ΠΟΣΟΤΙΚΕΣ ΜΕΘΟΔΟΙ

ΘΕΜΑΤΙΚΗ ΕΝΟΤΗΤΑ 7

Η Κανονική Κατανομή

ΧΡΙΣΤΙΝΑ Κ. ΜΠΕΝΕΚΗ, Αναπληρώτρια Καθηγήτρια

- Στην προηγούμενη ενότητα περιγράψαμε τις δύο κυριότερες διακριτές (ασυνεχείς) κατανομές, τη διωνυμική και την Poisson.
- Σε αυτή την ενότητα θα περιγράψουμε την πιο σημαντική κατανομή πιθανοτήτων της στατιστικής, την κανονική κατανομή.

Κανονική Κατανομή
(Normal distribution)

- Στην ιστορία της Θεωρίας των Πιθανοτήτων αναφέρεται ότι η κανονική κατανομή μελετήθηκε αρχικά ως προσέγγιση της διωνυμικής κατανομής. Στη συνέχεια, ο γερμανός μαθηματικός και φυσικός Gauss (1777-1855) έδειξε τη χρησιμότητά της στην περιγραφή της μεταβλητότητας των σφαλμάτων μέτρησης. Για αυτό η κανονική κατανομή αναφέρεται και ως «Νόμος των Σφαλμάτων» ή κατανομή του Gauss (Gaussian distribution).

- Εξάλλου η ονομασία της ως κανονική αποδίδεται στο ότι χρησιμοποιήθηκε σε πολλές ανθρωπομετρικές μελέτες (μετρήσεις του ύψους, βάρους, διαστάσεων διαφόρων οργάνων του ανθρώπινου σώματος, κ.λπ.), στις οποίες η μέση τιμή αναφέρεται ως τιμή του «μέσου» ή «κανονικού» ανθρώπου και η μεταβλητότητα γύρω απ’ αυτήν περιγράφεται από τον Νόμο των Σφαλμάτων.

- Για να υπολογίσουμε τις βασικές παραμέτρους (μέσο και τυπική απόκλιση), καθώς και τις πιθανότητες μιας συνεχούς κατανομής, χρειαζόμαστε τα εργαλεία του ολοκληρωτικού λογισμού (ορισμένα ολοκληρώματα, κ.λπ.).
- Όμως η διαδεδομένη χρήση της κανονικής κατανομής οδήγησε εδώ και πολλά χρόνια τους αναλυτές να πινακοποιήσουν τις πιθανότητες, έτσι ώστε οι απλοί χρήστες να μη χρειάζονται πολύπλοκους μαθηματικούς υπολογισμούς για να υπολογίσουν μια πιθανότητα. Οπότε είτε με χρήση έτοιμων πινάκων είτε με χρήση υπολογιστών μπορούμε να έχουμε γρήγορες και ακριβείς εκτιμήσεις.

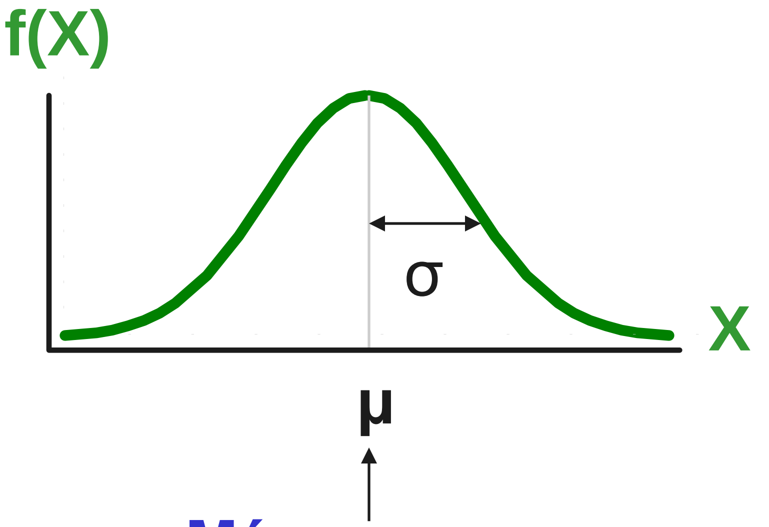
Κανονική Κατανομή

- Ακολουθεί το σχήμα καμπάνας (Bell Shaped)
- Συμμετρική (Symmetric)
- Μέσος, Διάμεσος και Σημείο Μέγιστης Συχνότητας (Επικρατούσα τιμή) είναι ίσες

