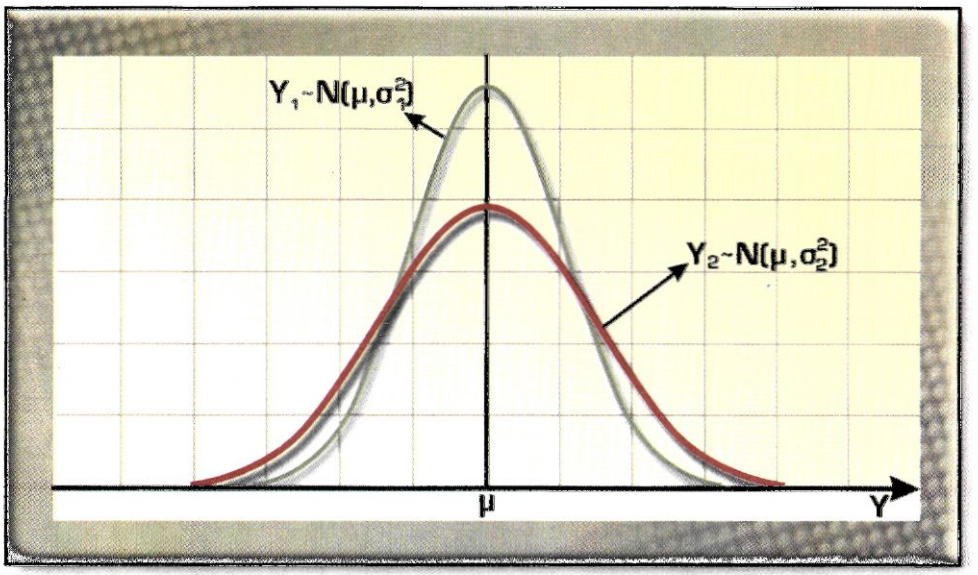
- Αν $X_1 \sim N(\mu_1,\sigma^2)$ και $X_2 \sim N(\mu_2,\sigma^2)$ με $\mu_1<\mu_2$, οι καμπύλες των κατανομών έχουν τη μορφή του Σχ. (5.2).



Σχ. (5.2) Καμπύλες κανονικών κατανομών με διαφορετικούς μέσους $\mu_1<\mu_2$ και ίσες διακυμάνσεις

Όπως φαίνεται στο σχήμα, οι κατανομές που διαφέρουν ως προς τον μέσο μ έχουν διαφορετική θέση στον άξονα των x, γι' αυτό άλλωστε ο μέσος αποτελεί μέτρο θέσης μιας κατανομής.

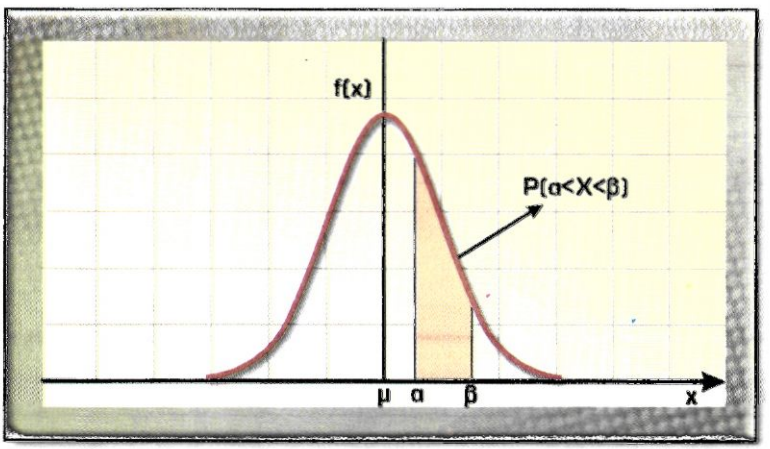
- Αν η $Y_1 \sim N(\mu,\sigma^2_1)$ και $Y_2 \sim N(\mu,\sigma^2_2)$ και $\sigma^2_1<\sigma^2_2$ από το Σχ. (5.3) βλέπουμε ότι οι δύο κατανομές είναι τοποθετημένες στο ίδιο σημείο του άξονα των x (έχουν δηλαδή ίδια μέτρα θέσης), αλλά **διαφορετικούς βαθμούς διασποράς**.



Σχ. (5.3) Καμπύλες κανονικών κατανομών με ίσους μέσους και διαφορετικές διακυμάνσεις $\sigma_1^2 < \sigma_2^2$

Όπως έχουμε αναφέρει η κανονική κατανομή χρησιμοποιείται για τη μελέτη της κατανομής των τιμών συνεχούς μεταβλητής και ως εκ τούτου ο υπολογισμός των πιθανοτήτων ισοδυναμεί με υπολογισμό εμβαδών που ορίζονται από την καμπύλη και από δύο τιμές α και β της μεταβλητής.

Αν λοιπόν $X \sim N(\mu,\sigma^2)$, τότε η πιθανότητα $P(\alpha<X<\beta)$ αντιστοιχεί, ως γνωστό, στο εμβαδόν του κίτρινου τμήματος του Σχ. (5.4).



Σχ. (5.4) Η πιθανότητα $P(\alpha<X<\beta)$*, όταν η τ.μ.* $X \sim N(\mu,\sigma^2)$

Η εύρεση των πιθανοτήτων της μορφής $P(\alpha<X<\beta)$ επιτυγχάνεται με τη χρησιμοποίηση πινάκων που έχουν κατασκευαστεί γι' αυτό ακριβώς τον σκοπό.

Από τους πίνακες αυτούς προκύπτει ότι:

- **Το εμβαδόν που περικλείεται κάτω από την καμπύλη και μεταξύ των αριθμών μ-σ και μ+σ είναι 68% περίπου.**
- **Το εμβαδόν που περικλείεται κάτω από την καμπύλη και μεταξύ των αριθμών μ-2σ και μ+2σ είναι 95% περίπου.**

Τον τρόπο υπολογισμού αυτών των εμβαδών όπως και άλλων θα δούμε αναλυτικά στην παράγραφο 5.3.

5.3 Η Τυποποιημένη Κανονική Κατανομή

Η κατασκευή των πινάκων που αναφέραμε πιο πάνω βασίστηκε στην ακόλουθη σημαντική ιδιότητα των τυχαίων μεταβλητών.

Αν η τ.μ. X έχει μέσο μ και διακύμανση σ^2, τότε η τ.μ. $Z = \frac{X-\mu}{\sigma}$ αποτελεί την τυποποιημένη μορφή της X και έχει μέσο ίσο με 0 και διακύμανση ίση με 1.

Εξειδικεύοντας την ιδιότητα αυτή στην περίπτωση της τ.μ. X για την οποία υποθέτουμε ότι έχει **κανονική κατανομή** μέσου μ και διακύμανσης σ^2, έχουμε το ακόλουθο :

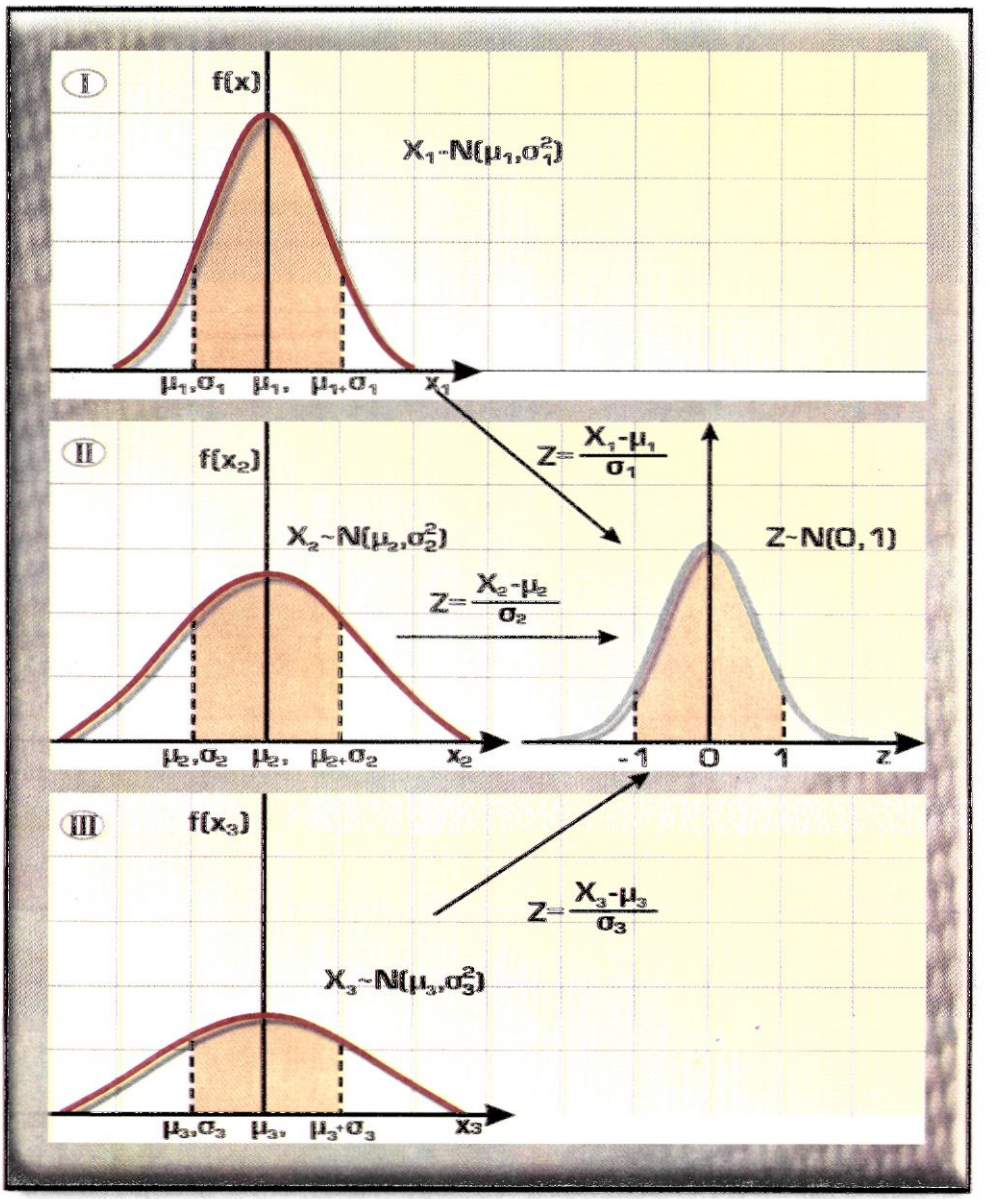
Αν η τ.μ. $X \sim N(\mu,\sigma^2)$, τότε η τ.μ. $Z \sim N(0,1)$

Το σημαντικό αυτού του αποτελέσματος είναι ότι **κάθε κανονική κατανομή μπορεί να μετασχηματιστεί σε κανονική κατανομή μηδενικού μέσου (μ=0) και μοναδιαίας διακύμανσης (σ^2=1)**.

Από την (5.1) για $Z = \frac{X-\mu}{\sigma}$, $\mu = 0$ και $\sigma^2=1$ προκύπτει η

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} \qquad \text{με} \quad -\infty < Z < \infty \qquad \textbf{(5.2)}$$

Παραδείγματα κανονικών κατανομών με διαφορετικούς μέσους και διαφορετικές διακυμάνσεις που έχουν μετασχηματιστεί στην αντίστοιχη τυποποιημένη κανονική κατανομή δίνονται στο Σχ. (5.5).



Σχ. (5.5) Τρεις κανονικές κατανομές διαφορετικών μέσων και διακυμάνσεων και ο μετασχηματισμός τους σε τυποποιημένη κανονική κατανομή

Παραδείγματα:

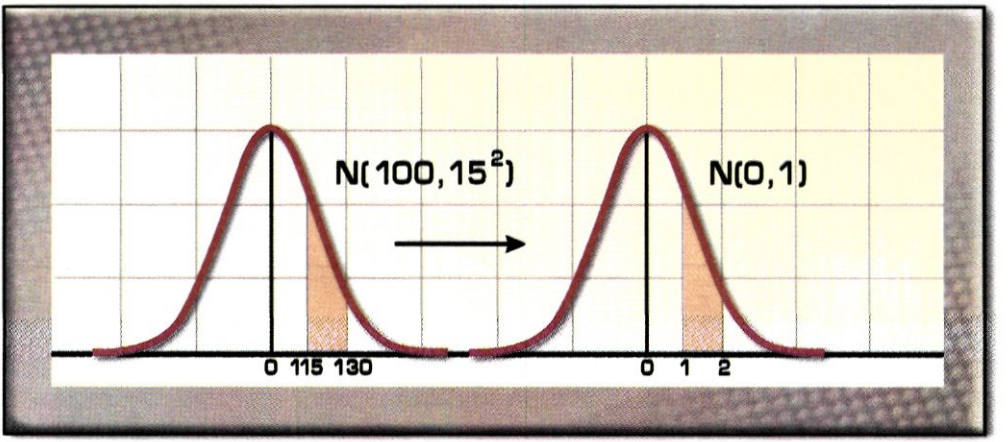
1. Η τ.μ. $X \sim N(100,15^2)$, να υπολογιστεί η πιθανότητα : $P(115<X<130)$

Απάντηση

Όπως ήδη γνωρίζουμε η τ.μ. $Z = \frac{X-\mu}{\sigma} = \frac{X-100}{15} \sim N(0,1)$, οπότε το κιτρινισμένο εμβαδόν της αρχικής κανονικής $N(100,15^2)$ κατανομής ισοδυναμεί με το κιτρινισμένο εμβαδόν της τυποποιημένης κανονικής $N(0,1)$ κατανομής. Άρα :

$$P(115<X<130) = P(\frac{115-100}{15} < \frac{X-100}{15} < \frac{130-115}{15}) = P(1<Z<2)$$

Σχέση από την οποία προκύπτει ότι η πιθανότητα που έχει η τ.μ. $X \sim N(100,15^2)$ να βρίσκεται μεταξύ των τιμών 115 και 130 ισούται με την πιθανότητα που έχει η τ.μ. $Z \sim N(0,1)$ να βρίσκεται μεταξύ των τιμών 1 και 2. (Σχ. 5.6)



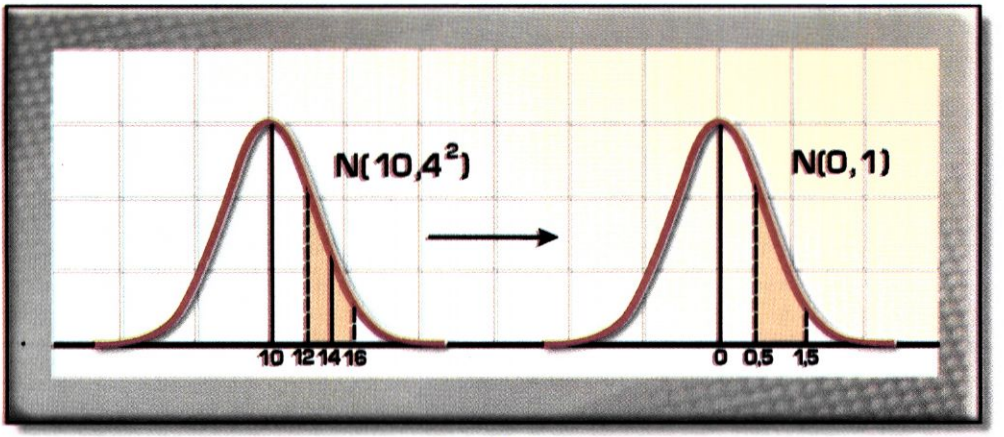
Σχ. (5.6)

2. Η τ.μ. $X \sim N(10,4^2)$ και θέλουμε να βρούμε την $P(12<X<16)$.

