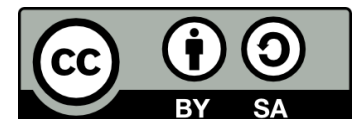




Στατιστική

4^ο Μάθημα: Θεωρητικές και Εμπειρικές - Δειγματοληπτικές Κατανομές

Γεώργιος Μενεξές
Τμήμα Γεωπονίας



Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.
- Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδειας χρήσης, η άδεια χρήσης αναφέρεται ρητώς.



Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «Ανοικτά Ακαδημαϊκά Μαθήματα στο Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.





4^ο Μάθημα

Θεωρητικές και Εμπειρικές-Δειγματοληπτικές Κατανομές

Η παρουσίαση βασίζεται σε υλικό από το έργο του:

- Παπαδημητρίου, Γ. (2001). *Περιγραφική Στατιστική*. Θεσσαλονίκη: Παρατηρητής.

**...και σε σχετική παρουσίαση (2005) του
εκδοτικού οίκου Brooks/Cole, a division of
Thomson Learning, Inc**



Τυχαίες Μεταβλητές

- Τυχαία μεταβλητή λέγεται η **συνάρτηση**, που επιτρέπει την απεικόνιση όλων των δυνατών **αποτελεσμάτων** ενός **πειράματος** στο σύνολο των **πραγματικών αριθμών**.
- Αν οι τιμές, που παίρνει μια τυχαία μεταβλητή X , είναι οποιοσδήποτε πραγματικός αριθμός ενός διαστήματος $[\alpha, \beta]$ ή (α, β) τότε αυτή λέγεται **συνεχής** τυχαία μεταβλητή.
- Αν οι τιμές που παίρνει μια τυχαία μεταβλητή X είναι ένας αριθμός, από κάποιο πεπερασμένο σύνολο διακεκριμένων αριθμών, τότε αυτή λέγεται **διακριτή** τυχαία μεταβλητή, αν δε οι τιμές της είναι από κάποιο άπειρο σύνολο αλλά αριθμήσιμο λέγεται **απαριθμητή**.



Συνάρτηση Πιθανότητας Διακριτής Τ.Μ. (1)

- Αν σε κάθε τιμή της μεταβλητής X αντιστοιχίσουμε την πιθανότητα με την οποία αυτή εμφανίζεται, έχουμε μία νέα συνάρτηση $f(x)$, που ονομάζεται **συνάρτηση πιθανότητας** ή **κατανομή πιθανότητας** με σύνολο ορισμού το πεδίο τιμών της X και πεδίο τιμών το σύνολο των πιθανοτήτων, με τις οποίες οι τιμές της X εμφανίζονται.
- Μία συνάρτηση $f(x)$ είναι συνάρτηση πιθανότητας ή κατανομή πιθανότητας της τυχαίας μεταβλητής X , όταν με την συνάρτηση αυτή f έχουμε την δυνατότητα ν' απεικονίσουμε τις τιμές της X στις πιθανότητες της.

$$f(x)=P(X=x)$$



Συνάρτηση Πιθανότητας Διακριτής Τ.Μ. (2)

- Η συνάρτηση πιθανότητας πρέπει να πληροί και τις συνθήκες:

$$1) \quad f(\mathbf{x}) \geq 0, \quad \forall \mathbf{x} \in \mathfrak{R} \quad 2) \quad \sum_{\mathbf{x}} f(\mathbf{x}) = 1$$



Συνάρτηση Κατανομής Διακριτής Τ.Μ. (1)

- Ονομάζουμε **συνάρτηση κατανομής** της διακριτής τυχαίας μεταβλητής X ή **αθροιστική συνάρτηση** και συμβολίζουμε με $F(x)$ τη συνάρτηση:

$$F(x_i) = P(-\infty < X \leq x_i) = \sum_{j=1}^i f(x_j)$$



Συνάρτηση Κατανομής Διακριτής Τ.Μ. (2)

Η συνάρτηση κατανομής της διακριτής τυχαίας μεταβλητής X

- είναι μη φθίνουσα
- είναι σταθερή σε κάθε διάστημα δύο διαδοχικών τιμών της X
- παίρνει την τιμή μηδέν, όταν $x < x_1$
- παίρνει την τιμή 1, όταν $x \geq x_k$
- είναι ασυνεχής από αριστερά σε κάθε x_i τιμή της X .



Μαθηματική Ελπίδα-Αναμενόμενη Τιμή Διακριτής Τ.Μ.

- Το άθροισμα των γινομένων των τιμών της τυχαίας μεταβλητής επί τις αντίστοιχες πιθανότητες τους ονομάζεται **μαθηματική ελπίδα** ή **μέση τιμή** ή **μέσος** ή **αναμενόμενη τιμή** της τυχαίας μεταβλητής X και συμβολίζεται με $E(X)$.

$$E(X) = \sum_{i=1}^v x_i \cdot p(x_i)$$



Ιδιότητες Μαθηματικής Ελπίδας

- Η μαθηματική ελπίδα σταθερού αριθμού ισούται με τον αριθμό αυτό: $E(a)=a$
- Η μαθηματική ελπίδα $E(X+Y)$ του αθροίσματος δύο μεταβλητών είναι ίση με το άθροισμα των μαθηματικών ελπίδων τους: $E(X+Y)=E(X)+E(Y)$
- Όταν δύο τυχαίες μεταβλητές είναι ανεξάρτητες τότε η μαθηματική ελπίδα του γινομένου τους είναι ίση με το γινόμενο των μαθηματικών τους ελπίδων:

$$E(X \cdot Y) = E(X) \cdot E(Y)$$



Διακύμανση Διακριτής Τ.Μ.

- **Διακύμανση** της τυχαίας μεταβλητής X , όταν αυτή υπάρχει, είναι η ποσότητα:

$$E[X - E(X)]^2 = \sum (x_i - E(X))^2 \cdot p(x_i)$$

και συμβολίζεται με $\text{Var}(X)$ ή $\sigma^2(X)$.

$$\text{Var}(X) = \sum (x_i - \mu)^2 \cdot p(x_i)$$

Η τετραγωνική ρίζα της διακύμανσης μιας τυχαίας μεταβλητής, ονομάζεται **τυπική απόκλιση** $\sigma(X)$ αυτής:

$$\sigma(X) = \sqrt{\text{Var}(X)}$$



Ιδιότητες της Διακύμανσης

$$\text{Var}(\alpha + \beta x) = \beta^2 \cdot \text{Var}(X)$$

Αν X και Y ανεξάρτητες τ.μ.

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$$

$$\text{Var}(\alpha X + \beta Y) = \alpha^2 \text{Var}(X) + \beta^2 \text{Var}(Y)$$



Συνδιακύμανση-Συμμεταβολή T.M.

- Συνδιακύμανση των τυχαίων μεταβλητών X και Y ονομάζεται η παράσταση:

$$E[(X - E(X)) \cdot (Y - E(Y))]$$

και συμβολίζεται $\text{cov}(X, Y)$

- Σύμφωνα με την σχέση αυτή η διακύμανση δύο μη ανεξάρτητων μεταβλητών γίνεται:

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{cov}(X, Y)$$

$$\text{Var}(Z) = \text{Var}(\alpha X + \beta Y) = \alpha^2 \text{Var}(X) + \beta^2 \text{Var}(Y) + 2\alpha\beta \text{cov}(X, Y)$$



Συνεχείς Τ.Μ.