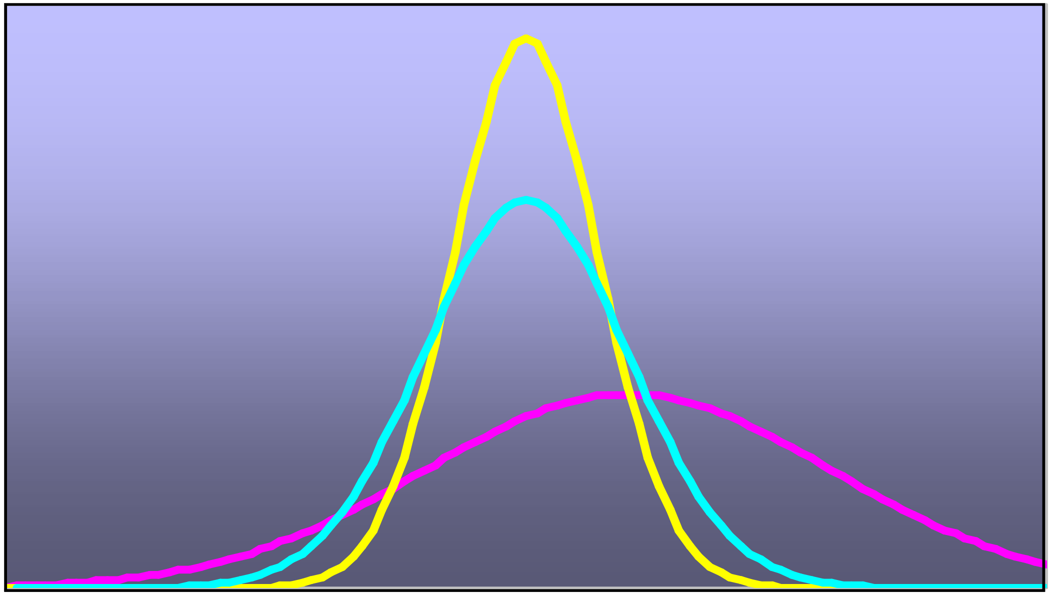
μ, ο μέσος του πληθυσμού

σ, η τυπική απόκλιση του πληθυσμού

Η τυχαία μεταβλητή X παίρνει τιμές στο διάστημα :
$+\infty$ έως $-\infty$



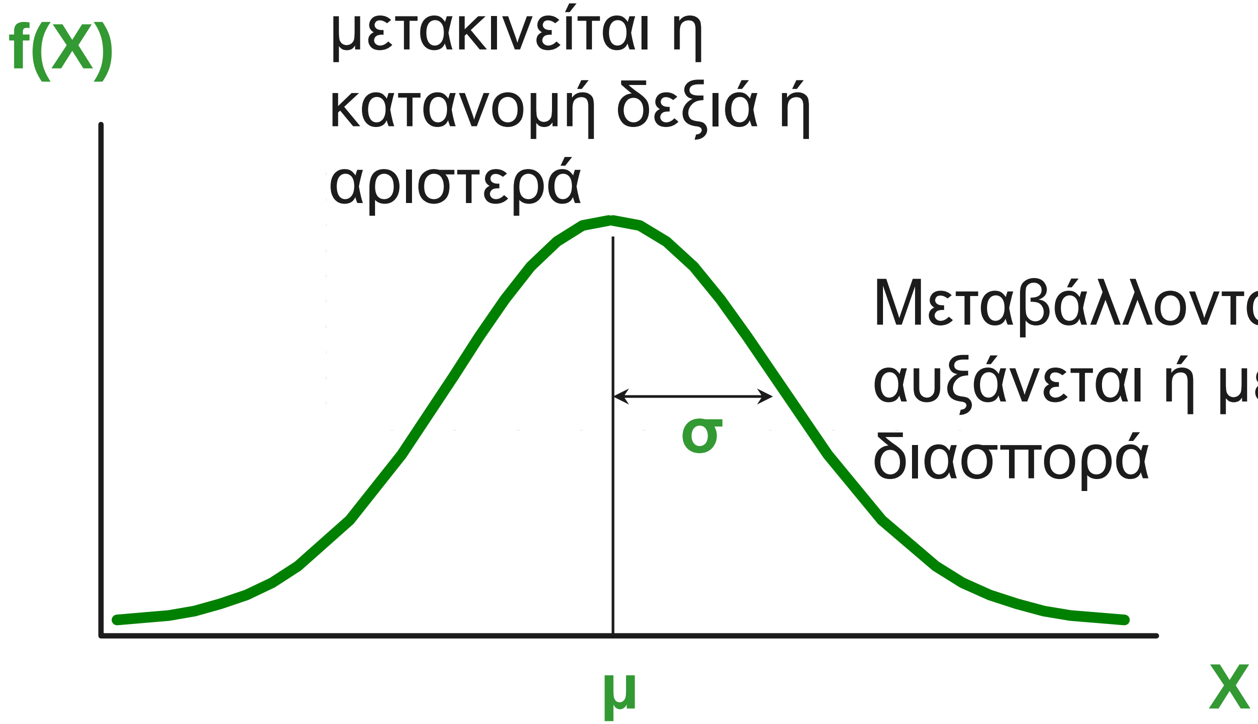
Μέσος
= Διάμεσος
= Επικρατούσα τιμή



Κάθε φορά που ορίζουμε έναν συγκεκριμένο συνδυασμό μ και σ, έχουμε και μια διαφορετική κανονική κατανομή

Αυτό που ενδιαφέρει δεν είναι η τιμή της X, αλλά η απόσταση της τιμής από τον μέσο σε όρους της μέσης απόκλισης σ

Μεταβάλλοντας το **μ** μετακινείται η κατανομή δεξιά ή αριστερά



Μεταβάλλοντας την σ αυξάνεται ή μειώνεται η διασπορά

Συνάρτηση πυκνότητας πιθανότητας της κανονικής κατανομής

- Η μαθηματική έκφραση της συνάρτησης πυκνότητας πιθανότητας της κανονικής κατανομής (normal probability density function) είναι η εξής:

$$f(X) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(1/2)[(X-\mu)/\sigma]^2}$$

όπου e = η βάση των νεπέρειων λογάριθμων 2.71828

π = η γνωστή μαθηματική σταθερά 3.14159

μ = ο μέσος του πληθυσμού

σ = τυπική απόκλιση του πληθυσμού

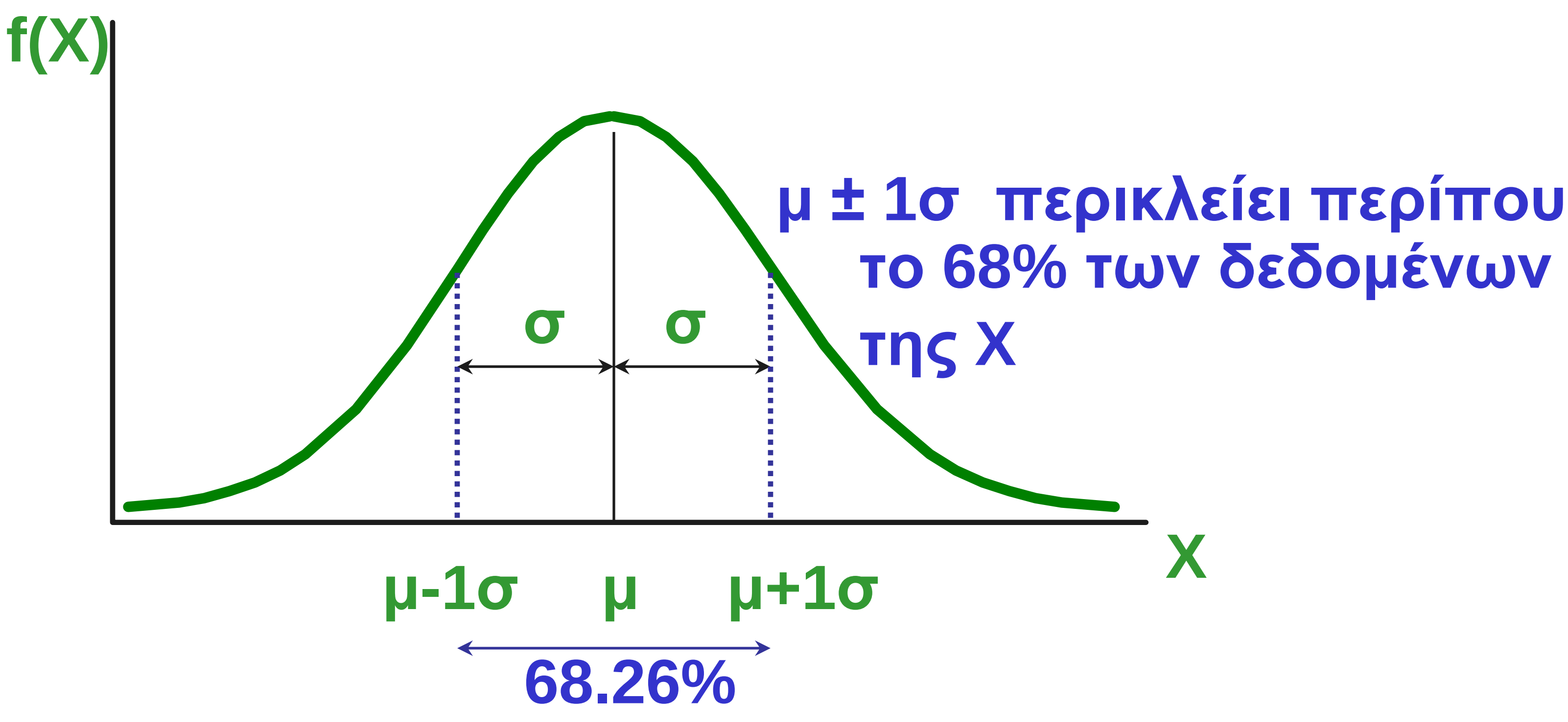
X = τιμές συνεχούς τυχαίας μεταβλητής

Τυποποιημένη Κανονική Κατανομή

- Το κύριο χαρακτηριστικό της κανονικής κατανομής είναι τα συγκεκριμένα ποσοστά συγκέντρωσης των τιμών γύρω από το μέσο που αντιστοιχούν σε συγκεκριμένες αποκλίσεις από τον αριθμητικό μέσο σε όρους της σ.
- Δηλ., αυτό που ενδιαφέρει δεν είναι η τιμή της X, αλλά η απόσταση της τιμής από τον μέσο σε όρους της μέσης απόκλισης σ.