Απάντηση

Σύμφωνα με τα παραπάνω έχουμε :

$$P(12 < X < 16) = P(\frac{2-10}{4} < \frac{X-10}{4} < \frac{16-10}{4}) = P(0,5 < Z < 1,5)$$ (Σχ. 5.7)



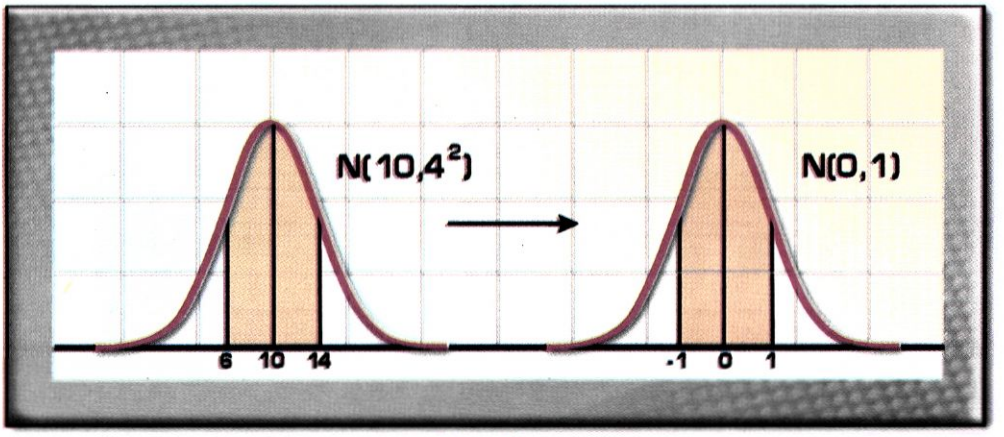
Σχ. (5.7)

3. Η τ.μ. $X \sim N(10,4^2)$ και ζητάμε την πιθανότητα $P(6<X<14)$.

Απάντηση

Κατά τον ίδιο τρόπο θα έχουμε :

$$P(6 < X < 14) = P(\frac{6-10}{4} < \frac{X-10}{4} < \frac{14-10}{4}) = P(-1 < Z < 1) \quad (\Sigma\chi.\ 5.8)$$



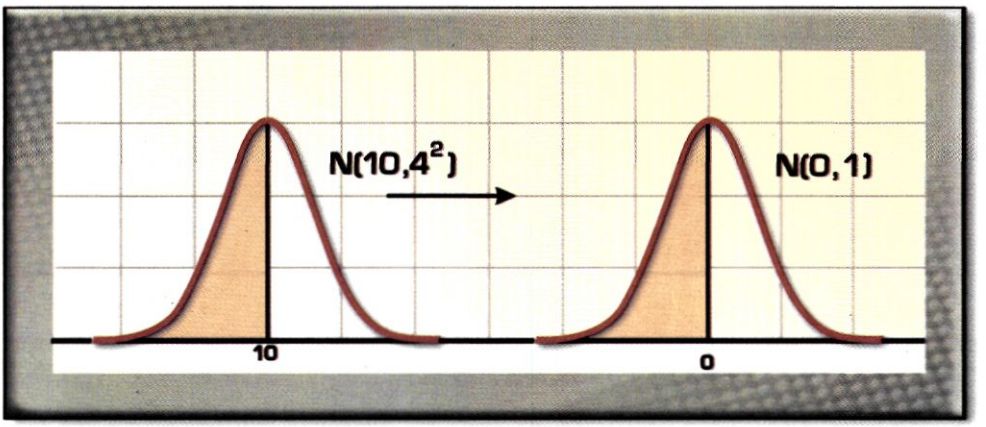
Σχ. (5.8)

4. Η τ.μ. $X \sim N(10,4^2)$ και ζητάμε την πιθανότητα $P(X<10)$.

Απάντηση

Ομοίως έχουμε :

$$P(X<10) = P\left(\frac{X-10}{4} < \frac{10-10}{4}\right) = P(Z<0)$$ (Σχ. 5.9)

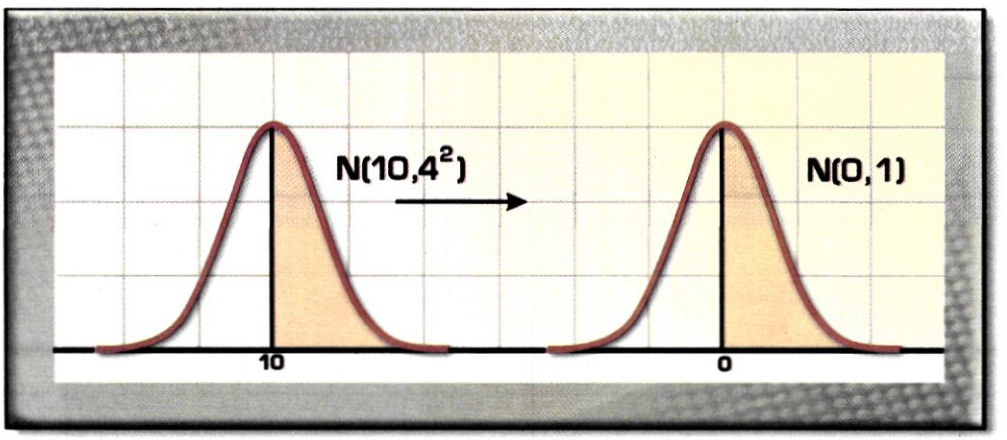


Σχ. (5.9)

5 Η τ.μ. $X \sim N(10,4^2)$ και ζητάμε την πιθανότητα $P(X>10)$

Απάντηση

Ομοίως : $$P(X>10) = P\left(\frac{X-10}{4} > \frac{10-10}{4}\right) = P(Z>0)$$ (Σχ. 5.10)

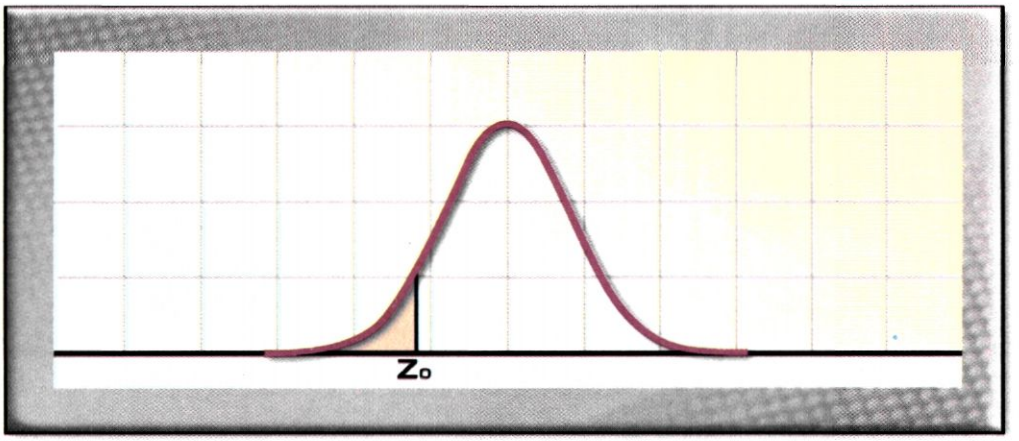


Σχ. (5.10)

Προκειμένου να υπολογίσουμε τις πιο πάνω πιθανότητες είναι ανάγκη να ορίσουμε τη συνάρτηση **αθροιστικής κατανομής** ή απλώς **συνάρτηση κατανομής** της τυποποιημένης κανονικής κατανομής N(0,1).

Η συνάρτηση κατανομής **(σ.κ) της N(0,1) συμβολίζεται με Φ(z) και ορίζεται από την $P(-\infty < Z < z) = P(Z < z)$.**

Δηλαδή, για κάποιο z_o η $P(-\infty<Z<z_o) = \Phi(z_o)$. (Σχ. 5.11)



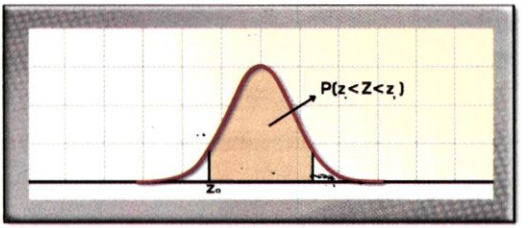
Σχ. (5.11) Η γραμμοσκιασμένη περιοχή αντιστοιχεί στην τιμή $\Phi(s_o)$

Βασιζόμενοι λοιπόν στον ορισμό της Φ(z) είναι φανερό ότι πιθανότητες της μορφής $P(\alpha<X<\beta)$, μετασχηματιζόμενες σε πιθανότητες της μορφής $P(z_\alpha<Z<z_\beta)$ υπολογίζονται ως ακολούθως :

$$P(\alpha < X < \beta) = P\left(\frac{\alpha-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{\beta-\mu}{\sigma}\right) = P(z_\alpha < Z < z_\beta)$$

$$P(-\infty < Z < z_\beta) - P(-\infty < Z < z_\alpha) = \Phi(z_\beta) - \Phi(z_\alpha),$$

όπου $z_\alpha = \frac{\alpha-\mu}{\sigma}$ **και** $z_\beta = \frac{\beta-\mu}{\sigma}$ **οι τυποποιημένες τιμές της μεταβλητής** $X \sim N(\mu,\sigma^2)$.



Σχ. (5.12) Η κίτρινη περιοχή αντιστοιχεί στην τιμή της διαφοράς $\Phi(z_\beta) - \Phi(z_\alpha)$

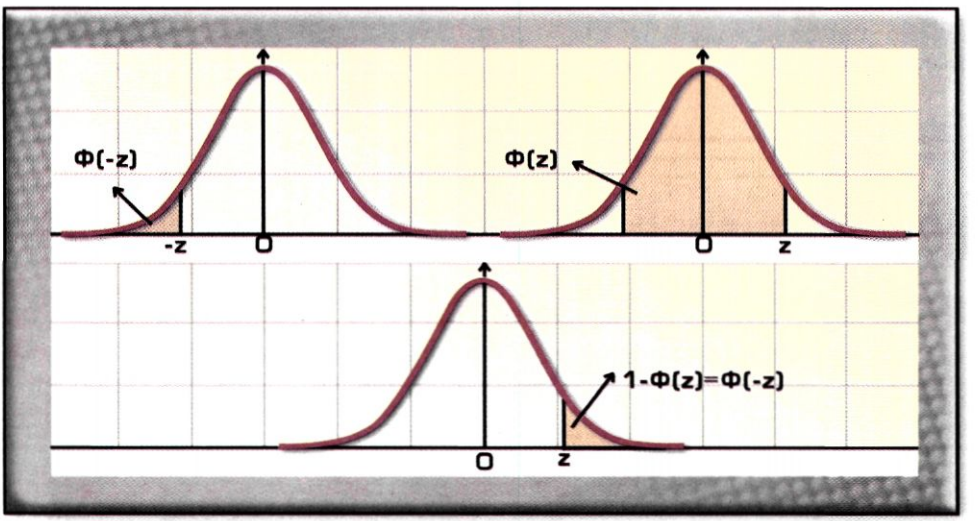
Επανερχόμενοι στο παράδειγμα (5.1) της τ.μ. $X \sim N(100,15^2)$, η πιθανότητα

$$P(115<X<130) = P\left(\frac{115-100}{15} < \frac{X-100}{15} < \frac{130-115}{15}\right) = P(1<Z<2) =$$
$$P(-\infty < Z < 2) - P(-\infty < Z < 1) = \Phi(2) - \Phi(1).$$

Η εύρεση των τιμών της Φ(z) διευκολύνεται εξαιρετικά από την ύπαρξη εκτεταμένων πινάκων με τις τιμές της.

Στον πίνακα (Β) του βιβλίου δίνονται οι τιμές της Φ(z) για z στο διάστημα (0,00 έως 4,00), έτσι για να υπολογίσουμε τις τιμές Φ(2) και Φ(1) ανατρέχουμε στον πίνακα (Β) και βρίσκουμε ότι Φ(2) = 0,9772 και Φ(1) = 0,8413, οπότε $P(-\infty < Z < 2) - P(-\infty < Z < 1) = \Phi(2)-\Phi(1) = 0,9772-0,8413=0,1359$.

Η πλήρης συμμετρία της κανονικής κατανομής επιτρέπει τον υπολογισμό πιθανοτήτων που αφορούν και αρνητικές τιμές της τ.μ. Z. [Σχ. (5.13)]



Σχ. (5.13) Υπολογισμός πιθανοτήτων αρνητικών τιμών της τ.μ. Z.

Αν λοιπόν έχουμε να υπολογίσουμε την πιθανότητα $P(-\infty<Z<-z) = \Phi(-z)$ καταφεύγουμε στη σχέση:

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

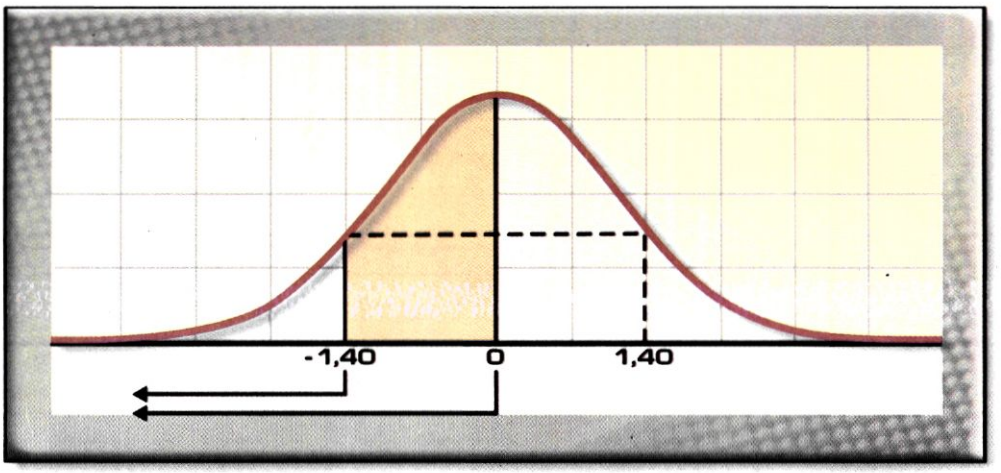
Άρα η $P(-\infty < Z < -1,4) = \Phi(-1,4) = 1-\Phi(1,4) = 1-0,9192 = 0,0808$.

Παραδείγματα :

6. Να βρεθεί το εμβαδόν κάτω από την τυποποιημένη κανονική καμπύλη μεταξύ $Z = -1,40$ και $Z = 0$.

Απάντηση

Όπως προκύπτει από το σχήμα (5.14) το εμβαδόν κάτω από την καμπύλη μεταξύ των τιμών $z = -1,40$ και $z = 0$, ισούται με το εμβαδόν κάτω από την καμπύλη μεταξύ των τιμών $z = 0$ και $z = 1,40$ (λόγω συμμετρίας).