- Η συνάρτηση κατανομής πιθανότητας τυχαίας συνεχούς κατανομής είναι:

$$F(X) = P(-\infty < X \leq x_1) = \int_{-\infty}^{x_1} f(x) dx$$

- Ιδιότητες:

$$1) f(x) \geq 0, \forall x \in \mathcal{R}$$

$$2) \int_{-\infty}^{+\infty} f(x) dx = 1$$

- Επίσης η πιθανότητα η X να έχει μια τιμή μεταξύ δύο πραγματικών αριθμών α και β είναι:

$$P(\alpha < x < \beta) = \int_{\alpha}^{\beta} f(x) dx$$

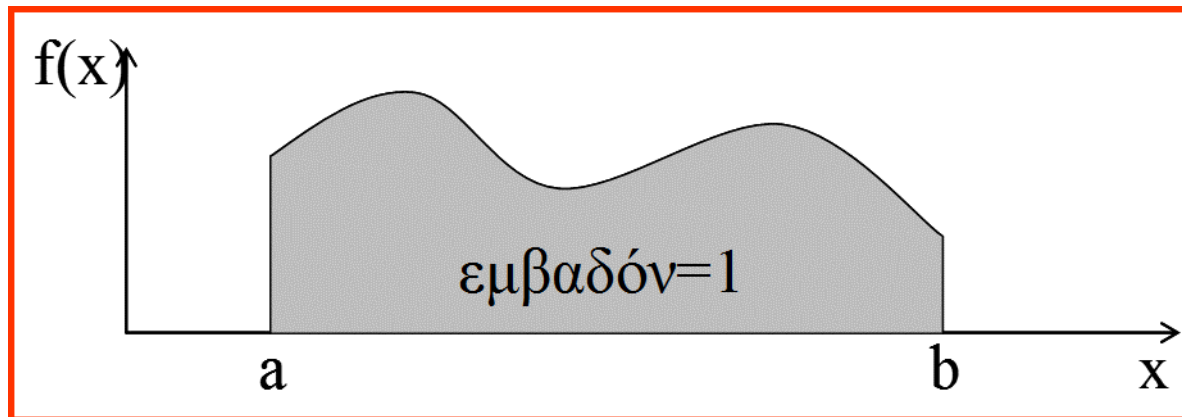
Συνάρτηση
Πυκνότητας
Πιθανότητας



Συνάρτηση Πυκνότητας Πιθανότητας

Μία συνάρτηση $f(x)$ καλείται **συνάρτηση πυκνότητας πιθανότητας (probability density function)** της τυχαίας μεταβλητής X , σε ένα διάστημα $a \leq x \leq b$, εάν ικανοποιεί τις ακόλουθες ιδιότητες:

- 1) $f(x) \geq 0$ για όλα τα x μεταξύ a και b , και



- 2) Το συνολικό εμβαδόν κάτω από την καμπύλη μεταξύ a και b είναι 1.0

Πηγή: Παπαδημητρίου (2001)



Μαθηματική Ελπίδα και Διακύμανση Συνεχών Τ.Μ.

$$E(X) = \int x \cdot f(x) dx$$

$$\text{Var}(X) = \int_x (x - E(x))^2 \cdot f(x) dx$$



Μονοδιάστατες Κατανομές Διακριτών Τ.Μ.



Πείραμα *Bernoulli* (1)

- Ας υποθέσουμε ότι πραγματοποιούμε ένα τυχαίο πείραμα το οποίο έχει δύο μόνον πιθανά (αμοιβαίως αποκλειόμενα) αποτελέσματα.
- Για παράδειγμα «ρίχνουμε» ένα νόμισμα και καταγράφουμε αν έχουμε κεφάλι ή γράμματα. Ή «ρίχνουμε» ένα ζάρι και καταγράφουμε εάν το αποτέλεσμα είναι «1» ή «όχι 1» (παρά το γεγονός ότι το ζάρι έχει 6 όψεις εμείς ενδιαφερόμαστε εάν το αποτέλεσμα ήταν «1» ή «όχι 1», κοινώς έχουμε δύο πιθανά αποτελέσματα).
- Για λόγους ευκολίας συνήθως αποκαλούμε (αυθαίρετα) το ένα εκ των δύο αποτελεσμάτων **Επιτυχία** (E) και το άλλο **Αποτυχία** (A).



Πείραμα *Bernoulli* (2)

- Η πιθανότητα το αποτέλεσμα του πειράματος να είναι E είναι p . Επομένως η πιθανότητα αποτυχίας θα είναι $q=1-p$.
- Ένα τυχαίο πείραμα με τα παραπάνω χαρακτηριστικά, αποκαλείται πείραμα *Bernoulli*, με πιθανότητα επιτυχίας p .
- Το παράδειγμα με την ρίψη του ζαριού, αποτελεί ένα πείραμα *Bernoulli*, με πιθανότητα επιτυχίας $1/6$.
- Η ρίψη ενός νομίσματος είναι επίσης ένα πείραμα *Bernoulli*, με πιθανότητα επιτυχίας $1/2$ (είτε ονομάσω επιτυχία το αποτέλεσμα να έρθει «γράμματα» ή «κεφαλή», αφού και τα δύο είναι *ισοπίθανα*).



Διωνυμική Κατανομή (1)

- Γενικά, όταν επαναλαμβάνουμε n φορές ένα τυχαίο πείραμα, που διενεργείται πάντοτε κάτω από τις **ίδιες ακριβώς συνθήκες** (δηλαδή οι πιθανότητες επιτυχίας p και αποτυχίας q είναι σταθερές και $p+q=1$), η πιθανότητα ένα ενδεχόμενο να εμφανισθεί k φορές στις n επαναλήψεις είναι ίση με:

$$P(X = k) = \binom{n}{k} p^k q^{n-k}$$



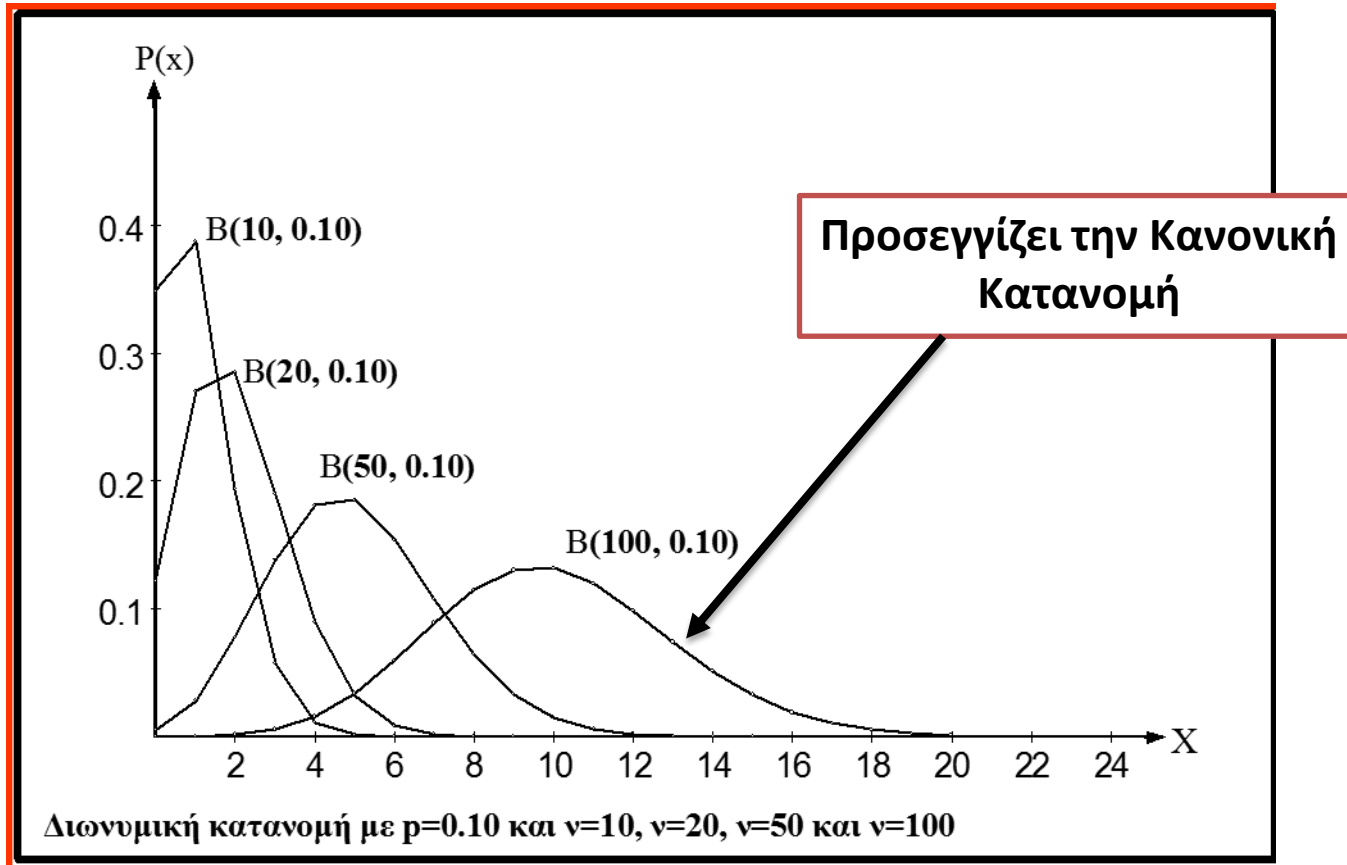
Διωνυμική Κατανομή (2)