Εμπειρικός κανόνας

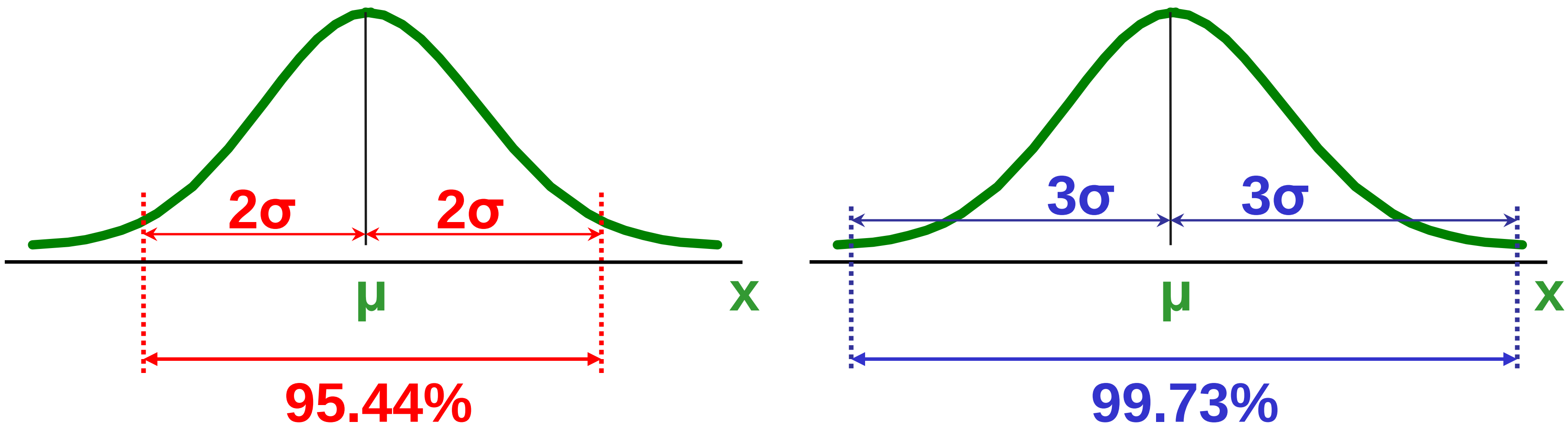
Εάν μια κατανομή συχνοτήτων είναι κανονική τότε ισχύουν τα ακόλουθα:



Εμπειρικός κανόνας

(συνέχεια)

- **μ ± 2σ περικλείει περίπου το 95.4% των δεδομένων της X**
- **μ ± 3σ περικλείει περίπου το 99.7% των δεδομένων της X**



Τυποποιημένη κανονική κατανομή

- Η παρακάτω σχέση δημιουργεί μια νέα μεταβλητή, τη Ζ, που ονομάζεται **τυποποιημένη κανονική κατανομή (standardized normal distribution)**.

$$Z = \frac{X - \mu}{\sigma}$$

Η Ζ κατανομή α) είναι ανεξάρτητη από τη μονάδα μέτρησης της μεταβλητής Χ, αφού δεν εκφράζεται σε καμία μονάδα μέτρησης, β) έχει μέσο = 0 και γ) τυπική απόκλιση = 1

Συνάρτηση πυκνότητας πιθανότητας της τυποποιημένης μεταβλητής

- Η συνάρτηση πυκνότητας πιθανότητας της τυποποιημένης μεταβλητής Z είναι:

$$f(Z) = \frac{1}{\sqrt{2\pi}} e^{-(1/2)Z^2}$$

όπου $e = 2.71828$

$\pi = 3.14159$

Z = τιμές της τυποποιημένης κανονικής κατανομής

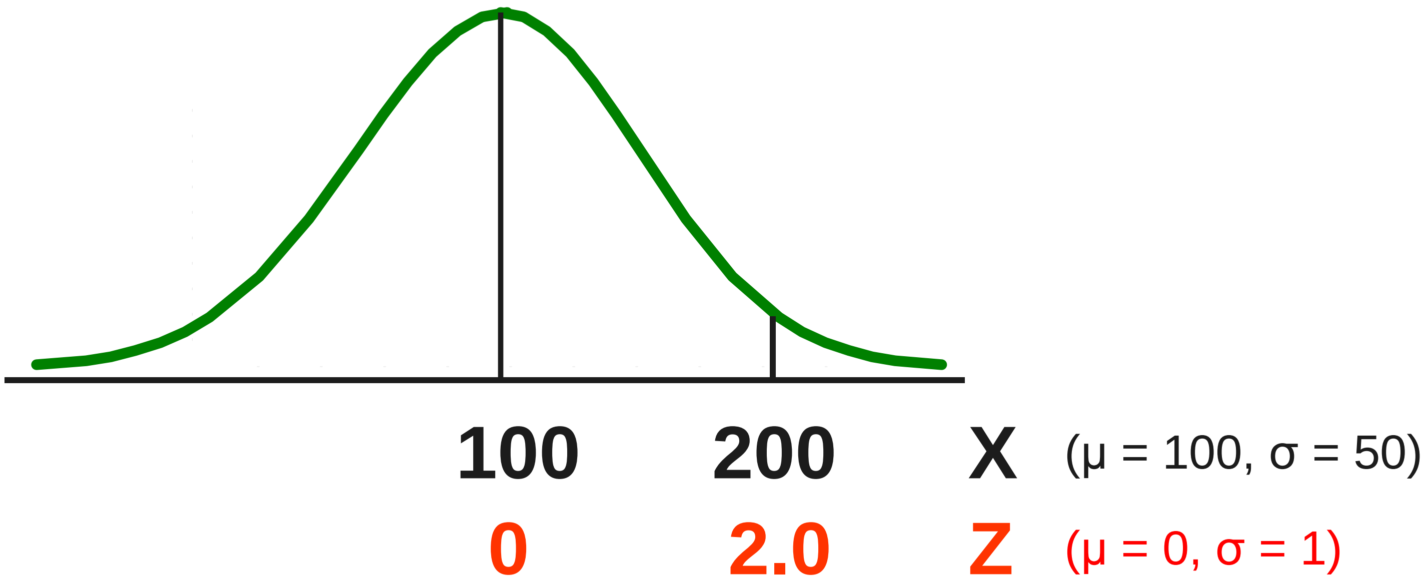
Παράδειγμα

- Εάν η X ακολουθεί την κανονική κατανομή με μέση τιμή 100 και τυπική απόκλιση 50, τότε η τυποποιημένη τιμή της Z που αντιστοιχεί στην τιμή X = 200 είναι

$$Z = \frac{X - \mu}{\sigma} = \frac{200 - 100}{50} = 2.0$$

- Αυτό μας λέει ότι η τιμή X = 200 είναι δύο τυπικές αποκλίσεις 2σ (μ+2σ=100+2*50) από την αντίστοιχη μέση τιμή 100.

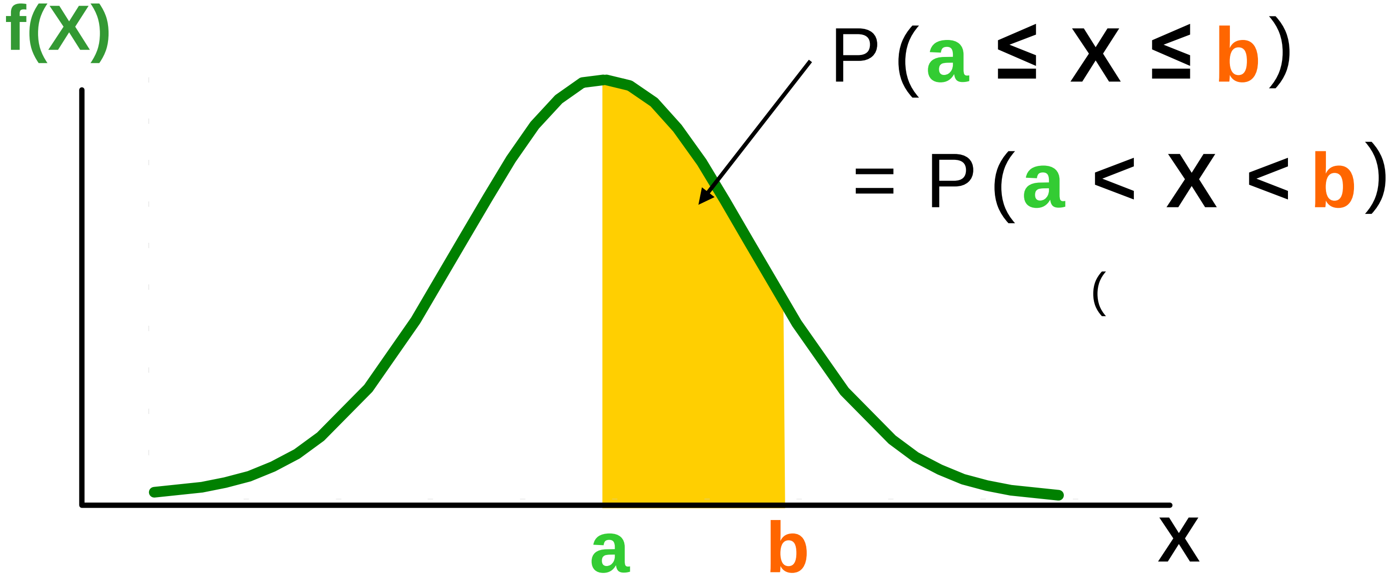
Συγκρίνοντας X και Z



ΣΗΜΕΙΩΣΗ: Η κατανομή είναι η ίδια.