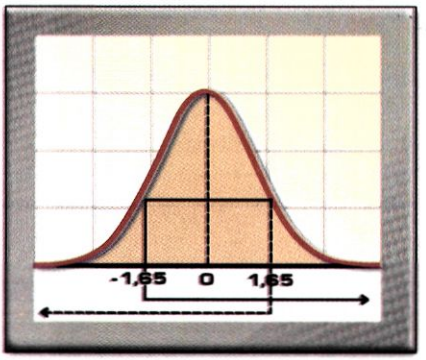


Σχ. (5.14) Πιθανότητα της τ.μ. Z να βρίσκεται μεταξύ −1,40 και 0

Από τον πίνακα I για $z = 1,40$ παίρνουμε $\Phi(1,40) = 0,9192$ και για $z = 0$, $\Phi(0) = 0,5$. Άρα : $P(-1,40< Z<0) = P(0<Z<1,40) =$
$= \Phi(1,40) - \Phi(0) = 0,9192-0,5000 = 0,4192$.

7. Να βρεθεί το εμβαδόν κάτω από την τυποποιημένη κανονική καμπύλη δεξιά του z = -1,65.

Απάντηση

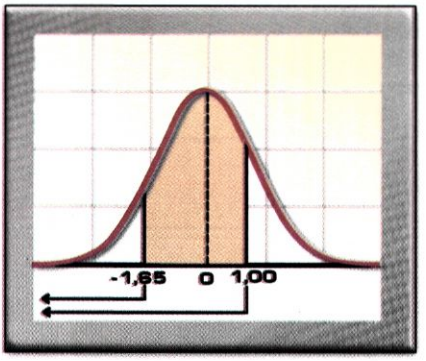


Από τη γνωστή συμμετρία της κανονικής κατανομής N(0,1) είναι φανερό ότι :
P(Z>-1,65)=P(Z<1,65) = 0,9505 (Σχ. 5.15)

Σχ. (5.15) Πιθανότητα της τ.μ. Z να είναι μικρότερη του -1,65

8. Να βρεθεί το εμβαδόν κάτω από την τυποποιημένη κανονική καμπύλη μεταξύ των τιμών z = -1,65 και z=1,00.

Απάντηση



Ζητάμε την P(-1,65<Z<1,00) =
= P(Z<1,00) - P(Z<-1,65) =
= 0,8413 – P(Z>1,65) (λόγω συμμετρίας)
= 0,8413 – [1 – P(Z<1,65)] =
= 0,8413 – 1 + 0,9505 =
= 0,7918
Άρα το εμβαδόν μεταξύ z = -1,65 και z=1,00 είναι 0,7918. (Σχ. 5.16)

Σχ. (5.16) Η πιθανότητα της τ.μ. Z να είναι μεταξύ -1,65 και 1,00

Το αποτέλεσμα ερμηνεύεται ως εξής :

1. Αν ένα χαρακτηριστικό (μια μεταβλητή) έχει την τυποποιημένη κανονική κατανομή N(0,1), τότε μεταξύ z = -1,65 και z = 1,00 βρίσκεται το 79,18% των τιμών του χαρακτηριστικού.

2. Η πιθανότητα μιας τ.μ. Z~N(0,1) να πάρει τιμές μεταξύ z=-1,65 και z=1,00 είναι 0,7918, δηλαδή

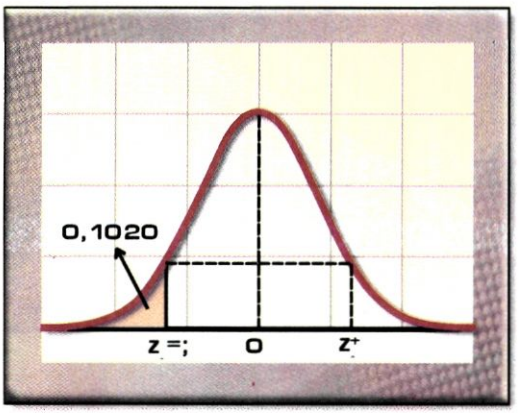
$$P(-1,65<Z<1,00) = 0,7918$$

Στα παραδείγματα που προηγήθηκαν ερμηνεύσαμε το εμβαδόν ως **ποσοστό** και ως **πιθανότητα**. Θα δούμε τώρα πώς γνωρίζοντας το ποσοστό ή την πιθανότητα μπορούμε να βρούμε το αντίστοιχο z.

Σημείωση: Αν η τ.μ. X έχει κανονική κατανομή $N(\mu, \sigma^2)$, λέμε ότι κατανέμεται κανονικά ή ότι ακολουθεί κανονική κατανομή με μέσο μ και διακύμανση σ^2.
Αν το χαρακτηριστικό X των ατόμων ενός πληθυσμού (δηλαδή η τ.μ. X, που μπορεί π.χ. να είναι το βάρος, το ύψος, το IQ κ.λ.π. των ατόμων) ακολουθεί κανονική κατανομή $N(\mu,\sigma^2)$, τότε λέμε ότι ο ***πληθυσμός είναι κανονικός*** *εξεταζόμενος ως προς το συγκεκριμένο χαρακτηριστικό.*
Έτσι αν π.χ. η μεταβλητή X = βάρος ακολουθεί $N(\mu, \sigma^2)$, ο πληθυσμός των βαρών είναι – λέγεται κανονικός.

Παραδείγματα, εύρεσης των τιμών του z, όταν πιθανότητες της μορφής P[Z<z] ή P[Z>z] είναι γνωστές

9. Σε ένα κανονικώς κατανεμημένο πληθυσμό το 10,2% των ατόμων έχουν Z μικρότερο κάποιου συγκεκριμένου z_o. Ποιο είναι το z_o ;



Σχ. (5.17)

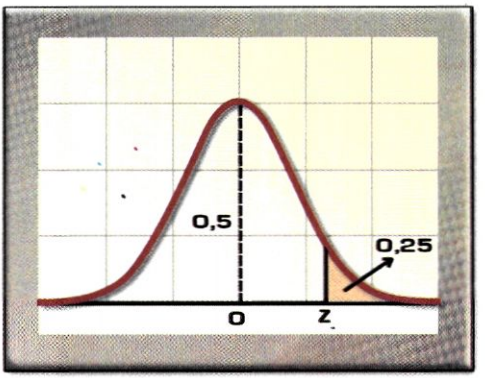
Η τιμή του αγνώστου z_o θα είναι αρνητική, διότι η $P(Z< z_o) = 0,1020 < 0,5$. Από τη γνωστή συμμετρία της N(0,1) έχουμε :

$P(Z< z_o) = P(Z> z_o^+) = 0,1020$

$= 1 - P(Z< z_o^+)$ ή $P(Z< z_o^+) = 0,8980$

Από τον πίνακα (Β) προκύπτει ότι η τιμή του Z που αντιστοιχεί στην πιθανότητα 0,8980 είναι $z_o^+ = 1,27$. Άρα το ζητούμενο $z_o = -1,27$.

10. Ένα z_o είναι τέτοιο ώστε $P(Z> z_o) = 0,025$. Ποια είναι η τιμή του z_o ;



Η τιμή του αγνώστου z_o θα είναι θετική διότι η $P(Z> z_o) = 1 - P(Z< z_o) = 0,025$ ή $P(Z< z_o) = 0,975>0,5$
Από τον πίνακα (Β) η τιμή του z_o που αντιστοιχεί στη συγκεκριμένη πιθανότητα είναι $z_o=1,96$.

Σχ. (5.18)

ΑΣΚΗΣΕΙΣ

Υπολογίστε τις πιθανότητες των παραδειγμάτων 1 έως και 5.

5.4 Εύρεση των τιμών X από τις τιμές Z

Στα περισσότερα προβλήματα που αφορούν κανονικές κατανομές, αναζητάμε πιθανότητες της αρχικής μεταβλητής X της μορφής $P[X < x]$ ή $P[X > x]$, με συνέπεια να καταφεύγουμε στην τεχνική της τυποποίησης μέσω του μετασχηματισμού $Z = \dfrac{X-\mu}{\sigma}$ και στην ακόλουθη διαδικασία :

1. Μετασχηματίζουμε την τ.μ. X σε $Z = \dfrac{X-\mu}{\sigma}$ και
2. Βρίσκουμε από τον πίνακα (Β) της κανονικής κατανομής N(0,1) την $P[Z<z]$ ή $P[Z>z] = 1-P[Z<z]$.

Αν τώρα υποθέσουμε ότι η τιμή των $P[Z<z]$ ή $P[Z>z]$ **είναι γνωστή**, πώς μπορούμε να βρούμε την τιμή x της τ.μ. X ;

1η Περίπτωση Η $P[Z<z]$ γνωστή.

1. Από τον πίνακα (Β) βρίσκουμε την τιμή z της τ.μ. Z που αντιστοιχεί στην πιθανότητα $P[Z<z]$.

2. Λύνοντας τη σχέση $Z = \frac{X-\mu}{\sigma}$ ως προς x, τιμή της τ.μ. X προκύπτει ότι **$x=\mu+z\sigma$** έκφραση που χρησιμοποιείται για το μετασχηματισμό του z σε x.

2η Περίπτωση Η P[Z>z] γνωστή.

1. Από τη σχέση P[Z>z] = 1-P[Z<z] βρίσκουμε την πιθανότητα P[Z<z].
2. Από τον πίνακα (Β) βρίσκουμε την τιμή z της τ.μ. Z που αντιστοιχεί στην πιθανότητα P[Z<z].
3. Μετασχηματίζουμε το z σε x χρησιμοποιώντας τη σχέση :

$$x = \mu+z\sigma$$

Παραδείγματα

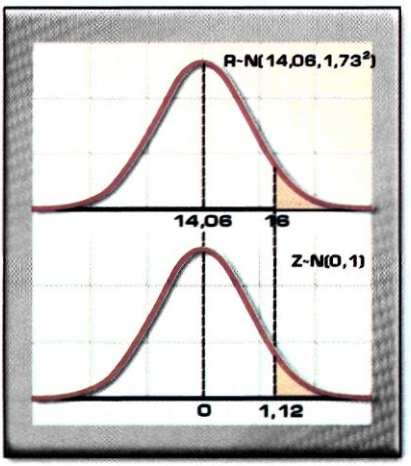
11. Ο τεχνικός ποιοτικού ελέγχου ενός εργοστασίου κατασκευής αντιστάσεων R = 14 OHMS, βρίσκει ότι η κατανομή των αντιστάσεων είναι κανονική με μέσο μ_R = 14,06 OHMS περίπου και σ_R=1,73 OHMS.

α) Ποια είναι η πιθανότητα τυχαία επιλεγόμενη αντίσταση να έχει R ≥ 16 OHMS.

β) Ποια είναι η πιθανότητα η αντίστοιχη R να έχει τιμές αποκλίνουσες από το μέσο μ_R το πολύ 1 OHM ;

Απάντηση

α) Ζητάμε την P(R ≥ 16) αν είναι γνωστό ότι η τ.μ. R~N(14,06,1,73^2).
Τυποποιώντας έχουμε :



Σχ. (5.19)

$$P\left(\frac{R-14,06}{1,73} \geq \frac{16-14,06}{1,73}\right) = P(Z \geq 1,12) =$$

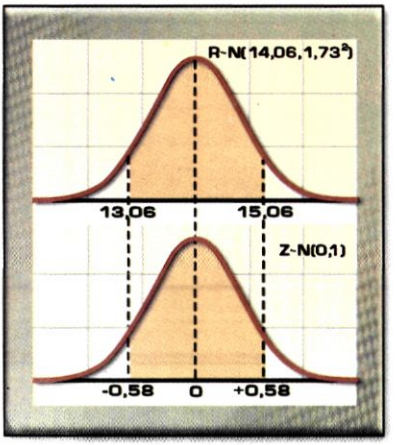
$$=1-P(Z<1,12) =1-0,8686 = 0,1314$$

Μπορούμε συνεπώς να χρησιμοποιήσουμε το R και να γράψουμε: P(R ≥ 16) = 0,1314, που σημαίνει ότι η πιθανότητα τυχαία επιλεγμένη αντίσταση να έχει R ≥ 16 OHMS είναι 0,1314 ή ακόμη ότι ποσοστό 13,14% των κατασκευαζόμενων αντιστάσεων έχουν R ≥ 16 OHMS.

β) Στο δεύτερο ερώτημα θέλουμε να υπολογίσουμε την

$$P(\mu_R - 1 < R < \mu_R + 1) = P(14{,}06 - 1 < R < 14{,}06 + 1) = P(13{,}06 < R < 15{,}06)$$

$$= P\left(\frac{13{,}06 - 14{,}06}{1{,}73} < \frac{R - 14.06}{1{,}73} < \frac{15{,}06 - 14{,}06}{1{,}73}\right) = P(-0{,}58 < Z < 0{,}58) =$$



Σχ. (5.20)

$= P(Z<0{,}58) - P(Z<-0{,}58) =$
$= P(Z<0{,}58) - P(Z>0{,}58) =$
$= P(Z<0{,}58) - [1-P(Z<0{,}58)]$
$= 2P(Z<0{,}58) - 1 =$
$= 2 \cdot 0{,}7190 - 1 = 0{,}438$

Άρα 43,8% των αντιστάσεων έχουν R μεταξύ 13,06 και 15,06 OHMS

12. Κατά τη διαδικασία του ποιοτικού ελέγχου του προηγούμενου παραδείγματος, δημιουργήθηκε η ανάγκη προσδιορισμού ενός R (σε OHMS) πέραν του οποίου να βρίσκεται το 5,48% των παραγόμενων αντιστάσεων.
Ποια είναι η τιμή του r_o αν όπως και στο παράδειγμα 11 υποθέσουμε ότι $R \sim N(14{,}06, 1{,}73^2)$.

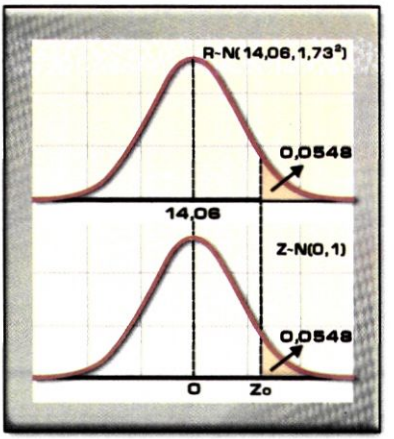
Απάντηση

Ζητάμε να βρούμε r_o τέτοιο ώστε :
$P(R>r_o) = 0.0548$, όπου
$R \sim N(14{,}06, 1{,}73^2)$

ή

$$P\left(\frac{R - 14{,}06}{1{,}73} > \frac{r_o - 14{,}06}{1{,}73}\right) =$$

$P(Z>z_o) = 0{,}0548$
ή
$P(Z<z_o) = 1 - 0{,}0548 = 0{,}9452$



Σχ. (5.21)

Συνεπώς το z_o είναι η τιμή του Z που αντιστοιχεί σε πιθανότητα 0,9452. Από τον πίνακα βρίσκουμε ότι $z_o = 1,6$ συνεπώς :

$$\frac{r_o - 14,06}{1,73} = 1,6 \quad \text{ή} \quad r_o = 14,06+1,6\cdot 1,73 = 16,828$$

Η τιμή του R πέρα της οποίας βρίσκεται το 5,48% των αντιστάσεων είναι r_o=16,823 OHMS.

13. Η τ.μ. X~N(19,49). Προσδιορίστε την τιμή του α για την οποία P[X<α] = 0,90.