- Η τ.μ. X που μετρά τον αριθμό επιτυχιών x στις n επαναλήψεις ενός πειράματος τύχης ακολουθεί τη **διωνυμική κατανομή** και συμβολίζεται με $B(n,p)$. Εξαρτάται από τον αριθμό n του δείγματος (των επαναλήψεων) και τη σταθερή πιθανότητα επιτυχίας p και γι' αυτό και στο συμβολισμό της εμφανίζονται αυτές οι παράμετροι $B(n,p)$. Η $P(x)$ συμβολίζεται με $b(x,n,p)$.

$$P(x) = b(x,n,p)$$



Διωνυμική Κατανομή (3)



Πηγή: Παπαδημητρίου (2001)



Διωνυμική Κατανομή (4)

- Χρήσιμη Ιδιότητα:

Εστω οι δύο ανεξάρτητες τυχαίες μεταβλητές X_1 και X_2 που ακολουθούν τις διωνυμικές κατανομές $B(v_1, p)$ και $B(v_2, p)$ αντίστοιχα με κοινή πιθανότητα επιτυχίας p τότε η τυχαία μεταβλητή $X = X_1 + X_2$ θα ακολουθεί τη διωνυμική κατανομή $B(v_1 + v_2, p)$



Μέση Τιμή και Διακύμανση T.M. που ακολουθεί τη Διωνυμική Κατανομή

$$E(X) = \text{Var}(x) = \lambda$$

Η κατανομή Poisson είναι πολύ χρήσιμη για τη μελέτη και διερεύνηση φαινομένων που είναι ανεξάρτητα του χρόνου και που ο αριθμός των επιτυχιών σε μεγάλο αριθμό επαναλήψεων είναι μικρός.



Παράδειγμα (1)

- Έστω ότι ρίχνουμε ένα ζάρι 5 φορές και ενδιαφερόμαστε αν το αποτέλεσμα κάθε ρίψης ήταν «1» ή «όχι 1».
 1. Ποια η πιθανότητα να μην έρθει ούτε μια φορά στις 5 προσπάθειες το «1»?
 2. Ποια η πιθανότητα να έρθει ακριβώς τρεις φορές στις 5 προσπάθειες το «1»?
 3. Ποια η πιθανότητα να έρθει τουλάχιστο δύο φορές στις 5 προσπάθειες το «1»?



Παράδειγμα (2)

- Ας θεωρήσουμε ως επιτυχία (E) το αποτέλεσμα της ρίψης να είναι το “1” και ως αποτυχία (A) το αποτέλεσμα να είναι «όχι 1» (κοινώς το «όχι 1» σημαίνει ότι το αποτέλεσμα μπορεί να είναι 2,3,4,5 ή 6).
- Η πιθανότητα να έρθει «1» όταν ρίχνουμε ένα ζάρι είναι $1/6$. Άρα η πιθανότητα επιτυχίας είναι $p=1/6$ και επομένως η πιθανότητα αποτυχίας θα είναι $q=1-p=5/6$.
- Αν με X συμβολίσουμε το συνολικό αριθμό επιτυχιών στις 5 επαναλήψεις του πειράματος, τότε $X \sim B(5, 1/6)$. Έχουμε λοιπόν:



Παράδειγμα (3)

$$1. \quad P(X=0) = \frac{5!}{0!(5-0)!} \left(\frac{1}{6}\right)^0 \left(\frac{5}{6}\right)^{5-0} \cong 0,4019$$

$$2. \quad P(X=3) = \frac{5!}{3!(5-3)!} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^{5-3} = \frac{5 \times 4 \times 3!}{3!2!} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^2 = 10 \times \frac{5^2}{6^5} \cong 0,03215$$

$$3. \quad P(X \geq 2) = 1 - P(X < 1) = 1 - [P(X=0) + P(X=1)] \quad (1)$$

όμως η πιθανότητα $P(X=0)$ υπολογίστηκε στο πρώτο ερώτημα ενώ για την $P(X=1)$ έχουμε:

$$P(X=1) = \frac{5!}{1!(5-1)!} \left(\frac{1}{6}\right)^1 \left(\frac{5}{6}\right)^{5-1} = 5 \times \frac{5^4}{6^5} \cong 0,4019$$

άρα η σχέση (1) γίνεται:

$$P(X \geq 2) \cong 1 - [0,4019 + 0,4019] = 0,1962$$



Κατανομή Poisson (1)

- Στις περιπτώσεις που η πιθανότητα επιτυχίας p είναι πολύ μικρή ($p < 0,05$) και ο αριθμός των δειγμάτων (ή των επαναλήψεων του πειράματος) μεγάλος ($n > 20$) ώστε ο αριθμός $\lambda = np$ να είναι μικρότερος του 10 η διωνυμική κατανομή προσεγγίζει σε μια άλλη θεωρητική κατανομή, την **Κατανομή Poisson**:

$$P(x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}$$

$$e = 2,71828$$



Κατανομή Poisson (2)

- Η κατανομή του **Poisson** έχει εφαρμογή στις περιπτώσεις που το ενδεχόμενο επιτυχίας μπορεί να θεωρηθεί **σπάνιο** και το μέγεθος των **επαναλήψεων** του πειράματος (ή το μέγεθος του δείγματος) είναι **αρκετά μεγάλο**.
 - Η κατανομή των βλαβών μιας μηχανής την ώρα.
 - Η κατανομή των ελαττωματικών αντικειμένων που παράγει μια μηχανή, σε κάθε ομοειδή και ισοπληθή ομάδα αυτών.
 - Η κατανομή των τυπογραφικών λαθών ανά σελίδα.
 - Η κατανομή των πελατών που εξυπηρετούνται από μια θυρίδα εκδόσεως εισιτηρίων.
 - Η κατανομή των αυτοκινήτων που περνούν ανά λεπτό από τα διόδια.
 - Η κατανομή των θανάτων που οφείλονται σε μία αρρώστια, ανά ημέρα σε μια πόλη.



Κατανομή Poisson (3)

- Χρήσιμη Ιδιότητα:

Έστω οι δύο ανεξάρτητες τυχαίες μεταβλητές X_1 και X_2 που ακολουθούν τις κατανομές του Poisson $P(\lambda_1)$ και $P(\lambda_2)$ αντίστοιχα. Η τυχαία μεταβλητή $X=X_1+X_2$ θα ακολουθεί την κατανομή του Poisson $P(\lambda_1+\lambda_2)$.



Μέση Τιμή και Διακύμανση T.M. που ακολουθεί την Κατανομή Poisson

$$E(X) = \text{Var}(x) = \lambda$$

Η κατανομή Poisson είναι πολύ χρήσιμη για τη μελέτη και διερεύνηση φαινομένων που είναι ανεξάρτητα του χρόνου και που ο αριθμός των επιτυχιών σε μεγάλο αριθμό επαναλήψεων είναι μικρός.