Υπολογισμός πιθανοτήτων της κανονικής κατανομής Χ

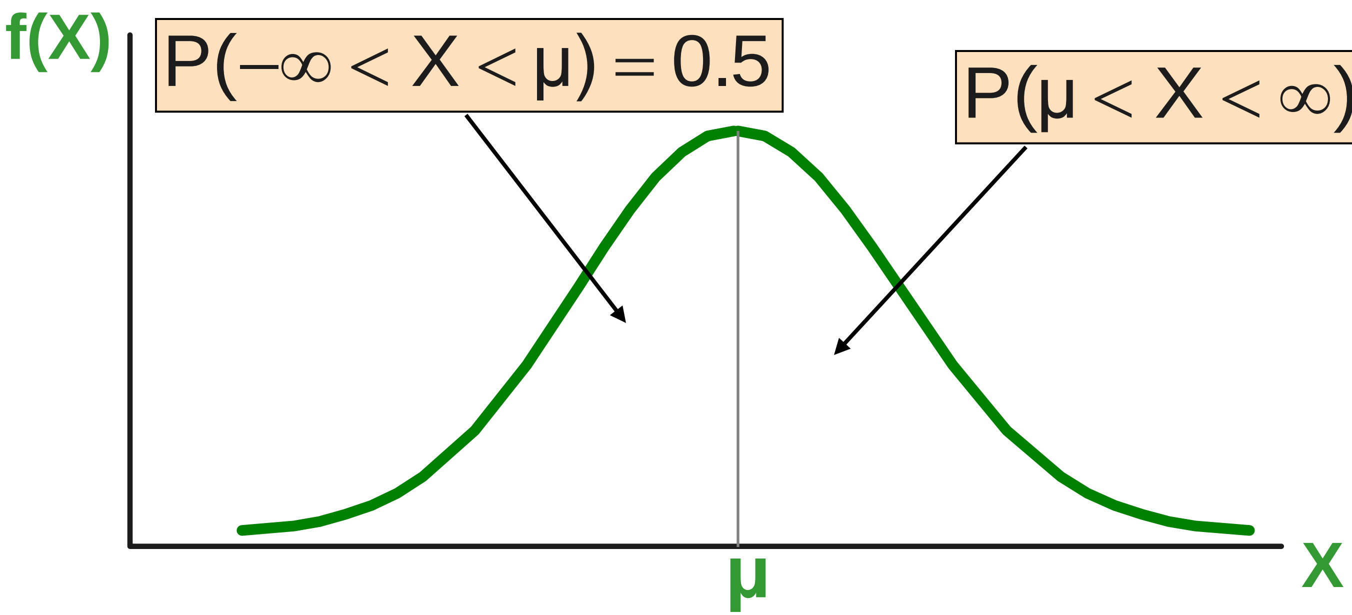
Η πιθανότητα υπολογίζεται ως το εμβαδό κάτω από την καμπύλη



Το συνολικό εμβαδό κάτω από την καμπύλη (total area under the curve) είναι 1, και η καμπύλη είναι συμμετρική, οπότε:

$$P(-\infty < X < \mu) = 0.5$$

$$P(\mu < X < \infty) = 0.5$$



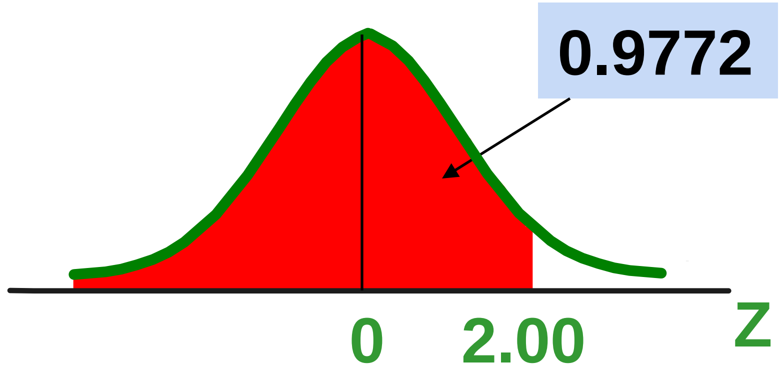
$$P(-\infty < X < \infty) = 1.0$$

Πίνακας τυποποιημένης κατανομής Z

- Ο Πίνακας στα βιβλία Στατιστικής δίνει τις αθροιστικές πιθανότητες της κατανομής της Z (*Cumulative Standardized Normal table*) από + ∞ έως – ∞, δηλ., δίνουν τις πιθανότητες P(Z<z) (με άλλα λόγια, από – ∞ έως Z)

Π.χ.:

$P(Z < 2.00) = 0.9772$



Πίνακας τυποποιημένης κατανομής Ζ

(συνέχεια)

Οι στήλες αναφέρονται στο δεύτερο δεκαδικό ψηφίο

Z	0.00	0.01	0.02 …
0.0			
0.1			
.			
.			
.			
2.0	**.9772**		

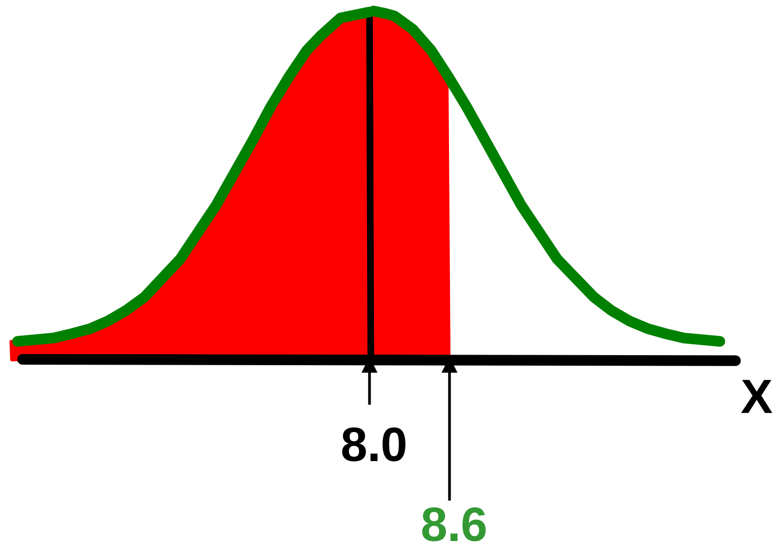
Οι γραμμές αναφέρονται στο πρώτο δεκαδικό ψηφίο

Το εμβαδό της τυποποιημένης κανονικής κατανομής Ζ από – ∞ έως την συγκεκριμένη Ζ τιμή, δηλ. z=2