Απάντηση

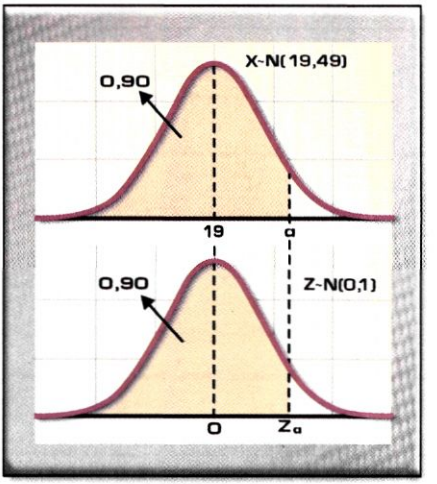
Η $P[X<\alpha] = P\left[\frac{X-19}{7} < \frac{\alpha-19}{7}\right] = 0,90$

Έτσι : $P[Z<z_\alpha] = 0,90$, όπου $z_\alpha = \frac{\alpha-19}{7}$

Από τον πίνακα (Ι) η τιμή του Z που αντιστοιχεί σε πιθανότητα 0,90 είναι z_α =1,28.

Άρα : $\frac{\alpha-19}{7} = 1,28$

ή $\alpha = 19+1,28\cdot 7 = 27,96$



Σχ. (5.22)

14. Μηχανή κοπής ξυλείας είναι ρυθμισμένη να κόβει δοκάρια μήκους 2 μέτρων. Η μηχανή όμως είναι παλιά και αν και τα κομμάτια που κόβει έχουν μέσο 2 μέτρων, το 10% της παραγωγής έχει μήκος μικρότερο των 1,95 μέτρων. Υποθέτουμε ότι τα παραγόμενα μήκη ακολουθούν κανονική κατανομή. Ποιο ποσοστό δοκαριών έχει μήκος μεγαλύτερο των 2,10 μέτρων ;

Απάντηση

Συμβολίζουμε με X το μήκος των δοκαριών για το οποίο έχουμε την πληροφορία ότι :
$X\sim N(2,\sigma^2)$ και $P(X<1,95) = 0,10$. Θέλουμε να βρούμε την πιθανότητα : $P(X>2,10)$
Πρώτα όμως πρέπει να υπολογίσουμε την σ.
Από την P[X<1,95]

$$= P\left[\frac{X-2}{\sigma} < \frac{1,95-2}{\sigma}\right] = 0,10$$

καθώς και από τον πίνακα (I) έχουμε ότι
$P[Z<-1,282] = 0,10$.

Άρα $\frac{1,95-2}{\sigma} = -1,282$

ή $\sigma = 0,0390$.

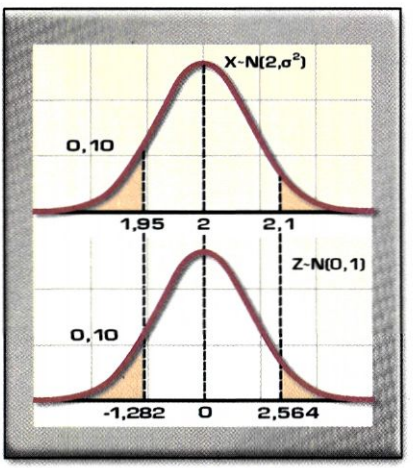
Ξέρουμε λοιπόν τώρα ότι:

$X \sim N(2, 0,0390^2)$

και ζητάμε την $P[X>2,1]$

$$= P\left[\frac{X-2}{0,039} > \frac{2,1-2}{0,039}\right] = P[Z>2,564] =$$

$1 - P[Z<2,564] = 1 - 0,9948 = 0,0052$

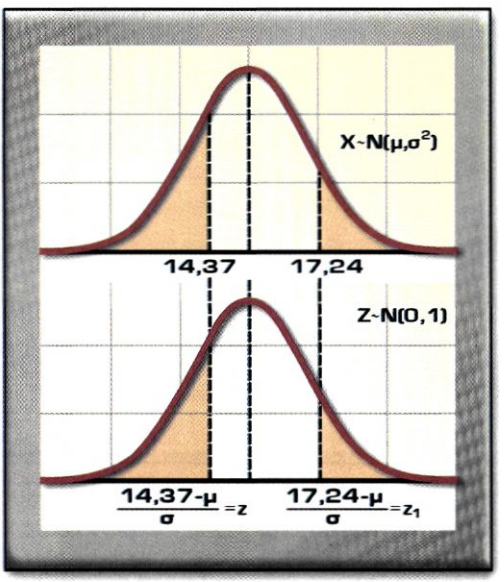


Σχ. (5.23)

Και συνεπώς 5,2% των δοκαριών έχει μήκος μεγαλύτερο των 2,1 μέτρων.

15. Η τ.μ. $X \sim N(\mu,\sigma^2)$. Αν είναι γνωστό ότι το 10% των τιμών της X είναι μεγαλύτερες του 17,24 και ότι το 25% των τιμών της είναι μικρότερες του 14,37, να βρεθούν οι τιμές των μ και σ.

Απάντηση



Σχ. (5.24)

$P[X>17,24] = 0,10$ και
$P[X<14,37] = 0,25$
Τυποποιώντας παίρνουμε :

$$P\left[\frac{X-\mu}{\sigma} > \frac{17,24-\mu}{\sigma}\right] = 0,10 \text{ και}$$

$$P\left[\frac{X-\mu}{\sigma} < \frac{14,37-\mu}{\sigma}\right] = 0,25.$$

Αν $z_1 = \frac{17,24-\mu}{\sigma}$ και $z_2 = \frac{14,37-\mu}{\sigma}$,

τότε : $P[Z>z_1] = 1 - P[Z<z_1] = 0,10$ ή
$P[Z<z_1] = 0,90$ και $P[Z<z_2] = 0,25$
Από τον πίνακα (B) έχουμε :
$z_1 = 1,282$ και $z_2 = -0,674$

Συνεπώς : $\frac{14,37-\mu}{\sigma} = -0,674$ και

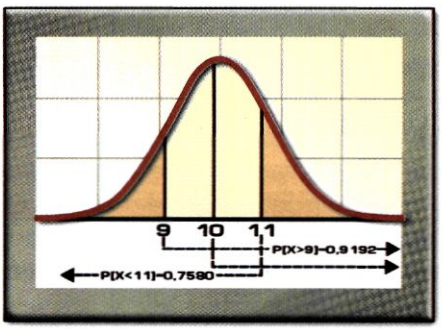
$\frac{17,24-\mu}{\sigma} = 1,282$

Λύνοντας το σύστημα ως προς μ και σ βρίσκουμε : μ =15,4 και σ =1,47

16. Η τ.μ. $X \sim N(\mu,\sigma^2)$. Είναι γνωστό ότι $P[X>9] = 0,9192$ και ότι $P[X<11] = 0,7580$. Να βρεθεί η $P[X>10]$.

Απάντηση

Οι τιμές των παραμέτρων μ και σ που καθορίζουν την κατανομή είναι άγνωστες. Από το σχήμα της καμπύλης προκύπτει ότι ο μέσος μ πρέπει να βρίσκεται μεταξύ 9 και 11 και επειδή 0,9192 > 0,7580, ο μέσος πρέπει να βρίσκεται λίγο δεξιότερα του 10 (γιατί;).



Σχ. (5.25)

Από την πιθανότητα $P[X>9] = 0,9192$

έχουμε $P\left[\frac{X-\mu}{\sigma} > \frac{9-\mu}{\sigma}\right] = 0,9192$

ή $P[Z>z_9]=0,9192$ ή

$1-P[Z<z_9]=0,9192$ και

$[Z<z_9] = 0,0808$ οπότε από τον πίνακα (Β) έχουμε $z_9 = -1,4$.

Ομοίως από την πιθανότητα

$P[X<11]=\left[\frac{X-\mu}{\sigma} < \frac{11-\mu}{\sigma}\right]=$

$P[Z<z_{11}] = 0,7580$

και τον πίνακα (Ι) παίρνουμε $z_{11} = 0,700$.

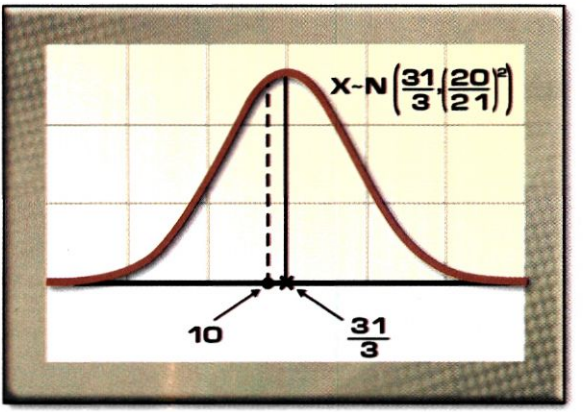
Έτσι έχουμε :

$$z_9 = \frac{9-\mu}{\sigma} = -1,4 \text{ και } z_{11} = \frac{11-\mu}{\sigma} = 0,700$$

ή $\begin{cases} 9-\mu = -1,4\sigma \\ 11-\mu = 0,7\sigma \end{cases}$ οπότε $\mu = \frac{31}{3}$ και $\sigma = \frac{20}{21}$

Μπορούμε λοιπόν τώρα να υπολογίσουμε τη ζητούμενη πιθανότητα. Πράγματι :

$$P[X>10] = P\left[\frac{X-31/3}{20/24} > \frac{10-31/3}{20/21}\right] = P[Z>-0,35] = P[Z<0,35] = 0,6368$$



Σχ. (5.26)

ΒΑΣΙΚΗ ΠΑΡΑΤΗΡΗΣΗ

Παρουσιάζοντας τις διάφορες κατανομές πιθανότητας θα πρέπει να παρατηρήσατε την αναφορά που γίνεται στις παραμέτρους της κάθε κατανομής.
Στην Bernoulli αναφέραμε την p, στη διωνυμική τη ν και p, στη γεωμετρική την p, στην Poisson τη λ, στην υπεργεωμετρική την α, ν και β και τώρα στην κανονική την μ και σ.

Ο ρόλος των παραμέτρων αυτών είναι να μας δίνουν για κάθε δυνατή τιμή τους μία και μόνο μία από τις δυνατές μορφές που μπορεί να έχει η κατανομή. Θυμηθείτε πόσο διαφορετική εικόνα έχει η B(ν, p) για $p<1/2$, $p=1/2$ και $p>1/2$ ή πόσο αλλάζει το διάγραμμα πιθανότητας της γεωμετρικής κατανομής, όταν δίνουμε διαφορετικές τιμές στο p.

Στην πράξη όμως, οι πραγματικές τιμές των παραμέτρων δεν είναι γνωστές και συνεπώς πρέπει με κάποιο τρόπο να «εκτιμηθούν».

Η εκτίμηση των αγνώστων παραμέτρων των κατανομών αποτελεί αντικείμενο ενός κλάδου της Στατιστικής της «ΕΚΤΙΜΗΤΙΚΗΣ».

Οι εκτιμήσεις των παραμέτρων επιτυγχάνονται με τη βοήθεια των «στατιστικών». **Καλούμε δε στατιστικό, κάθε συνάρτηση των δεδομένων ενός δείγματος.**

Παραδείγματα στατιστικών αποτελούν, μεταξύ πληθώρας άλλων, ο δειγματικός μέσος $\bar{X}=\frac{1}{\nu}\sum x_i$ που εκτιμά τον πληθυσμιακό μέσο μ, η δειγματική διακύμανση $S^2=\frac{1}{\nu-1}\sum(x_i-\bar{X})^2$ που εκτιμά τη σ^2 του πληθυσμού κ.λ.π.

Παρατηρούμε ότι τα παραπάνω παραδείγματα στατιστικών αποτελούν πράγματι συναρτήσεις των δεδομένων του δείγματος.

5.5 Δειγματική Κατανομή – Η Κατανομή του μέσου $\overline{X}$

Ας θεωρήσουμε όλα τα δυνατά δείγματα, μεγέθους ν, που σχηματίζονται τυχαία από έναν πληθυσμό. Η κατανομή των δυνατών τιμών ενός στατιστικού που μετράται σ' όλα τα τυχαία δείγματα μεγέθους ν που μπορούμε να πάρουμε από το πληθυσμό, ονομάζεται **δειγματική κατανομή** του συγκεκριμένου στατιστικού. Δειγματικές κατανομές μπορούν να ληφθούν εμπειρικά, όταν σχηματίζουμε δείγματα από έναν πεπερασμένο πληθυσμό. Για να δημιουργήσουμε **εμπειρική δειγματική κατανομή** ακολουθούμε την εξής διαδικασία :

1. Από ένα πεπερασμένο πληθυσμό μεγέθους ***N***, σχηματίζουμε κατά τυχαίο τρόπο* δείγματα.
2. Υπολογίζουμε το στατιστικό σε κάθε δείγμα.
3. Δημιουργούμε την κατανομή συχνοτήτων με το γνωστό τρόπο.

Αυτή η διαδικασία είναι δύσκολο να ακολουθηθεί για πληθυσμό μεγάλου μεγέθους, ενώ είναι αδύνατο για άπειρο πληθυσμό. Σ' αυτές τις περιπτώσεις είναι δυνατή η προσέγγιση της δειγματικής κατανομής με τη χρησιμοποίηση μεγάλου αριθμού δειγμάτων.

Τρία χαρακτηριστικά μιας δειγματικής κατανομής μας ενδιαφέρουν κατά κανόνα : **ο μέσος, η διασπορά** και η **γραφική της μορφή.**

Μια πολύ σημαντική δειγματική κατανομή είναι αυτή του μέσου $\overline{X}$. Ας δούμε πως μπορούμε να τη δημιουργήσουμε ακολουθώντας τα βήματα που προαναφέραμε.

Ας υποθέσουμε την ύπαρξη ενός πληθυσμού από N=5 παιδιά, με ηλικίες x_1=4, x_2=6, x_3=8, x_4=10, x_5=12, που ανήκουν σε κάποια κοινότητα. Ο μέσος του πληθυσμού είναι ίσος με

$$\mu = \frac{\sum_{j=1}^{5} x_j}{5} = \frac{4+6+8+10+12}{5} = 8$$

και η διασπορά του υπολογίζεται από την :

$$\sigma^2 = \frac{\sum_{j=1}^{5} (x_j - \mu)^2}{5} = \frac{(-4)^2 + (-2)^2 + 2^2 + 4^2}{5} = \frac{40}{5} = 8 \text{ και } \sigma = 2\sqrt{2}\,.$$

Ας σχηματίσουμε τώρα όλα τα δυνατά δείγματα μεγέθους ν=2 από τον πληθυσμό των 5 παιδιών. Αν χρησιμοποιήσουμε ως μέθοδο επιλογής του δείγματος, τη «δειγματοληψία με επανάθεση» (δηλαδή κάθε παιδί που επιλέγεται στο δείγμα, επανέρχεται στον πληθυσμό και επομένως μπορεί πάλι να επιλεγεί σε άλλο δείγ-

* Τα τυχαία δείγματα αναπτύσσονται στην παράγραφο 6.0 και 6.1 του 6ου κεφαλαίου.