Κατανομή Poisson (4)

- i. Σε κάθε διάστημα χρόνου ή χώρου ένα ενδεχόμενο μπορεί να συμβεί ή να μη συμβεί. Η εμφάνιση του ενδεχομένου καλείται συνήθως επιτυχία ενώ η μη εμφάνιση του αποτυχία.
- ii. Οι τυχαίες εμφανίσεις ενδεχομένων είναι ανεξάρτητες δηλαδή η εμφάνιση ενός ενδεχομένου σε ένα διάστημα χρόνου ή χώρου δεν επηρεάζει την πιθανότητα εμφάνισης του στο επόμενο διάστημα χρόνου ή χώρου.
- iii. Η πιθανότητα εμφάνισης (ή μη εμφάνισης) ενός ενδεχομένου σε ένα διάστημα χρόνου ή χώρου παραμένει σταθερή για όλη τη διάρκεια του φαινομένου.



Παράδειγμα

- Μια υπάλληλος η οποία εισάγει δεδομένα στον Η/Υ κάνει κατά μέσο όρο τρία λάθη ανά σελίδα. Να υπολογιστεί η πιθανότητα σε τυχαία επιλεγμένη σελίδα να βρεθούν δύο λάθη.
- Εστω X ο αριθμός των λαθών ανά σελίδα. Σύμφωνα με τα δεδομένα $X \sim P(\lambda=3)$.
- Ζητάμε την πιθανότητα $P(X=2)$

$$P(X = 2) = \frac{e^{-3} 3^2}{2!} = \frac{\frac{1}{e^3} * 3^2}{1 * 2} = \frac{(0,050) * 9}{2} = 0,225$$



Πολυωνυμική Κατανομή

- Αποτελεί **Γενίκευση** της Διωνυμικής:

$$P(y_1, y_2, \dots, y_k) = \frac{n!}{y_1! y_2! \dots y_k!} p_1^{y_1} p_2^{y_2} \dots p_k^{y_k}$$

$$\sum y_i = n \quad \sum p_i = 1$$

- Εκφράζει την πιθανότητα στις n επαναλήψεις του πειράματος να έχουμε y_i εκβάσεις από το κάθε είδος k δυνατών εκβάσεων.



Στατιστικοί Πίνακες 1



Πίνακας 3
Κατανομή Poisson
(συνέχεια)

λ										
x	7.10	7.20	7.30	7.40	7.50	7.60	7.70	7.80	7.90	8.00
15	.0037	.0041	.0046	.0051	.0057	.0062	.0069	.0075	.0083	.0090
16	.0016	.0019	.0021	.0024	.0026	.0030	.0033	.0037	.0041	.0045
17	.0007	.0008	.0009	.0010	.0012	.0013	.0015	.0017	.0019	.0021
18	.0003	.0003	.0004	.0004	.0005	.0006	.0006	.0007	.0008	.0009
19	.0001	.0001	.0001	.0002	.0002	.0002	.0003	.0003	.0003	.0004
20			.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0002
21									.0001	.0001
λ										
x	8.10	8.20	8.30	8.40	8.50	8.60	8.70	8.80	8.90	9.00
0	.0003	.0003	.0002	.0002	.0002	.0002	.0002	.0002	.0001	.0001
1	.0025	.0023	.0021	.0019	.0017	.0016	.0014	.0013	.0012	.0011
2	.0100	.0092	.0086	.0079	.0074	.0068	.0063	.0058	.0054	.0050
3	.0269	.0252	.0237	.0222	.0208	.0195	.0183	.0171	.0160	.0150
4	.0544	.0517	.0491	.0466	.0443	.0420	.0398	.0377	.0357	.0337
5	.0882	.0849	.0816	.0784	.0752	.0722	.0692	.0663	.0635	.0607
6	.1191	.1160	.1128	.1097	.1066	.1034	.1003	.0972	.0941	.0911
7	.1378	.1358	.1338	.1317	.1294	.1271	.1247	.1222	.1197	.1171
8	.1395	.1392	.1388	.1382	.1375	.1366	.1356	.1344	.1332	.1318
9	.1256	.1269	.1280	.1290	.1299	.1306	.1311	.1315	.1317	.1318
10	.1017	.1040	.1063	.1084	.1104	.1123	.1140	.1157	.1172	.1186
11	.0749	.0776	.0802	.0828	.0853	.0878	.0902	.0925	.0948	.0970
12	.0505	.0530	.0555	.0579	.0604	.0629	.0654	.0679	.0703	.0728
13	.0315	.0334	.0354	.0374	.0395	.0416	.0438	.0459	.0481	.0504
14	.0182	.0196	.0210	.0225	.0240	.0256	.0272	.0289	.0306	.0324
15	.0098	.0107	.0116	.0126	.0136	.0147	.0158	.0169	.0182	.0194
16	.0050	.0055	.0060	.0066	.0072	.0079	.0086	.0093	.0101	.0109
17	.0024	.0026	.0029	.0033	.0036	.0040	.0044	.0048	.0053	.0058
18	.0011	.0012	.0014	.0015	.0017	.0019	.0021	.0024	.0026	.0029
19	.0005	.0005	.0006	.0007	.0008	.0009	.0010	.0011	.0012	.0014
20	.0002	.0002	.0002	.0003	.0003	.0004	.0004	.0005	.0005	.0006
21	.0001	.0001	.0001	.0001	.0001	.0002	.0002	.0002	.0002	.0003
22					.0001	.0001	.0001	.0001	.0001	.0001
λ										
x	9.10	9.20	9.30	9.40	9.50	9.60	9.70	9.80	9.90	10.00
0	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0000



Μέση Τιμή και Διακύμανση

$$E(X) = v \cdot P$$

$$Var(X) = \sigma^2 = v \cdot P \cdot Q \cdot \frac{N - v}{N - 1}$$

$$P = \frac{N_1}{N}$$

$$Q = 1 - P$$



Αρνητική Διωνυμική Κατανομή-Κατανομή *Pascal*

- Ας υποθέσουμε ότι οι επιλογές, από έναν πληθυσμό όπου ισχύουν οι συνθήκες *Bernoulli*, συνεχίζονται μέχρι να έχουμε v επιτυχίες. Έστω ότι ο αριθμός των επιλογών, που χρειάζονται για να έχουμε ακριβώς v επιτυχίες, εκφράζεται από την τυχαία μεταβλητή X .
- Μια τυχαία μεταβλητή X με τιμές της, $v, v+1, v+2, \dots$ ($v > 1$), θα λέμε ότι ακολουθεί την Αρνητική Διωνυμική Κατανομή ή την κατανομή *Pascal*, όταν η πιθανότητα να λάβει η X την τιμή $x \geq v$ δίνεται από τον τύπο:

$$P(X = x) = \binom{x-1}{v-1} \cdot p^v \cdot q^{x-v}$$



Μέση Τιμή και Διακύμανση

$$E(X) = \frac{v}{p}$$

$$\text{Var}(x) = \frac{v \cdot q}{p^2}$$



Παρατήρηση

- Όπως αναφέραμε στη διωνυμική κατανομή, αυτή εφαρμόζεται όταν σ' έναν ορισμένο αριθμό επαναλήψεων ενός πειράματος (ή σ' έναν αριθμό επιλογών), εξετάζουμε τον αριθμό των επιτυχιών που εμφανίζονται σ' αυτές.
- Η διαφορά που υπάρχει, στην τυχαία μεταβλητή που ακολουθεί την κατανομή του Pascal, είναι ότι, σ' αυτήν εξετάζουμε τον αριθμό των επαναλήψεων που χρειάζονται για να εμφανιστεί σ' αυτές ένας ορισμένος αριθμός επιτυχιών.
- Ενώ, δηλαδή, σε μία διωνυμική κατανομή $B(n,p)$ εκτός από την πιθανότητα επιτυχίας p ορίζεται και ο αριθμός n των επαναλήψεων, στην κατανομή Pascal εκτός από την πιθανότητα επιτυχίας p ορίζεται ο αριθμός των επιτυχιών n που επιθυμούμε.