P(Z < 2.00) = 0.9772

Παράδειγμα

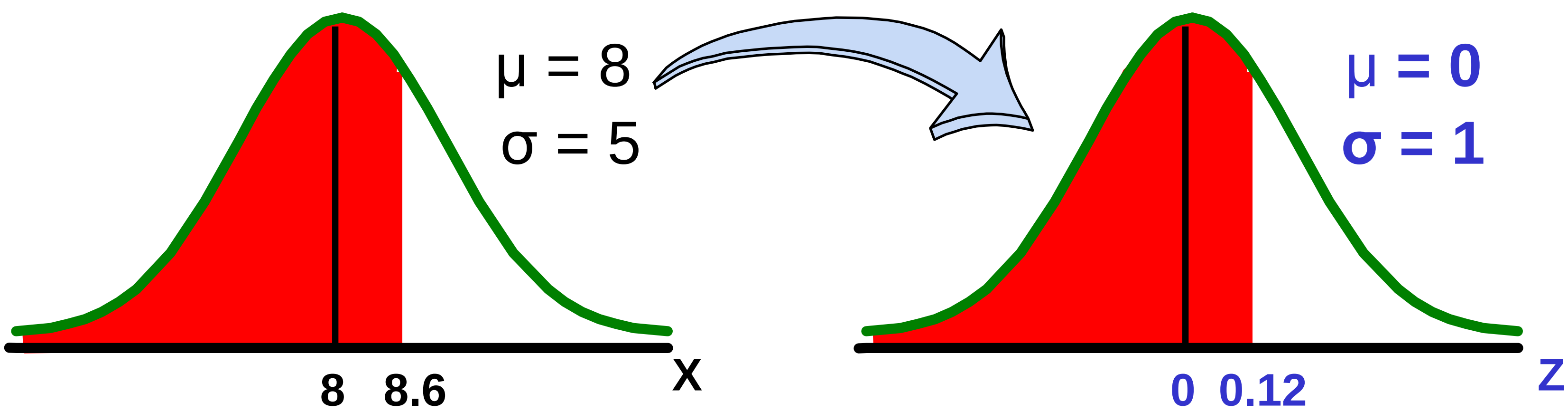
- Έστω ότι η X ακολουθεί την κανονική κατανομή με μ= 8 και σ=5
- Να βρεθεί: $P(X < 8.6)$



Παράδειγμα

(συνέχεια)

$$Z = \frac{X - \mu}{\sigma} = \frac{8.6 - 8.0}{5.0} = 0.12$$



P(X < 8.6)   P(Z < 0.12)

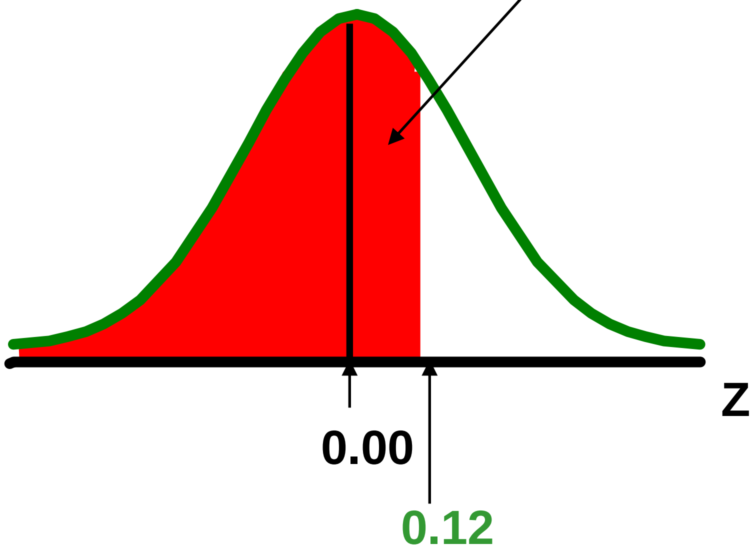
ΛΥΣΗ: Υπολογίζουμε P(Z < 0.12)

Πίνακας τυποποιημένης κατανομής Z

Z	.00	.01	.02
0.0	.5000	.5040	.5080
0.1	.5398	.5438	.5478
0.2	.5793	.5832	.5871
0.3	.6179	.6217	.6255

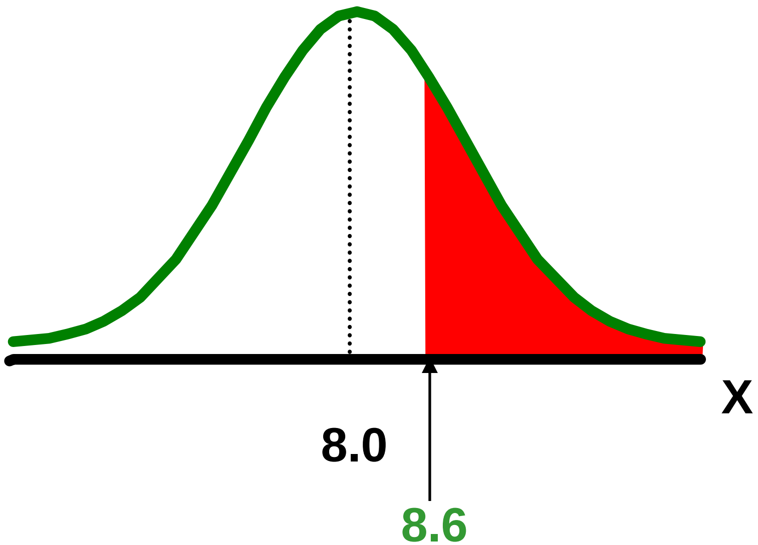
P(X < 8.6)
= P(Z < 0.12)

.5478



Παράδειγμα

- Έστω ότι η X ακολουθεί την κανονική κατανομή με μ= 8 και τυπική απόκλιση σ=5
- Να βρεθεί: $P(X > 8.6)$



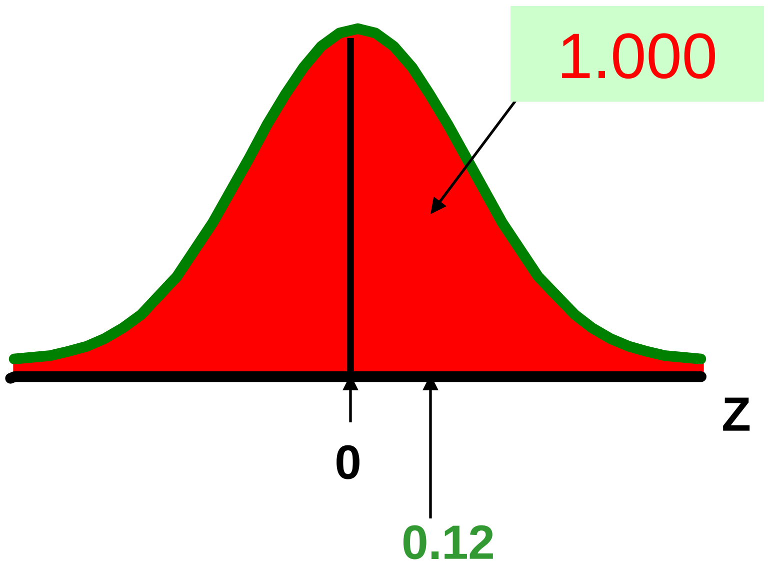
Παράδειγμα

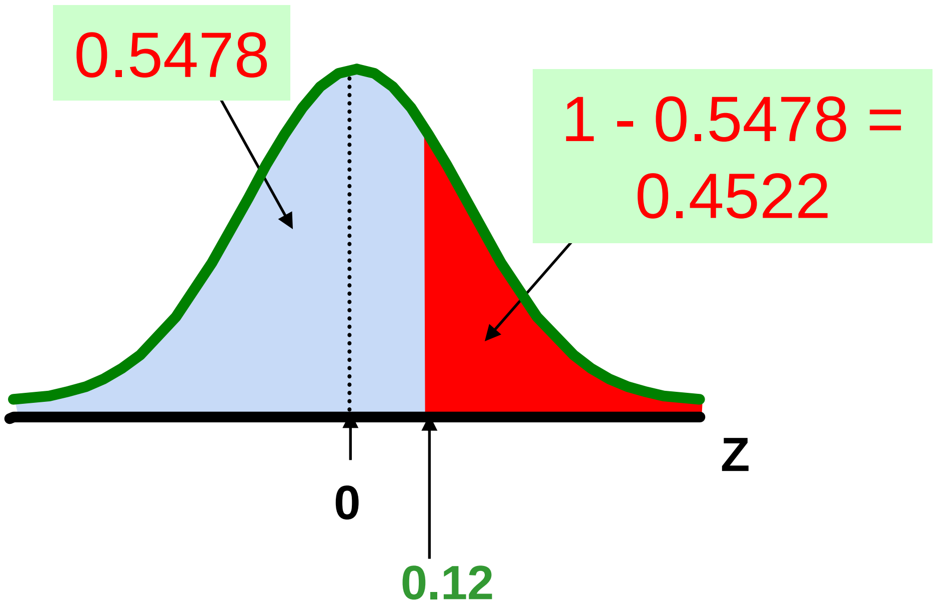
(συνέχεια)