μα), τότε το πλήθος των δυνατών δειγμάτων μεγέθους ν = 2 από πληθυσμό μεγέθους N = 5 είναι $5^2 = 25$ και δίνονται στον πιο κάτω πίνακα (5.1)

Πίνακας (5.1)

	4	6	8	10	12
4	4,4 (4)	4,6 (5)	4,8 (6)	4,10 (7)	4,12 (8)
6	6,4 (5)	6,6 (6)	6,8 (7)	6,10 (8)	6,12 (9)
8	8,4 (6)	8,6 (7)	8,8 (8)	8,10 (9)	8,12 (10)
10	10,4 (7)	10,6 (8)	10,8 (9)	10,10 (10)	10,12 (11)
12	12,4 (8)	12,6 (9)	12,8 (10)	12,10 (11)	12,12 (12)

Τα δυνατά δείγματα μεγέθους ν=2 από τον πληθυσμό των 5 τιμών : 4, 6, 8, 10, 12. Σε παρένθεση είναι ο μέσος όρος του δείγματος

Δημιουργούμε τη δειγματική κατανομή του μέσου $\overline{X}$, που δίνεται στον πίνακα (5.2)

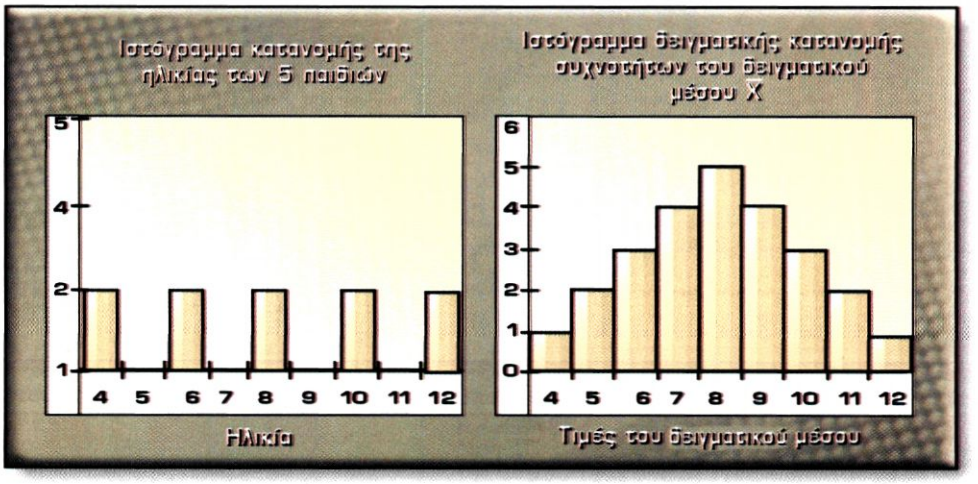
Πίνακας (5.2)

j	x_j	Συχνότητα v_j	Πιθανότητα $P_j=\frac{v_j}{v}$
1	4	1	1/25
2	5	2	2/25
3	6	3	3/25
4	7	4	4/25
5	8	5	5/25
6	9	4	4/25
7	10	3	3/25
8	11	2	2/25
9	12	1	1/25
	Σύνολο	**25**	**25/25=1**

Δειγματική κατανομή συχνοτήτων των 25 δειγμάτων για τον $\overline{X}$

Όπου j = 1, 2, 3, …, 9 δείκτης του αντίστοιχου μέσου και κ = 9 το πλήθος των τιμών των δειγματικών μέσων.

Παρακάτω δίνονται τα ιστογράμματα της κατανομής του πληθυσμού των 5 ηλικιών καθώς και της δειγματικής κατανομής του $\overline{X}$.



Σχ. (5.27) Κατανομή του πληθυσμού (αριστερά) και δειγματική κατανομή του μέσου για δείγματα μεγέθους ν=2 (δεξιά)

Η διαφορά ανάμεσα στο ιστόγραμμα της κατανομής του πληθυσμού και το ιστόγραμμα της δειγματικής κατανομής του $\overline{X}_j$ είναι εντυπωσιακή. Ενώ στην πρώτη οι τιμές είναι ομοιόμορφα κατανεμημένες, στη δεύτερη παρουσιάζεται κορυφή και είναι πλήρως συμμετρική.

Ας υπολογίσουμε το μέσο $\mu_{\bar{x}}$, της δειγματικής κατανομής με τα στοιχεία του πίνακα (5.2.)

$$\mu_{\bar{x}} = \sum_{j=1}^{9} (\bar{x}_j \cdot P_j) = \frac{\sum_{j=1}^{9} (\bar{x}_j \cdot \nu_j)}{\sum_{1}^{9} \nu_j} = \frac{4\cdot 1 + 5\cdot 2 + 6\cdot 3 + 7\cdot 4 + \ldots + 12\cdot 1}{25} = \frac{200}{25} = 8,$$

όπου $P_j = \dfrac{\nu_j}{\sum_{j=1}^{9} \nu_j}$

Βλέπουμε ότι ο μέσος $\mu_{\bar{x}}$ της δειγματικής κατανομής έχει την ίδια τιμή με τον μέσο μ του πληθυσμού. Τέλος υπολογίζουμε τη διασπορά του $\overline{X}$ που συμβολίζεται με $\sigma_{\bar{x}}^2$.

$$\sigma_{\bar{x}}^2 = \sum_{j=1}^{9} (\bar{x}_j - \mu_{\bar{x}})^2 \cdot P_j = \frac{\sum_{j=1}^{9} (\bar{X}_j - \mu_{\bar{x}})^2 \cdot v_j}{25} = \frac{(4-8)^2 \cdot 1 + (5-8)^2 \cdot 2 + ... + (12-8)^2}{25} = \frac{100}{25} = 4$$

Από το παραπάνω παράδειγμα μπορούμε να συμπεράνουμε ότι :

- Ο μέσος όρος $\mu_{\bar{x}}$ της δειγματικής κατανομής του $\overline{X}$ είναι ίσος με το μέσο μ του πληθυσμού.
- Η τυπική απόκλιση $\sigma_{\bar{x}}$ της δειγματικής κατανομής του $\overline{X}$ είναι μικρότερη της τυπικής απόκλισης σ του πληθυσμού.

Έτσι, για να περιγράψουμε κατά προσέγγιση τη δειγματική κατανομή του μέσου $\overline{X}$ σε ανάλογες περιπτώσεις μεγάλων πληθυσμών και δειγμάτων καταφεύγουμε σε προσομοίωση* με τη βοήθεια υπολογιστή. Αυτό σημαίνει ότι χρησιμοποιούμε τον υπολογιστή ως εξής : Παίρνουμε έναν ορισμένο αριθμό τυχαίων δειγμάτων από ένα πληθυσμό, υπολογίζουμε τους μέσους των δειγμάτων αυτών, και τέλος κατασκευάζουμε το ιστόγραμμα των συχνοτήτων τους.

5.6 Τυπικό σφάλμα του μέσου $\overline{X}$

Για να καθορίσουμε πόσο κοντά βρίσκεται ο μέσος $\overline{X}$ του δείγματος που διαθέτουμε στο μέσο μ του πληθυσμού, από τον οποίο προέρχεται το δείγμα, θα εφαρμόσουμε δυο θεωρήματα. Το πρώτο από τα δύο θεωρήματα εκφράζει ρητά αυτό που ουσιαστικά ανακαλύψαμε με τη βοήθεια του παραδείγματος στην προηγούμενη παράγραφο. Ο μέσος της δειγματικής κατανομής $\mu_{\bar{x}}$ είναι ίσος με το μέσο του πληθυσμού μ και η τυπική απόκλιση $\sigma_{\bar{x}}$ της δειγματικής κατανομής είναι

* Προσομοίωση ονομάζεται η απομίμηση πειράματος τύχης με τη βοήθεια μιας κατάλληλης συσκευής (προσομοιωτής). Πρέπει να προσεχθεί ώστε σε κάθε έκβαση του πειράματος τύχης να αντιστοιχεί ακριβώς μια έκβαση της προσομοίωσης και οι πιθανότητες των εκβάσεων του πειράματος να είναι ίσες με τις πιθανότητες των αντιστοίχων εκβάσεων της προσομοίωσης. Στην Θεωρία Πιθανοτήτων μιλάμε για εξαγωγή σφαιρών από κάλπες. Αυτός είναι ένας τρόπος προσομοίωσης πειραμάτων τύχης (Πρότυπο κάλπης). Στην πράξη πιο σημαντικό είναι το πρότυπο Μόντε Κάρλο η προσομοίωση δηλαδή με τη βοήθεια τυχαίων ψηφίων.

μικρότερη της τυπικής απόκλισης σ του πληθυσμού. Το θεώρημα μπορεί να διατυπωθεί ως εξής :

Η δειγματική κατανομή του μέσου $\overline{X}$ **δειγμάτων μεγέθους ν, που λαμβάνονται από ένα πληθυσμό με μέσο μ και τυπική απόκλιση σ, έχει μέσο** $\mu_{\overline{x}} = \mu$ ***και τυπική απόκλιση :***

- $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{\nu}}$ στην περίπτωση άπειρου πληθυσμού ή πεπερασμένου πληθυσμού και δειγματοληψίας με επανατοποθέτηση.
- $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{\nu}} \cdot \sqrt{\dfrac{N-\nu}{N-1}}$ στην περίπτωση πεπερασμένου πληθυσμού μεγέθους N και δειγματοληψία χωρίς επανατοποθέτηση.

Η τυπική απόκλιση της δειγματικής κατανομή $\sigma_{\overline{X}}$ ονομάζεται ***τυπικό σφάλμα του μέσου*** ή απλά ***τυπικό σφάλμα.*** Ο ρόλος του τυπικού σφάλματος είναι θεμελιώδης επειδή εκφράζει το βαθμό της τυχαίας απόκλισης των δειγματικών μέσων $\overline{X}$ από τον πληθυσμιακό μέσο $\mu_{\overline{x}} = \mu$.

Οπωσδήποτε κάποια γνώση για αυτή την τυπική απόκλιση είναι ουσιαστική στον καθορισμό του κατά πόσο καλά ο δειγματικός μέσος $\overline{X}$ εκτιμά το μέσο μ του πληθυσμού. Αντιλαμβανόμαστε ότι όσο μικρότερο είναι το τυπικό σφάλμα $\sigma_{\overline{X}}$ τόσο καλύτερη φαίνεται να είναι η εκτίμηση του μέσου, μ.

Η ακρίβεια της εκτίμησης μπορεί να διαπιστωθεί από την επισκόπηση των προηγούμενων τύπων του τυπικού σφάλματος που δόθηκαν. Συγκεκριμένα από τους τύπους του $\sigma_{\overline{X}}$ διαπιστώνεται ότι το τυπικό σφάλμα του μέσου ελαττώνεται, όταν η **μεταβλητότητα** του πληθυσμού ελαττώνεται και όταν αυξάνεται το δειγματικό μέγεθος ν.

Παραδείγματα

17. Όταν παίρνουμε δείγματα από ένα άπειρο πληθυσμό, πως μεταβάλλεται το τυπικό σφάλμα του μέσου $\overline{x}$ όταν το μέγεθος του δείγματος αυξηθεί από ν = 30 σε ν = 270 ;

Απάντηση

Παίρνοντας το πηλίκο των δύο τυπικών σφαλμάτων έχουμε :

$$\frac{\frac{\sigma}{\sqrt{270}}}{\frac{\sigma}{\sqrt{30}}} = \frac{\sqrt{30}}{\sqrt{270}} = \frac{1}{3}$$

Το τυπικό σφάλμα ελαττώνεται κατά τα 2/3 όταν το μέγεθος του δείγματος εννεαπλασιαστεί ($30 \cdot 9 = 270$).

18. Εφαρμόζοντας τον τύπο του τυπικού σφάλματος στην περίπτωση του παραδείγματος του Πίνακα (5.1) να επαληθευτεί ότι $\sigma_{\bar{X}} = 2$.

Απάντηση

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{2}} = \frac{2\sqrt{2}}{\sqrt{2}} = 2$$

5.7 Κεντρικό Οριακό Θεώρημα (Κ.Ο.Θ)

Στη βασική παρατήρηση της παραγράφου 5.4 είδαμε ότι τα **στατιστικά** χρησιμοποιούνται για να **εκτιμήσουν** άγνωστες **παραμέτρους.**

Κάθε **εκτίμηση** είναι συνδεδεμένη με **πιθανό σφάλμα**, το οποίο επιδιώκουμε να είναι **όσο γίνεται** μικρότερο.

Στη διαδικασία εκτίμησης του άγνωστου μέσου **μ** ενός πληθυσμού που αντιστοιχεί στις τιμές μιας τ.μ X το στατιστικό που χρησιμοποιείται είναι ως γνωστό, ο δειγματικός μέσος $\overline{X}$ και το **σφάλμα εκτίμησης E**, θα είναι η διαφορά μεταξύ του $\overline{X}$ και του μ.

Πρόβλημα

Με ποια πιθανότητα το σφάλμα εκτίμησης του μέσου μ από τον μέσο $\overline{X}$ είναι μικρότερο μιας δοσμένης τιμής κ;

Στο πρόβλημα αυτό μπορούμε να δώσουμε απάντηση μόνο κάτω από ορισμένες προϋποθέσεις τις οποίες εξασφαλίζει ένα από τα θεμελιώδη θεωρήματα της Στατιστικής γνωστό ως **Κεντρικό Οριακό Θεώρημα.**

Για μεγάλα δειγματικά μεγέθη $\nu \geq 30$ **η δειγματική κατανομή του μέσου** $\overline{X}$ **προσεγγίζεται από κανονική κατανομή με μέσο** $\mu_{\overline{X}} = \mu$ **και διακύμανση** $\sigma^2_{\overline{X}} = \frac{\sigma^2}{\nu}$

ή χρησιμοποιώντας σύμβολα

Αν $\nu \geq 30$, **τότε** $\overline{X} \sim N(\mu, \frac{\sigma^2}{\nu})$

Μέχρι τώρα παρατηρήσαμε ότι δεν έχει γίνει αναφορά στην κατανομή του πληθυσμού που αντιστοιχεί στις τιμές της μεταβλητής Χ.

Η «απαίτηση» που προβάλει το Κ.Ο.Θ είναι το μεγάλο δειγματικό μέγεθος ν.

Από το σημαντικό αυτό συμπέρασμα προκύπτει τώρα ότι:

Αν $\nu \geq 30$ **τότε** $\overline{X} \sim N(\mu, \frac{\sigma^2}{\nu})$ **και** $Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{\nu}} \sim N(0,1)$

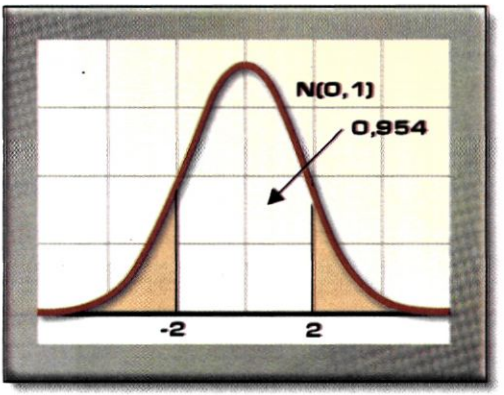
Παρατήρηση: Από την τελευταία σχέση που αφορά την κατανομή των Z προκύπτει ότι ο αριθμητής είναι το σφάλμα της εκτίμησης, $E = \overline{X} - \mu$ που αποτελεί τυχαία μεταβλητή αν ο μέσος μ είναι γνωστός.
Τα σφάλματα έχουν συνεπώς σύμφωνα με όσα είπαμε κανονική κατανομή με μέσο μηδέν και διακύμανση $\frac{\sigma^2}{\nu}$. Δηλαδή $E = \overline{X} - \mu \sim N(0, \frac{\sigma^2}{\nu})$.

Μετά τα όσα προηγήθηκαν είμαστε σε θέση να απαντήσουμε στο πρόβλημα δείχνοντας τον τρόπο με μία εφαρμογή.

Παράδειγμα 19ο :

Ποια είναι η πιθανότητα να έχουμε σφάλμα E < 4, όταν χρησιμοποιηθεί ο μέσος $\overline{X}$ δειγμάτων μεγέθους ν = 49 για την εκτίμηση του μέσου μ ενός άπειρου πληθυσμού με $\sigma^2 = 196$.