Γεωμετρική Κατανομή (1)

- Υποθέτουμε ότι σε μία διωνυμική κατανομή η πιθανότητα επιτυχίας είναι p και η πιθανότητα αποτυχίας q ($p+q=1$). Η πιθανότητα να έχουμε στην πρώτη επιλογή ενός ατόμου από τον πληθυσμό αυτό αποτυχία είναι q , η πιθανότητα να έχουμε και στη δεύτερη επιλογή αποτυχία είναι q^2 (υποθέτουμε ότι η δεύτερη επιλογή είναι ανεξάρτητη από την πρώτη). Η πιθανότητα αποτυχίας και στην τρίτη επιλογή είναι q^3 , γενικά δε η πιθανότητα να έχουμε αποτυχία και στην K τάξης επιλογή είναι:

$$q \cdot q \cdot \dots \cdot q = q^K$$

Η παραπάνω ιδιότητα ονομάζεται “Έλλειψη Μνήμης”



Γεωμετρική Κατανομή (2)

- Όταν η συνάρτηση πιθανότητας μιας τυχαίας μεταβλητής X δίνεται από τον τύπο:

$$P(X = x) = q^{x-1} \cdot p$$

με $p + q = 1$ και $x = 1, 2, 3, \dots, \nu, \dots$

λέμε ότι αυτή η τυχαία μεταβλητή ακολουθεί τη Γεωμετρική Κατανομή με παράμετρο p .



Μέση Τιμή και Διακύμανση

$$E(X) = \frac{1}{p}$$

$$Var(X) = \frac{q}{p^2}$$



Μονοδιάστατες Κατανομές Συνεχών Τ.Μ.

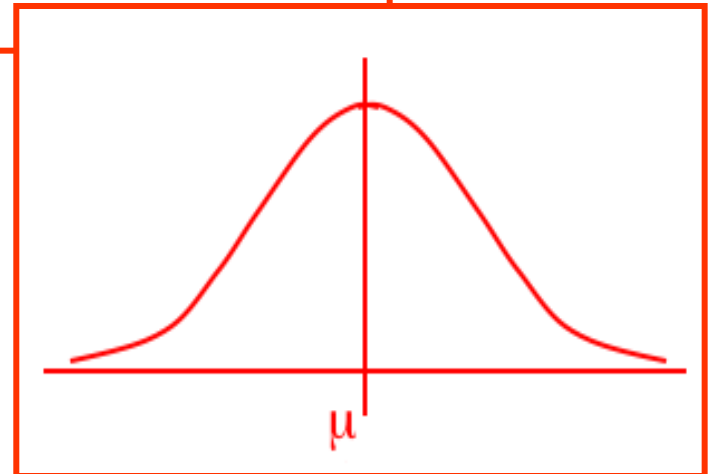


Η Κανονική Κατανομή (1)

Η **κανονική κατανομή** (*normal distribution*) είναι η πιο σημαντική από όλες τις κατανομές πιθανοτήτων. Η συνάρτηση κατανομής πιθανότητας είναι η **κανονική τυχαία μεταβλητή** και δίνεται από:

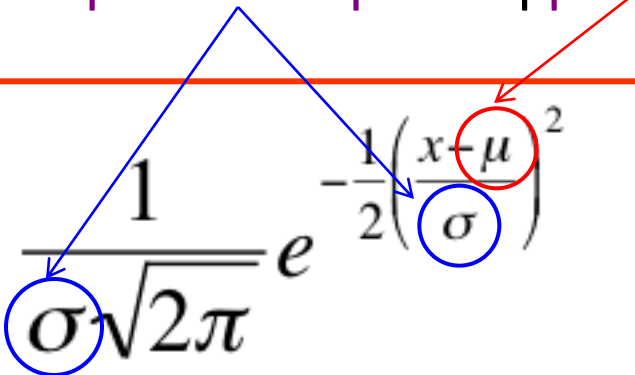
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad -\infty < x < \infty$$

Έχει το σχήμα Καμπάνας,
Είναι συμμετρική ως προς
τη μέση τιμή μ



Η Κανονική Κατανομή (2)

Η κανονική κατανομή ορίζεται πλήρως με **δύο παραμέτρους**:
την **τυπική απόκλιση** και τη **μέση τιμή** της

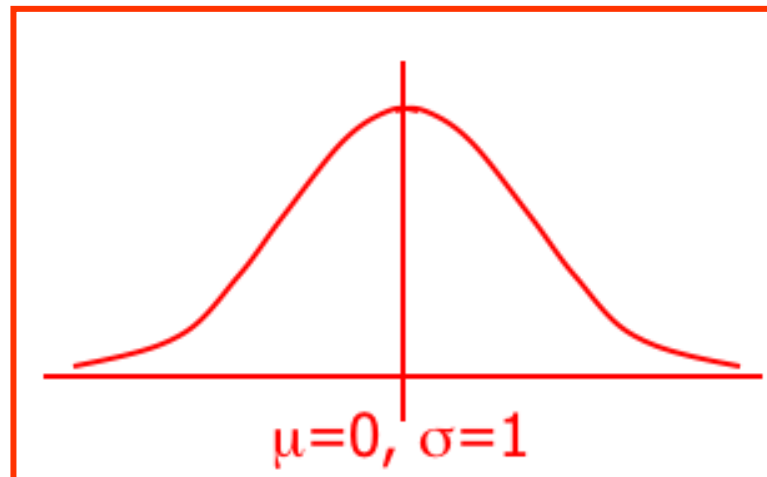
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad -\infty < x < \infty$$




Η Τυπική Κανονική Κατανομή

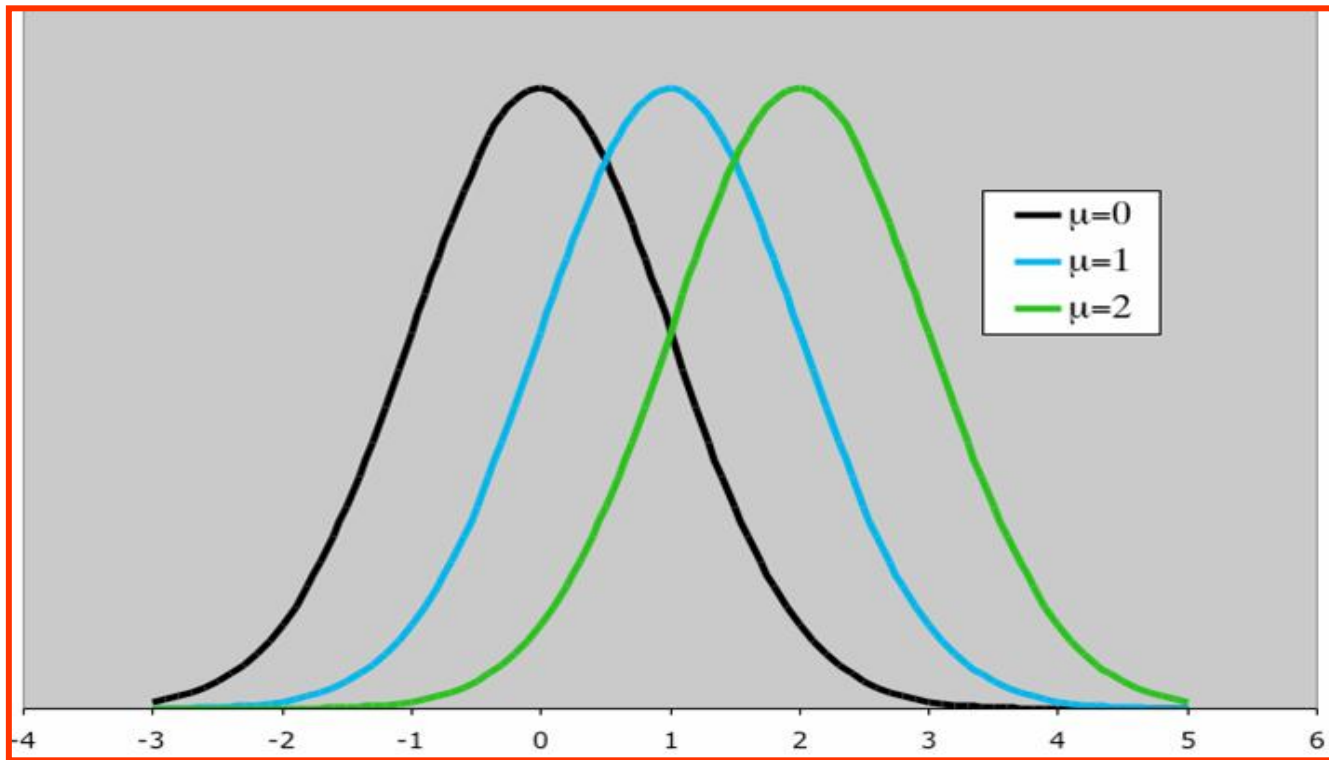
Η κανονική κατανομή στην οποία η μέση τιμή είναι μηδέν και η τυπική απόκλιση είναι ίση με τη μονάδα καλείται τυπική ή τυποποιημένη κανονική κατανομή.