- Να βρεθεί: P(X > 8.6)

P(X > 8.6) = P(Z > 0.12) = 1 - P(Z ≤ 0.12)

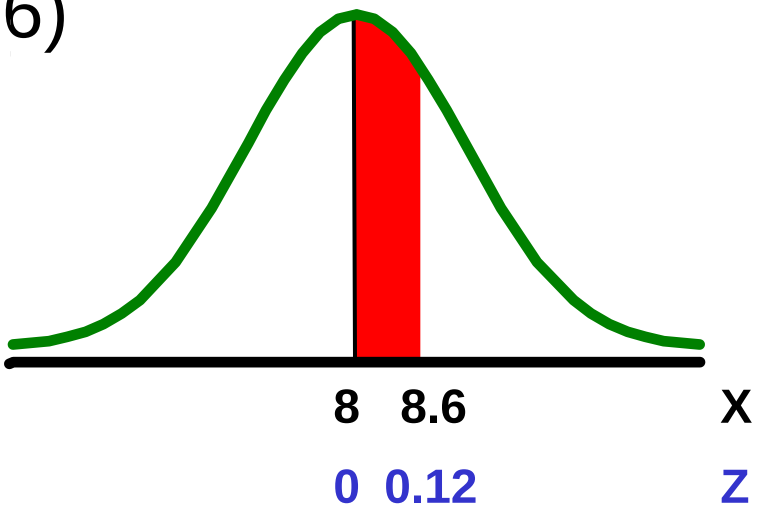
= 1 - 0.5478 = $\boxed{0.4522}$





Παράδειγμα

- Έστω ότι η X ακολουθεί την κανονική κατανομή με **μ= 8** και τυπική απόκλιση σ=5
- Να βρεθεί: P(8 < X < 8.6)



$$Z = \frac{X-\mu}{\sigma} = \frac{8-8}{5} = 0$$

$$Z = \frac{X-\mu}{\sigma} = \frac{8.6-8}{5} = 0.12$$

$$P(8 < X < 8.6)$$

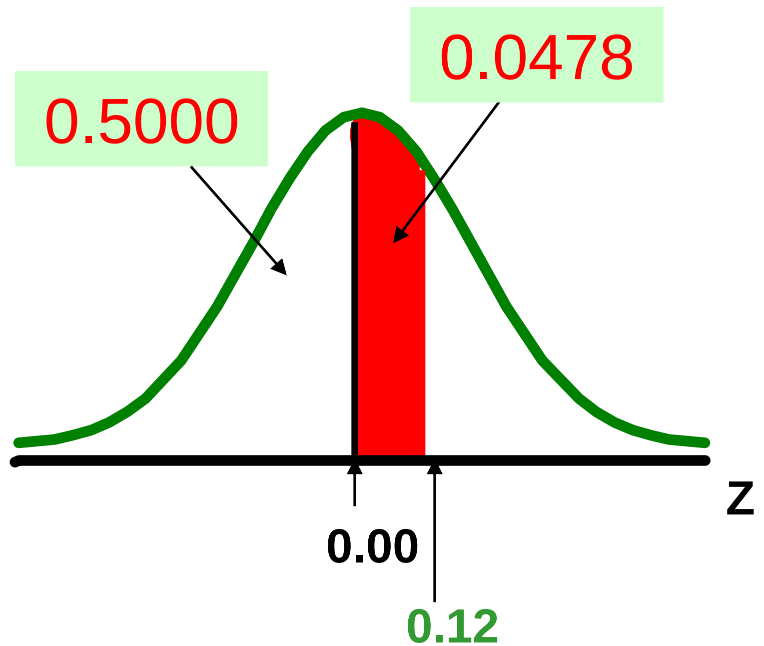
$$= P(0 < Z < 0.12)$$

ΛΥΣΗ: Υπολογίζοντας P(0 < Z < 0.12)

Πίνακας τυποποιημένης κατανομής Z

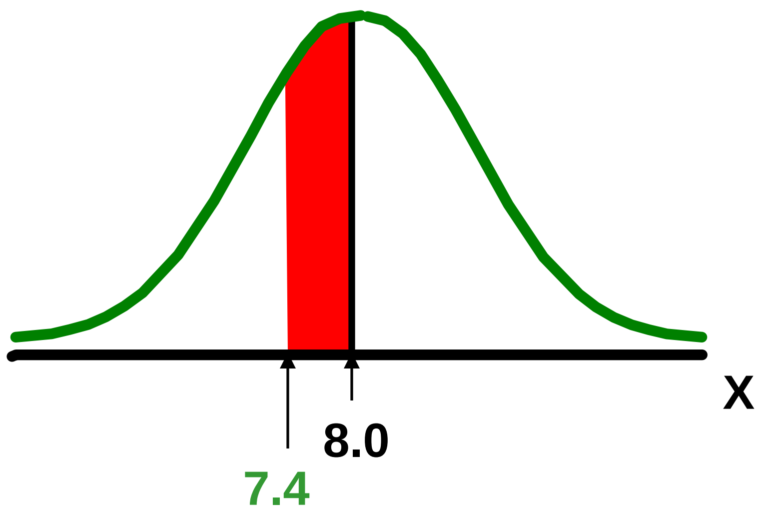
Z	.00	.01	.02
0.0	.5000	.5040	.5080
0.1	.5398	.5438	.5478
0.2	.5793	.5832	.5871
0.3	.6179	.6217	.6255

$P(8 < X < 8.6)$
$= P(0 < Z < 0.12)$
$= P(Z < 0.12) – P(Z \leq 0)$
$= 0.5478 - .5000 = 0.0478$



Παράδειγμα

- Έστω ότι η X ακολουθεί την κανονική κατανομή με **μ= 8** και τυπική απόκλιση σ=5
- Να βρεθεί: P(7.4 < X < 8)



Παράδειγμα

(συνέχεια)

Να βρεθεί: $P(7.4 < X < 8)$

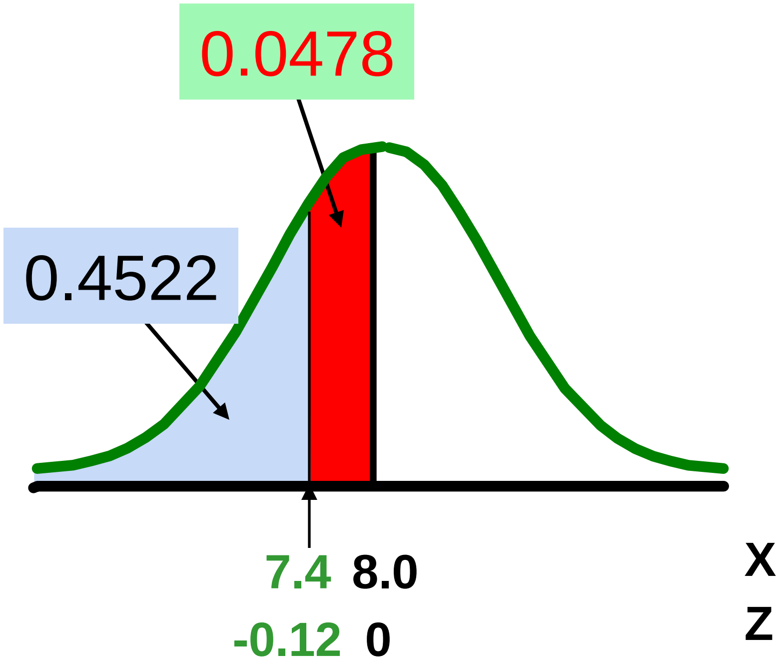
$P(7.4 < X < 8)$

$= P(-0.12 < Z < 0)$

$= P(Z < 0) – P(Z \leq -0.12)$

$= 0.5000 - 0.4522 = \boxed{0.0478}$

Η κανονική κατανομή είναι συμμετρική, οπότε η ζητούμενη πιθανότητα είναι ίδια με την $P(0 < Z < 0.12)$