Απάντηση



Σχ. (5.28)

Παρατηρούμε αρχικά ότι δε γίνεται αναφορά στην κατανομή που έχει η τ.μ. X στον πληθυσμό (δεν ξέρουμε δηλαδή την κατανομή του πληθυσμού των τιμών της τ.μ. X παρά μόνο την $\sigma^2 = 196$).

Το δειγματικό μέγεθος ν = 49 > 30 χαρακτηρίζεται όπως είπαμε μεγάλο και συνεπώς η κατανομή που δεν γνωρίζουμε προσεγγίζεται από κανονική κατανομή. Θα έχουμε δηλαδή ότι

$$\overline{X} \sim N(\mu, \sigma_{\overline{x}}^2) \quad \text{όπου} \quad \sigma_{\overline{x}}^2 = \frac{\sigma^2}{\nu} = \frac{196}{49}$$

και $\sigma_{\overline{x}} = 2$, οπότε :

$$Z = \frac{\overline{X} - \mu}{\sigma_{\overline{x}}} \sim N(0,1).$$

Στην εκφώνηση αναφέρεται ακόμη ότι το σφάλμα της εκτίμησης του αγνώστου μ από το $\overline{X}$ θέλουμε να είναι μικρότερο του 4. Αφού το $\overline{X}$ εκτιμά τον μ, το σφάλμα θα είναι η διαφορά $E = \overline{X} - \mu$ ή κατ' απόλυτη τιμή $|E| = |\overline{X} - \mu|$.

Ζητάμε την πιθανότητα :

$$P\left[|\overline{X} - \mu| < 4\right] = P\left[-4 < \overline{X} - \mu < 4\right] = P\left[\frac{-4}{\sigma_{\overline{x}}} < \frac{\overline{X} - \mu}{\sigma_{\overline{x}}} < \frac{4}{\sigma_{\overline{x}}}\right] = P\left[\frac{-4}{2} < z < \frac{4}{2}\right] =$$

$$P[-2 < z < 2] = \Phi(2) - \Phi(-2) = 2\Phi(2) - 1 = 0{,}9544$$

5.8 Η κατανομή Student- t

Γνωρίζουμε ήδη ότι αν η τ.μ $X \sim N(\mu,\sigma^2)$ τότε η τ.μ $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$.

Το **συμπέρασμα** αυτό **ισχύει** στην περίπτωση που ο μέσος **μ** και η τυπική απόκλιση **σ** είναι **γνωστά.** Στην πράξη το μ και το σ είναι συνήθως άγνωστα.

Τι συμβαίνει τότε στο πηλίκο $\frac{X-\mu}{\sigma}$;

*Το ερώτημα απασχόλησε τον Ιρλανδό Στατιστικό **(W. Gosset 1908)**, ο οποίος ύστερα από μελέτες κατέληξε σε ορισμένα συμπεράσματα τα οποία δημοσίευσε το 1908 με το ψευδώνυμο "Student".*

Σύμφωνα με αυτά, υποθέτοντας ότι η τ.μ X έχει κανονική κατανομή με γνωστό μέσο μ αλλά άγνωστη διακύμανση σ^2, παρατήρησε ότι:

- Η τ.μ $T = \frac{X-\mu}{S}$, όπου **S** η **δειγματική απόκλιση** που εκτιμά την άγνωστη σ μοιάζει με την Z, και οι καμπύλες των κατανομών των τιμών της τ.μ T είναι πλατύκυρτες δηλαδή 'πλατύτερες'
- Τις κατανομές αυτές ονόμασε **t-κατανομές.**

Παρατήρησε ακόμη ότι η μορφή των καμπυλών της κατανομής της τ.μ T, εξαρτάται από το πλήθος ν των παρατηρήσεων-δεδομένων.

Λαμβάνοντας υπόψη την τελευταία αυτή εξάρτηση της κατανομής από το ν όρισε την κατανομή κατά τρόπο ώστε το σχήμα της να μεταβάλλεται ανάλογα με την τιμή του ν. Εισήγαγε έτσι με τη βοήθεια του ν τη μοναδική **παράμετρο** της κατανομής την οποία συμβόλισε με **d** και ονόμασε **'βαθμούς ελευθερίας της κατανομής'.**

Αποδεικνύεται ότι η κατανομή της τ.μ $T = \frac{\overline{X}-\mu}{S/\sqrt{\nu}}$ έχει $d = \nu-1$ βαθμούς ελευθερίας.

5.9 Η δειγματική κατανομή του μέσου $\overline{X}$ για δείγματα από κανονικό πληθυσμό με άγνωστη διακύμανση σ^2

Ας θυμηθούμε ξανά ότι αν η τ.μ $X \sim N(\mu,\sigma^2)$, τότε η τ.μ $\overline{X} \sim N\left(\mu,\frac{\sigma^2}{\nu}\right)$.

Αν υποθέσουμε ότι ο μέσος μ είναι γνωστός και η διακύμανση σ^2 είναι άγνωστη, τότε:

$$T = \frac{\overline{X} - \mu}{S/\sqrt{\nu}}$$

είναι τυχαία μεταβλητή, που εξαρτάται από δύο άλλες τυχαίες μεταβλητές:

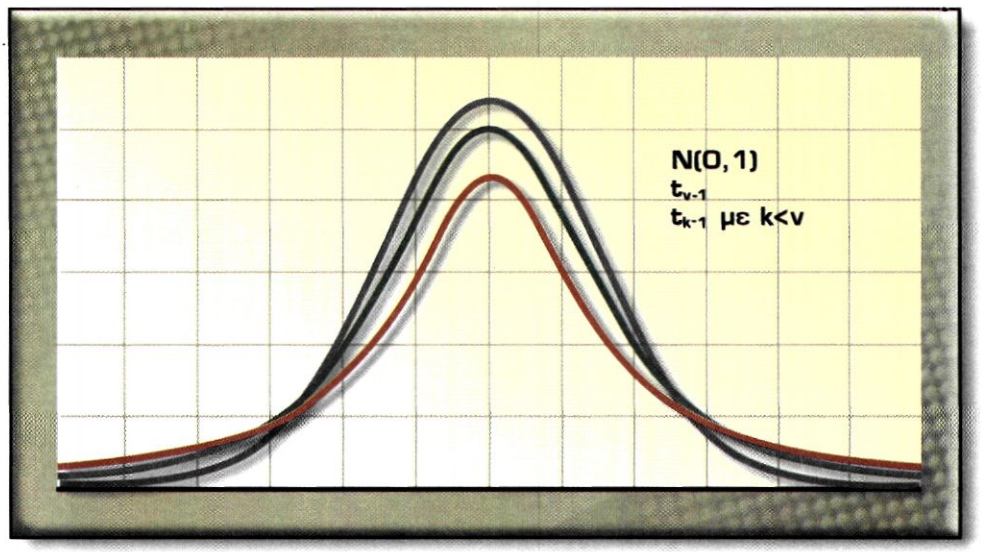
- την $\overline{X}$ στον αριθμητή
- την S στον παρονομαστή

Οι τιμές της τ.μ T μεταβάλλονται μεταξύ των δειγμάτων σταθερού μεγέθους ν, όχι μόνο εξ' αιτίας της μεταβολής του μέσου $\overline{X}$,αλλά και λόγω της μεταβολής της δειγματικής τυπικής απόκλισης S,που και αυτή μεταβάλλεται μεταξύ των δειγμάτων.

Η κατανομή της τ.μ T είναι συμμετρική με μέσο το μηδέν και διακύμανση άμεσα εξαρτώμενη από τους βαθμούς ελευθερίας d, δηλαδή από το ν.

Για ευκολία αντικαθιστούμε την φράση «μία t-κατανομή d βαθμών ελευθερίας» από τη φράση «μία $t_{(d)}$ – κατανομή».

Αποδεικνύεται ότι η τ.μ T έχει $t_{(\nu-1)}$ – κατανομή.



Σχ. (5.29) Δύο t-κατανομές σε σύγκριση με την N(0,1)

5.9.1 Ιδιότητες της $t_{(d)}$ – κατανομής

Αν η τ.μ Τ έχει $t_{(d)}$ – κατανομή τότε ισχύουν τα εξής :

- Η E[T]=0
- Η $\text{Var}[T] = \dfrac{d}{d-2}$ με $d > 2$
- Η κατανομή είναι συμμετρική γύρω από το μέσο
- Η κατανομή είναι διαφορετική για διαφορετικές τιμές του d που ισούται με ν-1 στη συγκεκριμένη περίπτωση.

5.9.2 Πίνακας της $t_{(d)}$ – κατανομής

Οι τιμές t της τ.μ Τ που έχει $t_{(d)}$ – κατανομή, για τις οποίες η πιθανότητα $P[T < t] = a$, βρίσκονται με τη βοήθεια του πίνακα της κατανομής με d βαθμούς ελευθερίας και για επιλεγμένες τιμές της πιθανότητας α.

Εδώ δίνουμε ένα τμήμα του πίνακα της $t_{(d)}$ – κατανομής για να δείξουμε τον τρόπο εύρεσης των τιμών t της τ.μ Τ για δοσμένη α.

1-α / d	0,75	0,90	0,95	0,975	0,990	0,995
1	1,000	3,078	6,314	12,71	31,82	63,66
2	0,816	1,886	2,940	4,303	6,965	9,925
3	0,765	1,638	2,353	3,182	4,541	5,841
...						
...						
...						
	0,674	1,282	1,645	1,960	2,326	2,576

Παράδειγμα 20ο:

Η τυχαία μεταβλητή Τ έχει t-κατανομή με 3 βαθμούς ελευθερίας.
Ποια είναι η τιμή t της Τ για την οποία :

α) $P[T < t] = 0,99$

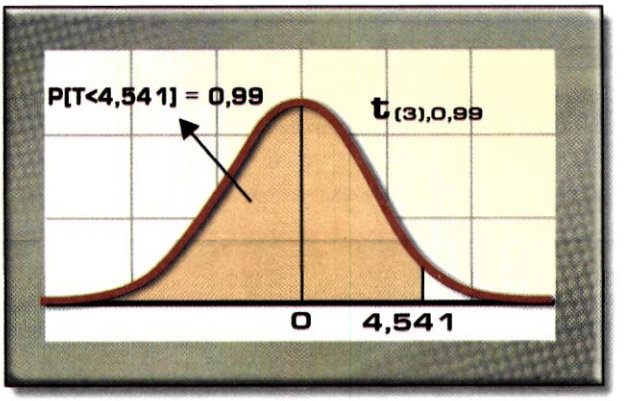
β) $P[T < t] = 0,25$

γ) $P[|T| < t] = 0,98$

δ) $P[|T| > t] = 0,05$

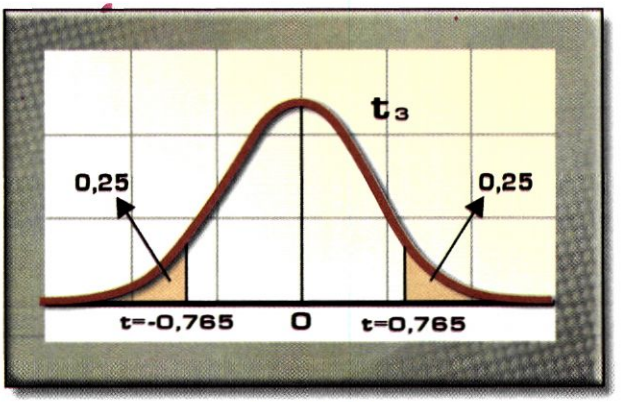
Απάντηση

α) Από τη στήλη του 0,99 και την γραμμή των 3 βαθμών ελευθερίας βρίσκουμε ότι t = 4,541.



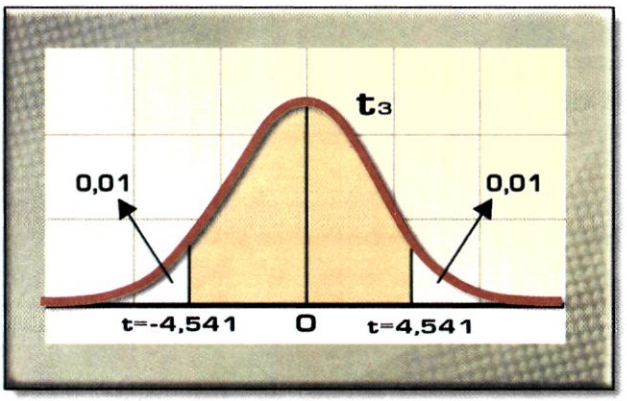
Σχ. (5.30)

β) Από τη στήλη του 0,75 και την γραμμή των 3 βαθμών ελευθερίας βρίσκουμε ότι $P[T < 0,765] = 0,75$ και συνεπώς η $P[T > 0,765] = 0,25$. Από τη συμμετρία της κατανομής, όπως φαίνεται από το σχήμα η $P[T < -0,765] = 0,25$ και η ζητούμενη τιμή t είναι –0,765.



Σχ. (5.31)

Η $P[|T| < t] = 0{,}98$ γράφεται ως $P[-t < T < t] = 0{,}98$ και συνεπώς η θετική τιμή t έχει δεξιά της πιθανότητα ίση με 0,01και συνεπώς αριστερά της πιθανότητα ίση με 0,99. Άρα η τιμή t βρίσκεται στην τομή της στήλης 0,99 και της γραμμής των 3 βαθμών ελευθερίας και είναι η 4,541.



Σχ. (5.32)

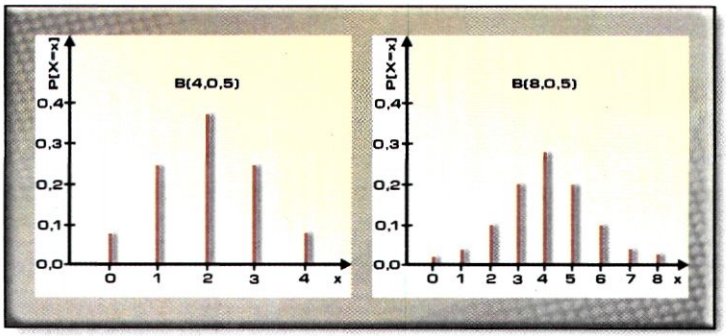
δ) Κατά ανάλογο τρόπο αντιμετωπίζεται και η περίπτωση δ την οποία αφήνουμε στους μαθητές για άσκηση.