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-0}{1}\right)^2} \quad -\infty < x < \infty$$



Η Κανονική Κατανομή (4)

Αυξάνοντας τη μέση τιμή μετακινείται η καμπύλη προς τα δεξιά...
Ίδια διακύμανση αλλά διαφορετικές μέσες τιμές.



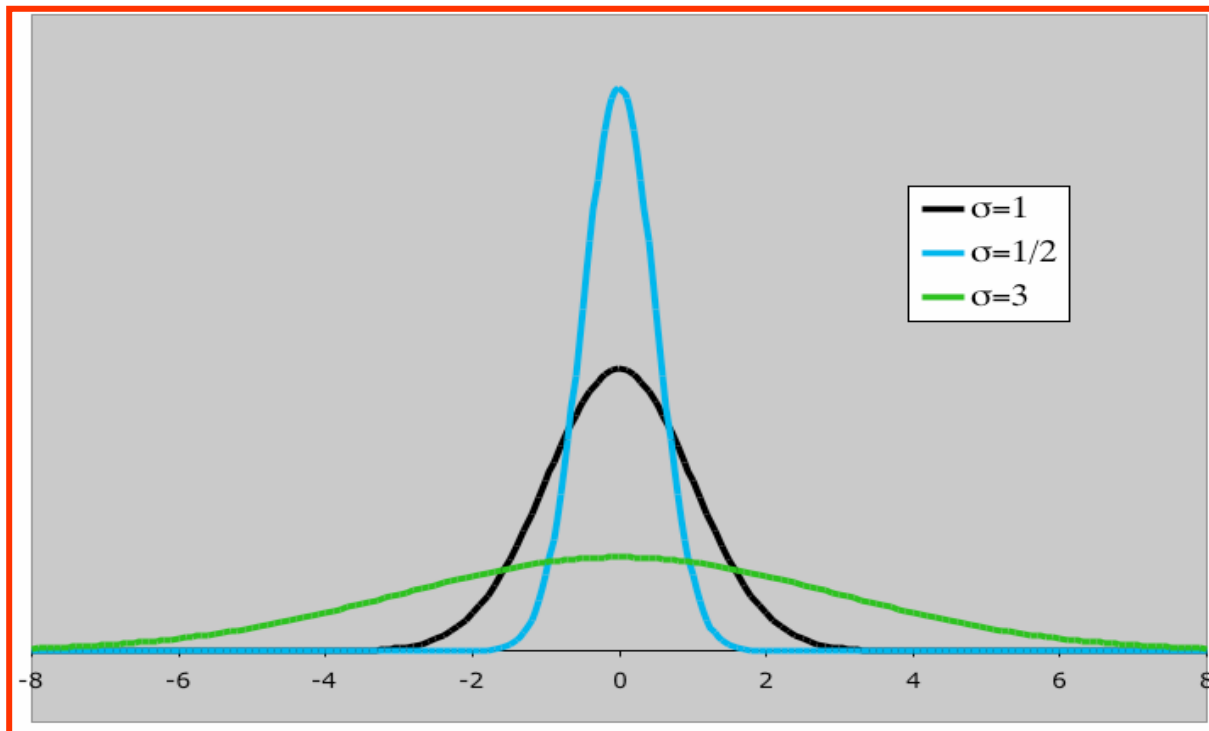
Πηγή: Brooks/Cole (2005)



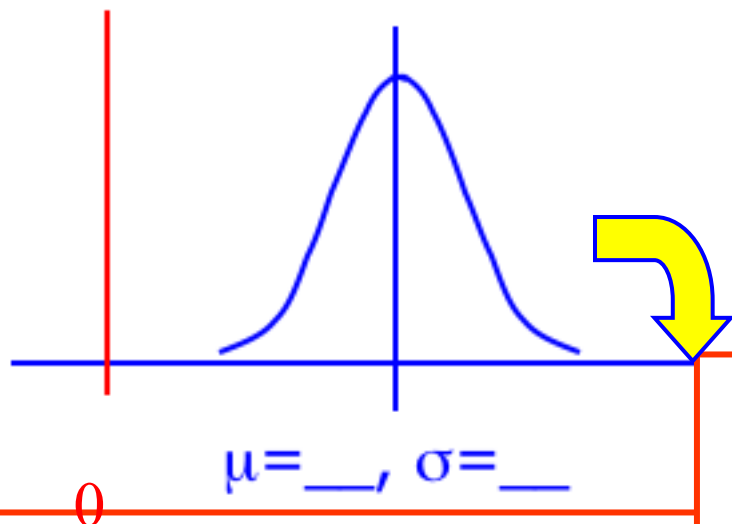
Η Κανονική Κατανομή (5)

Αυξάνοντας την τυπική απόκλιση, η καμπύλη «απλώνεται» ...

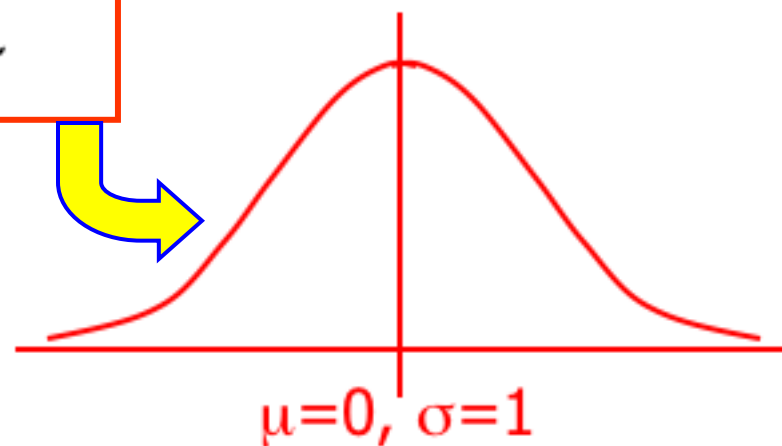
Ίδιες Μέσες Τιμές, Διαφορετικές Τυπικές Αποκλίσεις



Υπολογίζοντας Πιθανότητες (1)