Παράδειγμα

- Μια εταιρία κατασκευής μπαταριών για κινητά τηλέφωνα παράγει ένα συγκεκριμένο τύπο μπαταρίας με την προδιαγραφή ότι ο μέσος χρόνος διάρκειας της μπαταρίας σε κατάσταση αναμονής είναι 191 ώρες με τυπική απόκλιση περίπου 5 ώρες (δηλ. στην πράξη είναι απίθανο μια μπαταρία να εξαντληθεί πριν από μ-3*σ=191-3*5=176 ώρες ή να διαρκέσει πάνω από μ+3*σ=191+3*5=206 ώρες όταν το τηλέφωνο είναι μόνο σε κατάσταση αναμονής).

Παράδειγμα

(συνέχεια)

- Το τμήμα ποιοτικού ελέγχου έχει συλλέξει σε διάστημα 6 μηνών 5000 παρατηρήσεις από εργαστηριακούς ελέγχους ισάριθμων μπαταριών.

- Από τον πίνακα κατανομής συχνοτήτων των παρατηρήσεων του δείγματος προκύπτει ότι η κατανομή είναι συμμετρική. Θέλουμε να βρούμε:

Παράδειγμα

(συνέχεια)

- (α) Ποια η πιθανότητα μια μπαταρία να έχει διάρκεια μικρότερη από 185 ώρες;

Να βρεθεί: $P(X < 185)=P(Z < (185-191)/5)$

$= P(Z < -1.2) = 0.11507$ (ή 11.507%)

- (β) Ποια η πιθανότητα μια μπαταρία να έχει διάρκεια μεγαλύτερη από 195 ώρες;

Να βρεθεί: $P(X > 195)=1- P(Z < (195-191)/5)$

$=1- P(Z < 0.78814) = 0.21186$ (ή 21.186%)

Παράδειγμα

(συνέχεια)

- (γ) Ποια η πιθανότητα μια μπαταρία να διαρκέσει μεταξύ 185 και 195 ώρες;

Να βρεθεί: $P(185 < X < 195) =$

$P(Z < (195-191)/5) - P(Z < (185-191)/5) =$

$P(Z < 0.8) - P(Z < -1.2) = 0.78814 - 0.11507 =$

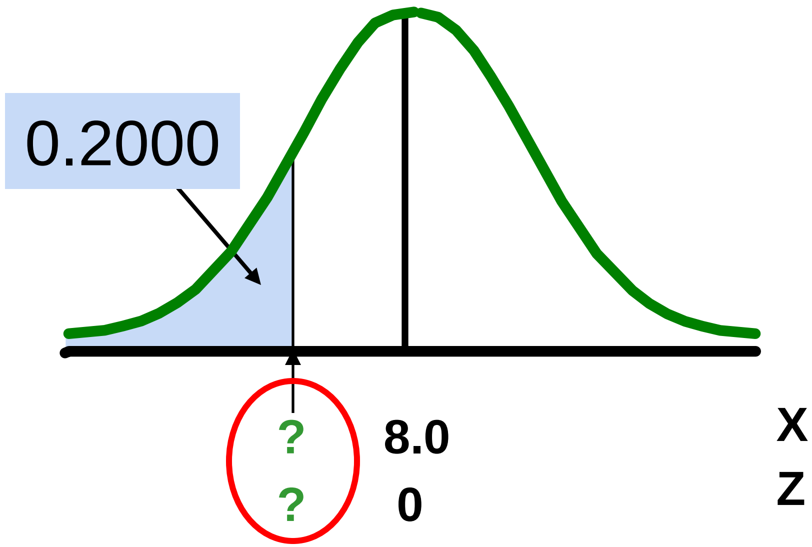
0.67307 (ή 67.307%)

Βρίσκοντας την τιμή x για δοσμένη πιθανότητα

(συνέχεια)

Π.χ.:

- Έστω ότι η X ακολουθεί την κανονική κατανομή με **μ= 8** και τυπική απόκλιση σ=5
- Να βρεθεί η τιμή της X, έτσι ώστε μόνο το 20% όλων των τιμών να είναι κάτω από αυτή την τιμή



Βρίσκοντας την τιμή x για δοσμένη πιθανότητα

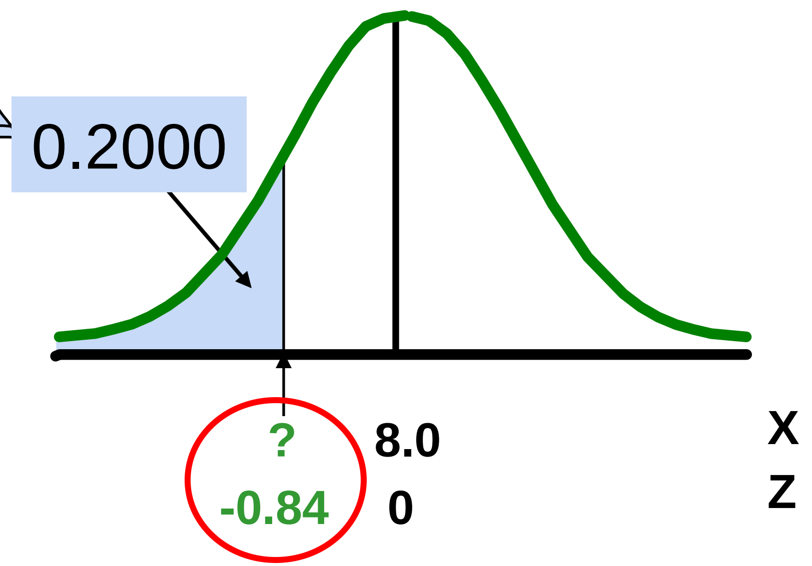
(συνέχεια)

1. Βρίσκουμε την τιμή Z για δοσμένη πιθανότητα

Πίνακας τυποποιημένης κατανομής Z

Z	…	.03	.04	.05
-0.9	…	.1762	.1736	.1711
-0.8	…	.2033	.2005	.1977
-0.7	…	.2327	.2296	.2266

- 20% του εμβαδού αντιστοιχεί στην τιμή Z -0.84



Βρίσκοντας την τιμή x για δοσμένη πιθανότητα

(συνέχεια)

2. Βρίσκουμε το X από τον τύπο:

$$X = \mu + Z\sigma$$

$$= 8.0 + (-0.84)5.0$$

$$= 3.80$$

Συνεπώς, το 20% των τιμών μιας κανονικής κατανομής με μέσο 8.0 και τυπική απόκλιση 5.0 βρίσκεται αριστερά του 3.80.