5.10 Κανονική προσέγγιση στη Διωνυμική Κατανομή

Στην παράγραφο των ειδικών διακριτών κατανομών μελετήσαμε λεπτομερώς τη διωνυμική κατανομή B(ν, p), όπου ν ο αριθμός των επαναλήψεων του πειράματος και p η σταθερή πιθανότητα επιτυχίας. Θυμίζουμε ότι αν X η διακριτή τ.μ. που ακολουθεί διωνυμική κατανομή, τότε : $P[X = x] = \binom{\nu}{x} p^x q^{\nu - x}$ $x = 0, 1, 2, \ldots, \nu$ και

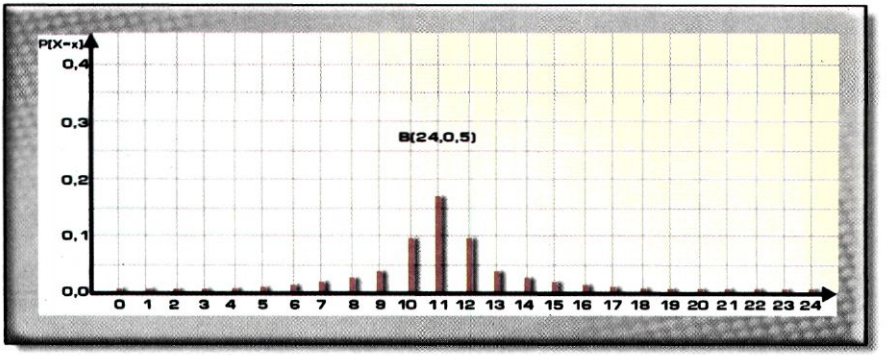
[B(ν, p)] $q = 1 - p$

Θα δούμε τώρα, κάτω από ποιες προϋποθέσεις και με ποιο τρόπο, πιθανότητες που υπολογίζονται από την B(ν, p), μπορούν να εκτιμηθούν με τη χρήση κανονικής κατανομής $N(\mu,\sigma^2)$. Ας δούμε τα διαγράμματα πιθανότητας μερικών διωνυμικών κατανομών με (p = 0,5 , ν = 4), (p = 0,5 , ν = 8), (p = 0,5 , ν = 24), Σχήματα (5.33).



x	P[X=x]
0	$\binom{4}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^4=0,062$
1	$\binom{4}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^3=0,250$
2	$\binom{4}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2=0,375$
3	$\binom{4}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^1=0,250$
4	$\binom{4}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^0=0,062$

x	P[X=x]
0	$\binom{8}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^4=0,004$
1	$\binom{8}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^3=0,031$
2	$\binom{8}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2=0,109$
3	$\binom{8}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^1=0,219$
4	$\binom{8}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^0=0,273$
5	$\binom{8}{5}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^3=0,219$
6	$\binom{8}{6}\left(\frac{1}{2}\right)^6\left(\frac{1}{2}\right)^2=0,109$
7	$\binom{8}{7}\left(\frac{1}{2}\right)^7\left(\frac{1}{2}\right)^1=0,031$
8	$\binom{8}{8}\left(\frac{1}{2}\right)^8\left(\frac{1}{2}\right)^0=0,004$

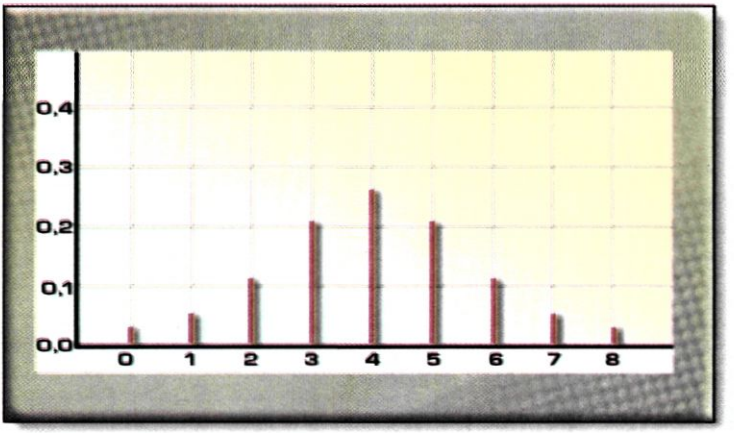


Σχ. (5.33)

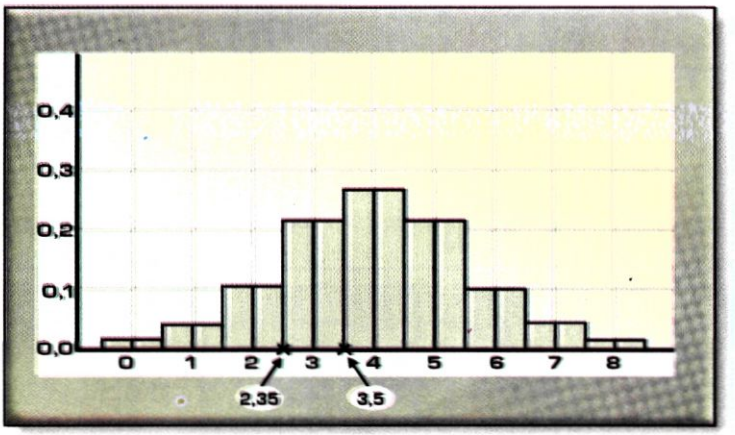
Κάθε μία αυτών των διωνυμικών κατανομών έχει p = 0,5, ενώ όσο το ν αυξάνεται η κατανομή εμφανίζεται να πλησιάζει τη μορφή της κανονικής κατανομής.

Για να επιτύχουμε την επιθυμητή προσέγγιση λαμβάνουμε υπόψη μας ότι η διωνυμική μεταβλητή X είναι διακριτή, ενώ η κανονική μεταβλητή είναι συνεχής.

Ας πάρουμε για ευκολία το διάγραμμα πιθανότητας στη διωνυμική κατανομή $B(8, \frac{1}{2})$ και ας υποθέσουμε ότι οι τιμές του X αποτελούν τα κέντρα κλάσεων μοναδιαίου πλάτους, με συνέπεια το διάγραμμα να μπορεί να αντικατασταθεί από το αντίστοιχο ιστόγραμμα.



Σχ. (5.34) Διάγραμμα πιθανότητας της B(8, 1/2)



Σχ. (5.35) Ιστόγραμμα κατά την υπόθεση ότι οι x είναι κέντρα κλάσεων

Αν υπολογίσουμε π.χ. την P[X=3] χρησιμοποιώντας την B(8, $\frac{1}{2}$), βρίσκουμε :

$$P[X=3]=\binom{8}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^5=\frac{8!}{3!5!}\cdot\frac{1!}{2^8}=0{,}219$$

Στην περίπτωση του ιστογράμματος η πιθανότητα που αντιστοιχεί στην τιμή x=3, δίνεται από το εμβαδόν του ιστού με βάση πλάτους 1 μονάδας και μέσο της βάσης την τιμή x = 3.

Το εμβαδόν του συγκεκριμένου ιστού ισούται με το γινόμενο της βάσης x ύψος = 1 · 0,219 = 0,219.

Να παρατηρήσουμε ότι για x=3 ο ιστός αρχίζει στο 2,5 και καταλήγει στο 3,5.

Έτσι, η P[X=3] με βάση το ιστόγραμμα, υπολογίζεται από την P[2,5≤x≤3,5].

Η προσθαφαίρεση του 0,5 στην τιμή της τ.μ. X καλείται «**διόρθωση συνέχειας**», αποτελεί δε τη **μέθοδο μετατροπής** διακριτής μεταβλητής σε συνεχή.

Ποια όμως είναι η κανονική κατανομή που σχετίζεται με αυτή την περίπτωση;

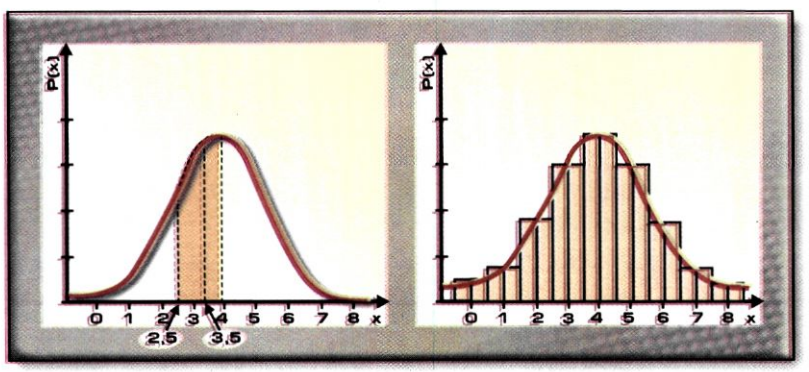
Αφού προσεγγίζουμε διωνυμική κατανομή με κανονική είναι επιθυμητό ο μέσος μ της κανονικής να ισούται με το μέσο E[X] = np της διωνυμικής, καθώς και η διακύμανση της κανονικής κατανομής να ισούται με τη διακύμανση Var[X]=npq της διωνυμικής.

Για το παράδειγμα μας έχουμε:

$\mu = np = 8\cdot 0{,}5 = 4$

$\sigma = \sqrt{npq} = \sqrt{8\cdot 0{,}5\cdot 0{,}5} = \sqrt{2}$

και η πιθανότητα της x = 3 προσεγγίζεται από το εμβαδόν κάτω από την κανονική καμπύλη μεταξύ x = 2,5 και x = 3,5, όπως φαίνεται στο Σχ. (5.36).



Σχ. (5.36)
Η N(np, npq) = N(4, 2)

Σχ. (5.37)
Κανονική κατανομή ως προσέγγιση διωνυμικής

Το Σχ. (5.37) δείχνει τη συνολική κατανομή της διωνυμικής μεταβλητής X, μαζί με κανονική κατανομή ίδιου μέσου και διακύμανσης. Σημειώστε ότι το ιστόγραμμα και η καμπύλη καλύπτουν το ίδιο σχεδόν εμβαδόν.

Η πιθανότητα η τ.μ. X να βρίσκεται μεταξύ 2,5 και 3,5 βρίσκεται κατά το γνωστό τρόπο:

$$P[2,5 < X < 3.5] = P\left[\frac{2,5-4}{\sqrt{2}} < \frac{X-4}{\sqrt{2}} < \frac{3.5-4}{\sqrt{2}}\right] =$$

$$= P[-1,06 < Z < -0,35] = P[Z < -0,35] - P[Z < -1,06] =$$
$$= \{1 - P[Z < 0,35]\} - \{1 - P[Z < 1,06]\} =$$
$$= P[Z < 1,06] - P[Z < 0,35] =$$
$$= 0,8554 - 0,6368$$
$$= 0,2186$$

Συγκρίνοντας τη διωνυμική πιθανότητα P[X=3] = 0,219 με την κανονική πιθανότητα P[2,5≤X≤3,5] = 0,2186, διαπιστώνουμε την εξαιρετική προσέγγιση που παρέχει η κανονική κατανομή.

Η κανονική προσέγγιση στη διωνυμική κατανομή είναι χρήσιμη όχι μόνο στην περίπτωση που η p της διωνυμικής είναι ίση με $\frac{1}{2}$.

Για την αποτελεσματική χρήση της κανονικής προσέγγισης στη διωνυμική κατανομή χρησιμοποιούμε τον ακόλουθο ΚΑΝΟΝΑ :

Η κανονική κατανομή προσεγγίζει με ικανοποιητικό βαθμό ακριβείας τη διωνυμική B(ν,p), όταν νp και ν(1-p) είναι και τα δύο μεγαλύτερα ή ίσα του 5.

Παράδειγμα 21°:

Απρόβλεπτη μη εντοπισμένη μηχανική βλάβη μηχανής παραγωγής μικροαντιστάσεων παράγει σε 5000 μονάδες το 1/3 ελαττωματικές.
Ποια είναι η πιθανότητα ο μηχανικός ποιοτικού ελέγχου να βρει όχι περισσότερες από 3 ελαττωματικές μικροαντιστάσεις σε τυχαίο δείγμα ν = 25;

Απάντηση

Αν Χ η τ.μ. που συμβολίζει τον αριθμό των ελαττωματικών στο δείγμα των ν = 25 μικροαντιστάσεων και p = P[ελαττωματικής μικροαντίστασης] = $\frac{1}{3}$, τότε :

P[όχι περισσότερες από 3 ελαττωματικές μικροαντιστάσεις σε δείγμα ν = 25] =

$$= P[X = 0] + P[X = 1] P[X = 2] + P[X = 3] =$$

$$= \binom{25}{0}\left(\frac{1}{3}\right)^0\left(\frac{2}{3}\right)^{25} + \binom{25}{1}\left(\frac{1}{3}\right)^1\left(\frac{2}{3}\right)^{24} + \binom{25}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^{23} + \binom{25}{3}\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^{22}$$

Είναι φανερό ότι ο υπολογισμός των επι μέρους πιθανοτήτων είναι χρονοβόρος και επίπονος, λόγω του μεγάλου εκθέτη που αντιστοιχεί στην

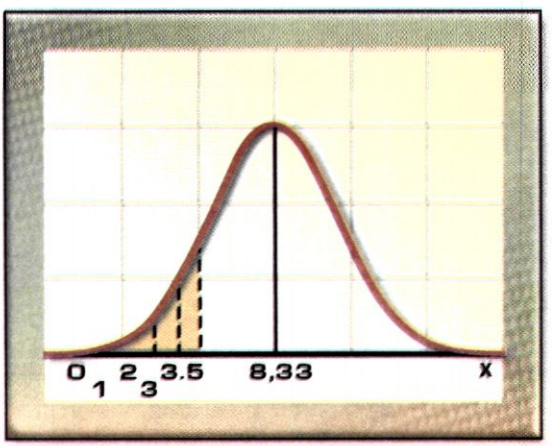
$q = 1 - p = \frac{2}{3}$.

Παρατηρούμε όμως ότι : $np = 25 \cdot \frac{1}{3} = 8{,}33 > 5$ και

$$nq = 25 \cdot \frac{2}{3} = 16{,}66 > 5$$

Σύμφωνα με τον κανόνα, είναι επιτρεπτή η χρήση της κανονικής κατανομής ως προσέγγιση της διωνυμικής $B(25, \frac{1}{3})$.

Άρα η P[X ≤ 3] θα βρεθεί με βάση την N(np, npq), δηλαδή τη $N(25 \cdot \frac{1}{3}, 25 \cdot \frac{1}{3} \cdot \frac{2}{3})$, N(8,33 , 5,55).



Έτσι έχουμε :

P[X≤3]=P[X<3,5]=

$$P\left[\frac{X - 8{,}33}{\sqrt{5{,}55}} < \frac{3{,}55 - 8{,}33}{\sqrt{5{,}55}}\right] =$$

= P[Z < -2,05] =

= 1 – 0,9798 = 0,0202

(☆) 5.11 Η κατανομή $X^2_{(\nu)}$

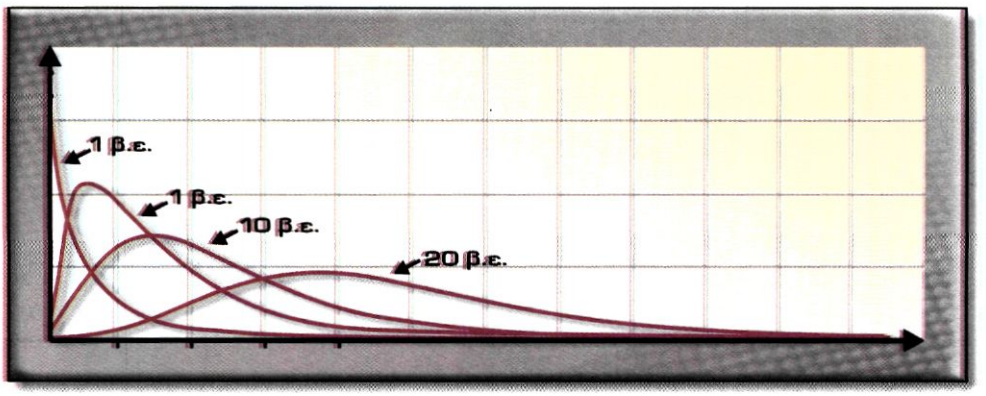
Αποτελεί κατανομή που παράγεται από τη τυποποιημένη κανονική κατανομή με τον ακόλουθο τρόπο.