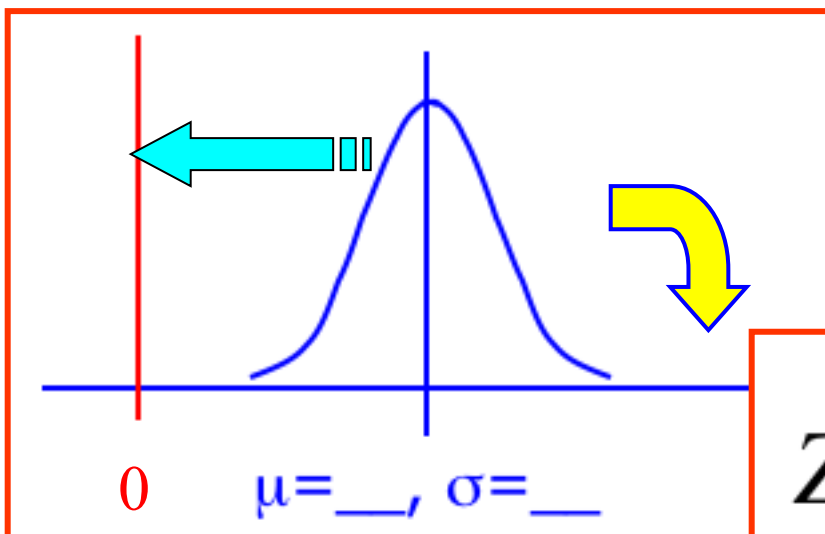
$$Z = \frac{X - \mu}{\sigma}$$



**Συμβουλή: πάντα να
σχεδιάζετε διάγραμμα!**

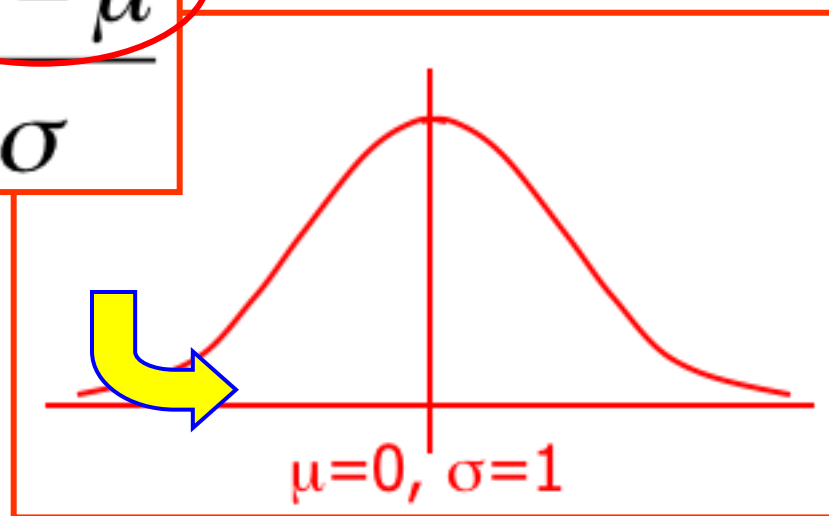


Υπολογίζοντας Πιθανότητες (2)

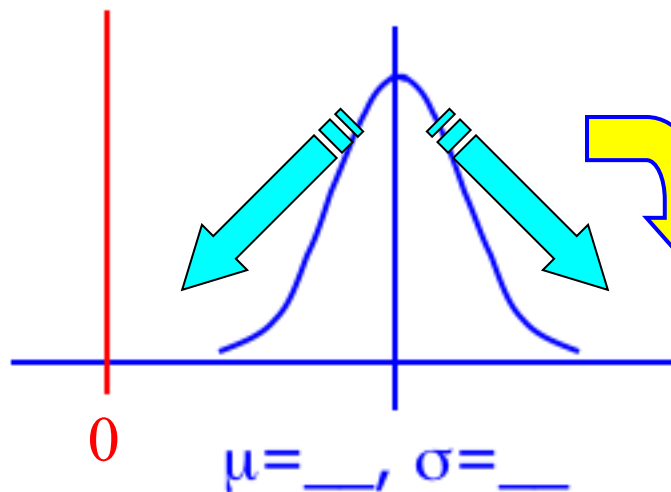


Αυτό μετακινεί την μέση τιμή της X στο 0

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

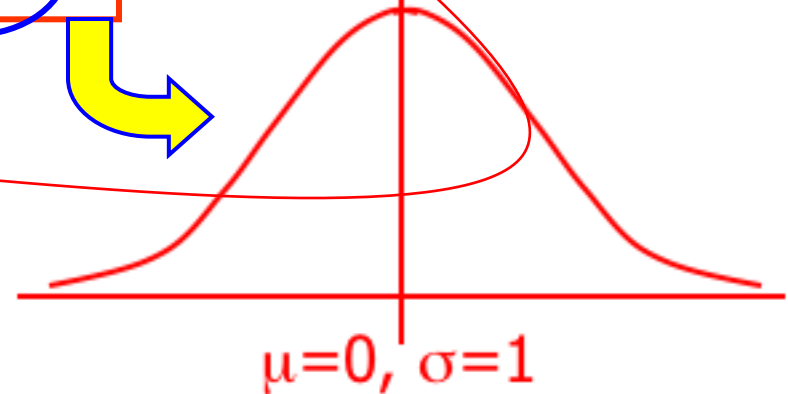


Υπολογίζοντας Πιθανότητες (3)



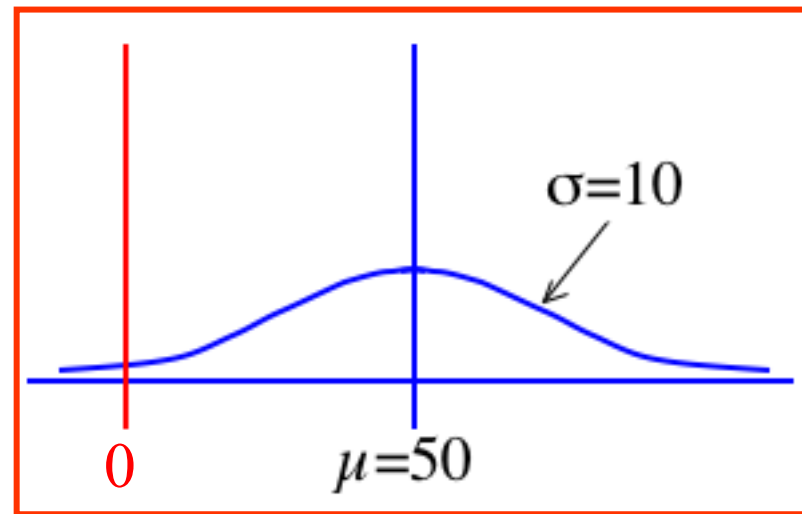
$$Z = \frac{X - \mu}{\sigma}$$

Αυτό αλλάζει το σχήμα
της καμπύλης ...



Παράδειγμα

Ο χρόνος που απαιτείται για την συναρμολόγηση ενός υπολογιστή είναι κανονικά κατανομημένος με μέση τιμή 50 λεπτά και τυπική απόκλιση 10 λεπτά:

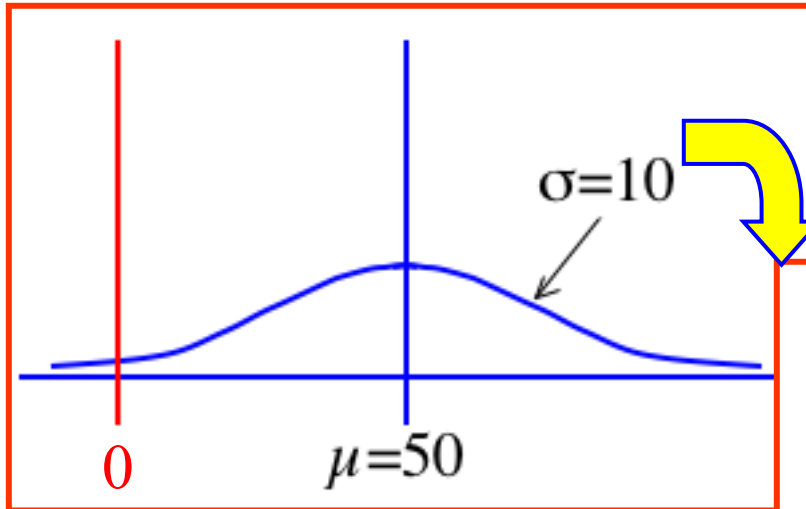


Ποια είναι η πιθανότητα ένας υπολογιστής να συναρμολογηθεί σε χρόνο μεταξύ 45 και 60 λεπτών;

Αλγεβρικά: Ποια είναι η $P(45 < X < 60)$;