Παράδειγμα

- Αν το ύψος Χ, σε εκατοστά, ενός ενήλικα άνδρα ακολουθεί την κανονική κατανομή με μέση τιμή 170 εκατ. και τυπική απόκλιση 7,28 εκατ., ποια είναι η τιμή $X=x^*$ για την οποία $P(X>x^*)=0.025$;
- ΛΥΣΗ
- Επειδή $P(X>x^*)=0.025 \leftrightarrow P(X\leq x^*)=1- 0.025$

$= 0.975$ από τον Πίνακα τυποποιημένης κατανομής βρίσκουμε αυτή η πιθανότητα αντιστοιχεί στην τιμή $z = 1.96$

Παράδειγμα

(συνέχεια)

- Επομένως έχουμε:

$X = \mu + Z\sigma$

$= 170 + 1.96 * 7.28$

$= 184.27$

Συνεπώς, η πιθανότητα για έναν ενήλικα άνδρα να έχει ύψος μεγαλύτερο από **184.27 εκατ.** είναι ίση με 2.5%

Παράδειγμα

- Οι βίδες που παράγονται μαζικά από μία μηχανή έχουν διάμετρο Χ, η οποία ακολουθεί την κανονική κατανομή με μέση τιμή ίση με 0.502 εκατ. και τυπική απόκλιση ίση με 0.005 εκατ. Αν ο σκοπός για τον οποίο προορίζονται οι βίδες επιτρέπει μια ανοχή στη διάμετρο από 0.496 εκατ. έως 0.508 εκατ. να προσδιοριστεί το ποσοστό των ελαττωματικών βιδών.

Παράδειγμα

(συνέχεια)

- **Μη ελαττωματικές** είναι οι βίδες στις οποίες η διάμετρος ικανοποιεί τη σχέση $0.496<X<0.508$.
- Η πιθανότητα αυτού του ενδεχομένου είναι:
- $P(0.496<X<0.508) =$

$P((0.496-0.502)/0.005 < Z < (0.508-0.502)/0.005)=$

$P(-1.2 < Z < 1.2) = P(Z < 1.2) – P(Z > 1.2) =$

$P(Z < 1.2) – [1 – P(Z \leq 1.2)] =$

$2P(Z < 1.2) – 1 = 0.7698$, δηλ. το **ποσοστό των μη ελαττωματικών στο σύνολο της παραγωγής** ισούται με 76.98% και άρα το ποσοστό των μη ελαττωματικών ισούται με 23.02%.

Η κανονική κατανομή ως προσέγγιση της διωνυμικής

- Ένα από τα βασικά πλεονεκτήματα της κανονικής κατανομής είναι ότι μπορεί να προσεγγίσει με μεγάλη ακρίβεια τις ασυνεχείς κατανομές, παρά το γεγονός ότι ως συνεχής κατανομή περιγράφει μόνο συνεχείς μεταβλητές. Βέβαια, η προσέγγιση αυτή ισχύει για τις περιπτώσεις εκείνες που και οι ασυνεχείς κατανομές τείνουν να πάρουν το «κανονικό» σχήμα.
- Για παράδειγμα η διωνυμική τείνει στην κανονική για μέγεθος δείγματος (n) μεγαλύτερο του 20. Ενώ για μικρότερα δείγματα η πιθανότητα p θα πρέπει να είναι κοντά στο 0.5
- Επίσης, η Poisson πλησιάζει την κανονική κατανομή όσο αυξάνει ο μέσος λ.

Η κανονική κατανομή ως προσέγγιση της διωνυμικής

(συνέχεια)

- Γενικός κανόνας:
  - Η κανονική κατανομή μπορεί να προσεγγίσει τη διωνυμική, εάν:

$$np \geq 5$$

και

$$n(1 - p) \geq 5$$

Η κανονική κατανομή ως προσέγγιση της διωνυμικής

(συνέχεια)

- Παράδειγμα:
  - X είναι διακριτή (ασυνεχής) σε μια διωνυμική κατανομή, οπότε η $P(X = 4)$ μπορεί να προσεγγιστεί με μια συνεχή κανονική κατανομή, υπολογίζοντας:

$$P(3.5 < X < 4.5)$$

Η κανονική κατανομή ως προσέγγιση της διωνυμικής *(συνέχεια)*

- Ο μέσος και η τυπική απόκλιση μιας διωνυμικής κατανομής, υπολογίζονται ως εξής:

$$\mu = np$$

$$\sigma = \sqrt{np(1-p)}$$

- Μετατρέπουμε τη διωνυμική σε κανονική, χρησιμοποιώντας τον τύπο:

$$\boxed{Z = \frac{X-\mu}{\sigma} = \frac{X-np}{\sqrt{np(1-p)}}}$$

Η κανονική κατανομή ως προσέγγιση της διωνυμικής *(συνέχεια)*

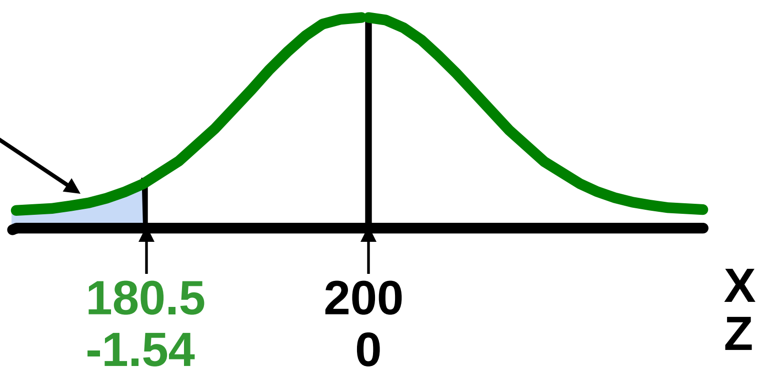
- Εάν n = 1000 και p = 0.2, να βρεθεί $P(X \leq 180)$.
- Για να βρούμε την $P(X \leq 180)$, θα υπολογίσουμε:

$$P(X \leq 180.5)$$

- Μετατρέπουμε τη διωνυμική σε κανονική, χρησιμοποιώντας τον τύπο:

$$Z = \frac{X - np}{\sqrt{np(1-p)}} = \frac{180.5 - (1000)(0.2)}{\sqrt{(1000)(0.2)(1-0.2)}} = -1.54$$

So $P(Z \leq -1.54) = \boxed{0.0618}$



ΕΝΔΕΙΚΤΙΚΗ ΒΙΒΛΙΟΓΡΑΦΙΑ

- 1. ΧΑΛΙΚΙΑΣ ΙΩΑΝΝΗΣ (2010). ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗΣ ΓΙΑ ΕΠΙΧΕΙΡΗΜΑΤΙΚΕΣ ΑΠΟΦΑΣΕΙΣ, ΕΚΔΟΣΕΙΣ ROSILI
- 2. ΖΑΧΑΡΟΠΟΥΛΟΥ ΧΡΥΣΟΥΛΑ (2012). ΣΤΑΤΙΣΤΙΚΗ, ΜΕΘΟΔΟΙ-ΕΦΑΡΜΟΓΕΣ, ΕΚΔΟΣΕΙΣ ΣΟΦΙΑ
- 3. ΚΙΟΧΟΣ ΠΕΤΡΟΣ ΚΑΙ ΚΙΟΧΟΣ ΑΠΟΣΤΟΛΟΣ (2010). ΣΤΑΤΙΣΤΙΚΗ ΓΙΑ Τις ΕΠΙΧΕΙΡΗΣΕΙΣ ΚΑΙ ΤΗΝ ΟΙΚΟΝΟΜΙΑ, ΕΚΔΟΣΕΙΣ ΕΛΕΝΗ ΚΙΟΧΟΥ
- 4. BERENSON M.L., LEVINE D.M., KREHBIEL T.C., Basic Business Statistics (12th Edition), PRENTICE HALL

https://docs.google.com/file/d/0Bxyykc61IUF_b1Zrc2taTnNGbUk/edit

- 5. [για το Εργαστήριο] Χαλικιάς, Μ., Λάλου, Π., Μανωλέσου, Α. 2015. *Μεθοδολογία έρευνας και εισαγωγή στη Στατιστική Ανάλυση Δεδομένων με το IBM SPSS STATISTICS*. [ηλεκτρ. βιβλ.] Αθήνα:Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών. Διαθέσιμο στο: http://hdl.handle.net/11419/5075