Αν $X_1, X_2, \ldots, X_\nu$, ν συνεχείς τ.μ. *που κάθε μία έχει* την τυποποιημένη κανονική κατανομή N(0,1), τότε :

Η τ.μ. $Y=X_1^2+X_2^2+\ldots+X_\nu^2$ *ακολουθεί* κατανομή γνωστή ως «*χι*-τετράγωνο κατανομή» με ν βαθμούς *ελευθερίας και* συμβολίζεται ως $Y\sim X^2_{(\nu)}$. *Αν μια* τ.μ Y έχει $X^2_{(\nu)}$ κατανομή, *αποδεικνύεται ότι*:

$$E[Y]=\nu \text{ και } \mathrm{Var}[Y]=2\nu$$

Στο σχήμα *5*.39 δίνονται μορφές της $X^2_{(\nu)}$ *για* μερικές τιμές του ν. *Για* ν>2, η κορυφή της κατανομής έχει προβολή το σημείο (ν-2, 0).



Σχ. (5.39) Η μορφή της $X^2_{(\nu)}$ *για μερικές τιμές του ν*

Ιδιότητες της $X^2_{(\nu)}$ κατανομής

- *Αν* η τ.μ. $Z\sim N(0, 1)$, τότε η $Y=Z^2\sim \mathbf{X^2_{(1)}}$
- *Αν* η τ.μ. $T\sim X^2_{(\nu)}$ *και* η $Y\sim X^2_{(\lambda)}$ τότε $T+Y\sim X^2_{(\nu+\lambda)}$
- *Από τον ορισμό της* $X^2_{(\nu)}$, *γίνεται φανερό ότι οι τιμές* της τ.μ. Y *που έχει* $X^2_{(\nu)}$ *κατανομή είναι* θετικές. Άρα, η γραφική της παράσταση βρίσκεται στους δύο θετικούς ημιάξονες του συστήματος συντεταγμένων.

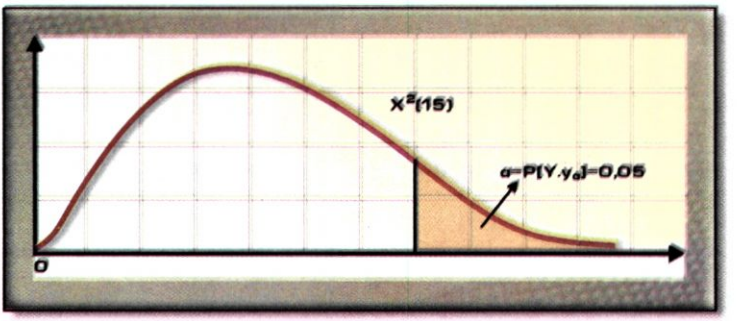
- Παράμετρος της κατανομής είναι το ν. Κάθε τιμή του ν δίνει και διαφορετική μορφή της $X^2_{(\nu)}$ (Σχ. 5.39)
- Η κατανομή είναι ασύμμετρη, με δεξιά ασυμμετρία για ν > 2. (Σχ. 5.39)

Πίνακες της κατανομής $X^2_{(\nu)}$

Είναι αρκετές φορές απαραίτητος στην πράξη, ο υπολογισμός των τιμών y της τ.μ. $Y \sim X^2_{(\nu)}$ σε περιπτώσεις στις οποίες, οι πιθανότητες P[Y>y], P[Y<y] ή $P[y_1<Y<y_2]$ είναι γνωστές.

Ο υπολογισμός αυτός, είναι δυνατός αν γνωρίζουμε τους βαθμούς ελευθερίας ν της κατανομής και έχουμε στη διάθεσή μας κατάλληλους πίνακες, όπως ο πίνακας (Δ) του παραρτήματος.

Ας πάρουμε για παράδειγμα την $X^2_{(15)}$, η μορφή της οποίας είναι :



Σχ. (5.40)

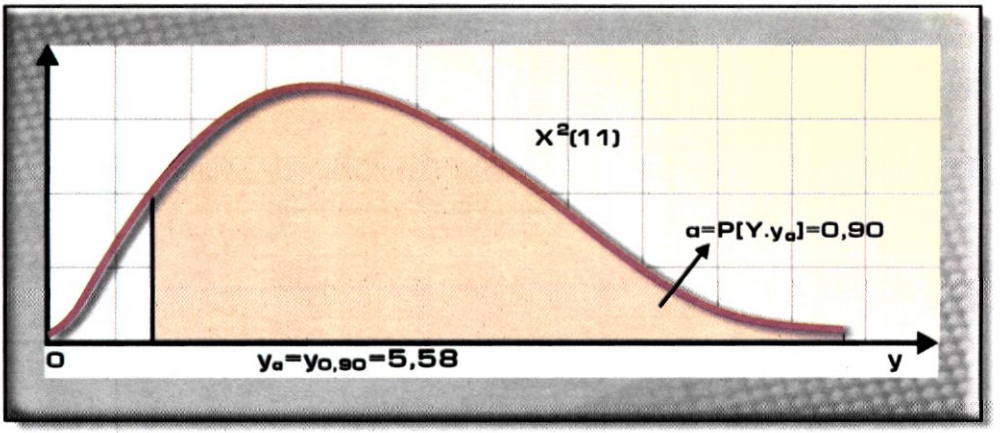
και ας βρούμε για ν=15 και α=0,05 την τιμή $y_{0,05}$ της τ.μ. $Y \sim X^2_{(15)}$.

Πίνακας (5.3) $X^2_{(\nu)}$, α

β.ε.	α ...	0,050	...	0,90
⋮				
11				5,58
⋮				
15		25		
⋮				

Από τον πίνακα (5.3) στην τομή της στήλης α=0,050 και της γραμμής β.ε =15, βρίσκουμε την τιμή y = 25.

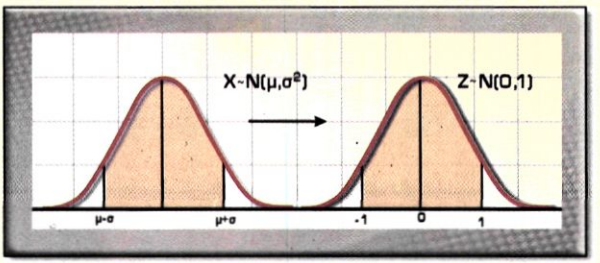
Ας επιχειρήσουμε τώρα χρησιμοποιώντας τον πίνακα, να βρούμε για ν = 11 και α = 0,90 την τιμή του y. Με ανάλογη διαδικασία βρίσκουμε y = 5,58 που είναι η τιμή της στήλης α = 0,90 και της γραμμής β.ε =11.



Στο σχήμα δίνεται η μορφή της $X^2_{(11)}$ και η θέση του y=5,58 στον άξονα.

ΠΕΡΙΛΗΨΗ ΚΕΦΑΛΑΙΟΥ

- $X \sim N(\mu, \sigma^2)$ $\qquad Z = \dfrac{X-\mu}{\sigma} \sim N(0, 1)$
- $E[X] = \mu$, $Var[X] = \sigma^2$ $\qquad E[Z] = 0$, $Var[Z] = 1$
- $P[X < x] = P\left[\dfrac{X-\mu}{\sigma} < \dfrac{x-\mu}{\sigma}\right] = P[Z < z] = \Phi(z)$



Σχ. (5.41 Σχ. (5.40)

- $X \sim N(\mu, \sigma^2)$, $\qquad \overline{X} \sim N(\mu, \dfrac{\sigma^2}{\nu})$
- $E[X] = \mu$, $Var[X] = \sigma^2$ $\qquad E[\overline{X}] = \mu$, $Var[\overline{X}] = \dfrac{\sigma^2}{\nu}$
- $T = \dfrac{\overline{X}-\mu}{S/\sqrt{\nu}} \sim t_d$ – κατανομή , όπου $d = \nu - 1$
- $E[T] = 0$, $Var[T] = \dfrac{d}{d-2}$, $\quad d > 2$
- Αν $X \sim B(\nu, p)$ με $\nu p > 5$ και $\nu(1-p) > 5$, τότε :

  $P[X = x] \cong P[x - \dfrac{1}{2} < X < x + \dfrac{1}{2}]$

  όπου $X \sim N(\nu p, \nu pq)$ με $q = 1 - p$

ΑΣΚΗΣΕΙΣ

ΟΜΑΔΑ Α

Στις ασκήσεις από 1 έως 5 η τυχαία μεταβλητή Z ακολουθεί την τυποποιημένη κανονική κατανομή με μέσο $\mu = 0$ και διακύμανση $\sigma^2 = 1$.
Χρησιμοποιήστε τον πίνακα της N(0,1) Πιν. (Β) σελ. ... για να απαντήσετε στις ακόλουθες ερωτήσεις.

1. Να βρεθούν :
   α) $P(Z < 1,1)$ β) $P(Z < 1,2)$ γ) $P(Z > 1,8)$
   δ) $P(Z < 0)$ ε) $P(Z > 0)$ στ) $P(Z < -1,4)$
   ζ) $P(Z < -0,8)$ η) $P(Z > -1,5)$

2. Να βρεθούν :
   α) $P(2 < Z < 2,6)$ β) $P(2,5 < Z < 2,8)$
   γ) $P(-1,2 < Z < 0,8)$ δ) $P(-1,8 < Z < -0,2)$

3. Να βρεθούν :
   α) $P(|Z| < 0,6)$ β) $P(|Z| > 1,2)$ γ) $P(0,6 < |Z| < 2,2)$

4. Να βρεθεί το α έτσι ώστε :
   α) $P(Z<\alpha) = 0,9192$ β) $P(Z<\alpha) = 0,3446$
   γ) $P(Z>\alpha) = 0,8849$ δ) $P(Z>\alpha) = 0,0047$
   ε) $P(1<Z<\alpha) = 0,1039$ στ) $P(\alpha<Z<-0,8) = 0,1760$

5. Να βρεθεί το α έτσι ώστε :
   α) $P(|Z| <\alpha) = 0,4514$
   β) $P(|Z| >\alpha) = 0,1096$

6. Η τ.μ. X~N(12, 9). Να βρεθούν :
   α) $P(X > 15)$ β) $P(X < 16,8)$
   γ) $P(X < 8,4)$ δ) $P(X > 9,6)$

7. Η τ.μ. X~N(50, 100). Να βρεθούν :
   α $P(36 < X < 62)$ β) $P(40 < X < 50)$
   γ) $P(56 < X < 70)$ δ) $P(38< X <, 42)$

8. Η τ.μ. X~N(16, 4). Να βρεθούν :
   α) $P(X>0)$ β) $P(X<-1,6)$
   γ) $P(|X|<2)$ δ) $P(0<X<2)$

9. Η τ.μ. X~N(-4, 25). Να βρεθούν :
   α) $P(X>0)$ β) $P(-5<X<-2)$
   γ) $P(-2<X<1)$ δ) $P(|X|>1)$

10. Το IQ αποτελεί δείκτη ευφυΐας των ατόμων και ακολουθεί κανονική κατανομή με μέσο μ = 100 και διασπορά $\sigma^2 = 225$.
    Να βρεθεί η αναλογία των ατόμων με IQ :
    1) μικρότερο του 118
    2) μεγαλύτερο του 112
    3) μικρότερο του 94
    4) μεγαλύτερο του 73
    5) μεταξύ 100 και 112
    6) μεταξύ 73 και 118
    7) μεταξύ 73 και 94

ΟΜΑΔΑ Β

1. Αν η τ.μ. X~N(μ, 25) και η $P(X>3,5) = 0,970$ να βρεθεί ο μέσος μ.

2. Αν η τ.μ. X~N(μ, 0,5) και η $P(X<-1,2) = 0,05$ να βρεθεί ο μέσος μ.

3. Αν η τ.μ. X~N(32,4 , σ^2) και η $P(X>45,2) = 0,300$ να βρεθεί η διακύμανση σ^2.

4. Αν η τ.μ. X~N(μ, σ^2) και η $P(X>0) = 0,800$ ενώ $P(X<5) = 0,700$. Να βρεθούν τα μ και σ^2.

5. Η μάζα Μ των κουτιών συγκεκριμένης μάρκας καφέ είναι τυχαία μεταβλητή που ακολουθεί κανονική κατανομή με μέσο μ = 250 gr και $\sigma^2 = 100$.
   Να βρεθεί
   α) η διάμεσος και
   β) το τρίτο τεταρτημόριο Q_3.

6. Συγκεκριμένη ποικιλία καλαμποκιού αναπτύσσεται πολύ περισσότερο από άλλες ποικιλίες.

   Αν το ύψος του καλαμποκιού σε μέτρα ακολουθεί $N(\mu, \sigma^2)$, να βρεθούν τα μ και σ^2, αν είναι γνωστό ότι : $P(Y < 1,83) = 0,30$ και $P(Y < 2,31) = 0,70$.

7. Το μήκος σε cm μεταλλικών κυλίνδρων που κατασκευάζει μεταλλουργική βιομηχανία ακολουθεί κανονική κατανομή αγνώστου μέσου και άγνωστης διασποράς σ^2.

   Σε μεγάλο δείγμα μεταλλικών κυλίνδρων βρέθηκε ότι 10% των κυλίνδρων είναι μακρύτεροι των 3,68 cm και 3% κοντύτεροι των 3,52 cm. Να βρεθούν τα μ και σ^2.

8. Οι πόρτες που χρησιμοποιεί μεγάλη κατασκευαστική εταιρία είναι τυποποιημένες σταθερού ύψους 1,83 m.

   Τα ύψη των ανδρών ακολουθούν κανονική κατανομή με μέσο μ = 1,73 m και τυπική απόκλιση σ = 0,064 m.

   α) Ποιο είναι το ποσοστό των ανδρών που είναι ψηλότεροι από την πόρτα ;

   β) Ποιο είναι το ύψος που πρέπει να έχουν οι πόρτες έτσι ώστε το 1/1000 των ανδρών να είναι ψηλότεροι από τις πόρτες ;

   Οι πόρτες πρόκειται να χρησιμοποιηθούν στο χτίσιμο ενός super – market για το οποίο είναι γνωστό ότι στους 20 πελάτες οι 19 είναι γυναίκες.

   Αν η αναλογία των γυναικών που είναι ψηλότερες της πόρτας είναι 0,00069

   γ) Ποιο είναι το ποσοστό πελατών για τους οποίους το ύψος των 1,83 m θα είναι πολύ χαμηλό ;

9. Από το αρχείο καλά οργανωμένου οδοντιατρείου προκύπτει ότι η πιθανότητα αναμονής ενός ασθενή για χρόνο περισσότερο των 20 λεπτών είναι 0,0239.

   Αν ο χρόνος αναμονής ακολουθεί κανονική κατανομή με σ = 3,75 λεπτών

   α) Ποιος είναι ο μέσος χρόνος αναμονής στο οδοντιατρείο ;

   β) Τι ποσοστό ασθενών περιμένει στο οδοντιατρείο μεταξύ 10 και 15 λεπτών;

10. Ο μέσος χρόνος που διαρκεί το ταξίδι για την Αίγινα με πλοίο ανοικτού τύπου (ferry boat) είναι 65 λεπτά με τυπική απόκλιση σ_1 =8 λεπτών και ο μέσος χρόνος για το Αγκίστρι είναι 85 λεπτά με τυπική απόκλιση σ_2 = 9 λεπτών.

    Αν υποθέσουμε ότι οι χρόνοι για τους δύο προορισμούς ακολουθούν κανονικές κατανομές, ποιο ποσοστό ταξιδιών προς την Αίγινα διαρκεί περισσότερο από το μέσο χρόνο ταξιδιού προς το Αγκίστρι;

    Ποιο ποσοστό ταξιδιών προς το Αγκίστρι διαρκεί λιγότερο από το μέσο χρόνο προς την Αίγινα;