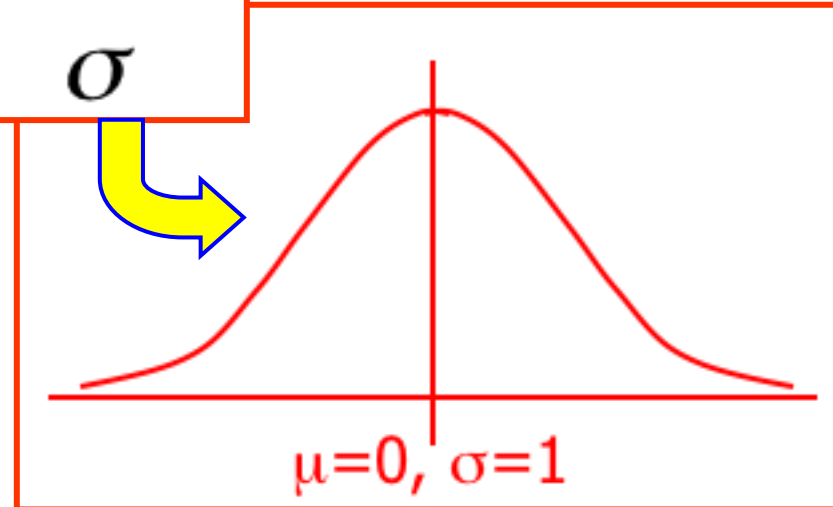


Παράδειγμα (1)



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

$$P(45 < X < 60) =$$
$$P\left(\frac{45 - 50}{10} < \frac{X - \mu}{\sigma} < \frac{60 - 50}{10}\right) =$$
$$P(-.5 < Z < 1)$$



Παράδειγμα (2)

Έχουμε μετατρέψει την $P(45 < X < 60)$ για μία κανονική κατανομή με μέση τιμή = 50 και τυπική απόκλιση = 10

σε

$P(-0,5 < Z < 1)$ [τυπική κανονική κατανομή με μέση τιμή = 0 και τυπική απόκλιση = 1]



Παράδειγμα (3)

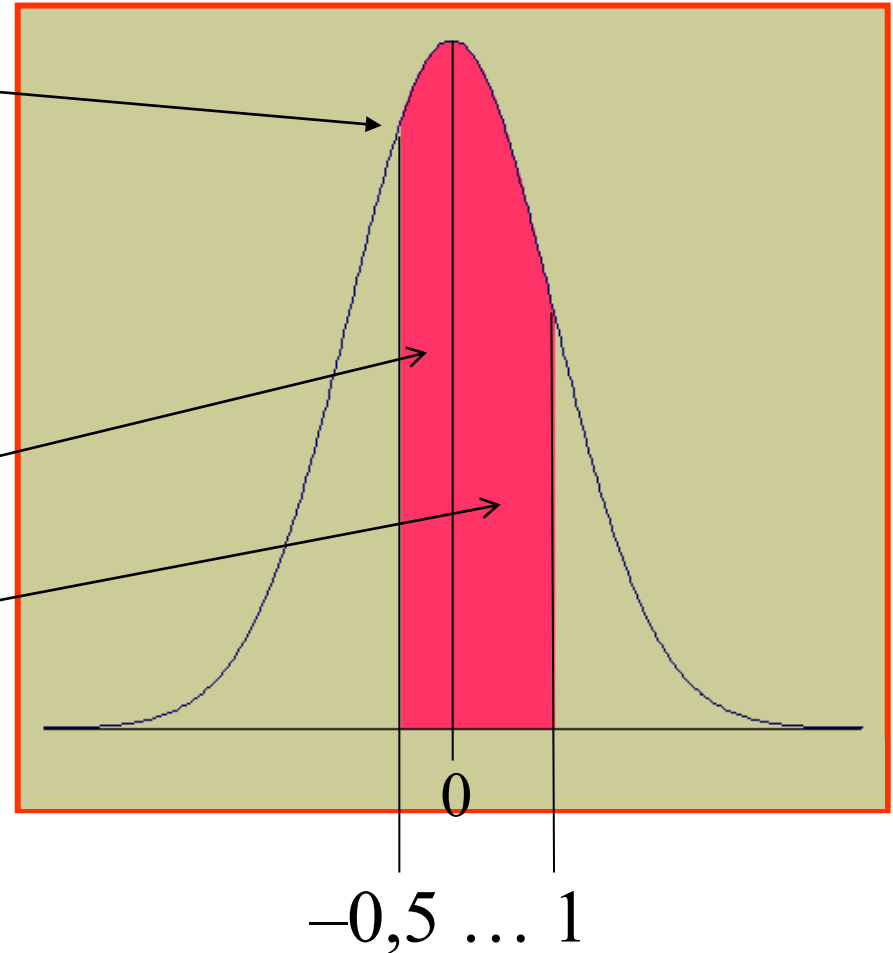
$$P(-0,5 < Z < 1)$$

Η πιθανότητα είναι το εμβαδόν
κάτω από την καμπύλη ...

Προσθέτουμε τα δύο μέρη:

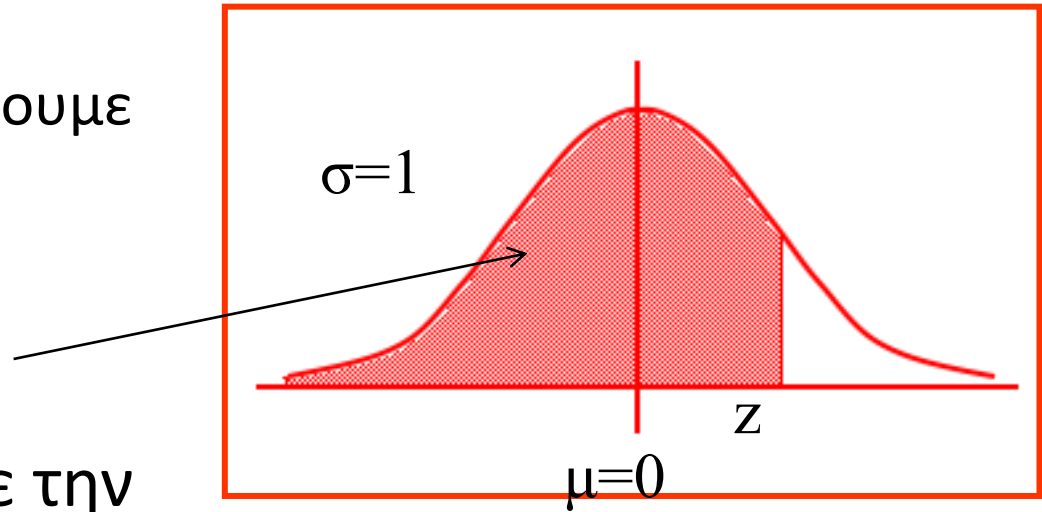
$$P(-0,5 < Z < 0) \text{ και}$$

$$P(0 < Z < 1)$$



Παράδειγμα (4)

Μπορούμε να χρησιμοποιήσουμε
Πίνακες για να
βρούμε πιθανότητες $P(Z < z)$



Μπορούμε να γράψουμε την

$$P(-0,5 < Z < 1) \text{ σε: } P(Z < 1) - P(Z < -0,5)$$

και από τον Πίνακα βρίσκουμε $P(Z < 1) = 0,8413$ και

$P(Z < -0,5) = 0,3085$. Επομένως:

$$P(-0,5 < Z < 1) = 0,8413 - 0,3085 = 0,5328$$



Χρήση Πίνακα (1)

Αυτός ο Πίνακας δίνει τις πιθανότητες
 $P(Z < z)$

Πρώτη στήλη = ακέραιος + πρώτο
 δεκαδικό ψηφίο

Πρώτη γραμμή = δεύτερο δεκαδικό
 ψηφίο

$$P(Z < 1) = 0,8413$$

$$P(Z < 0,5) = 0,3085$$

$$P(-0,5 < Z < 1) = 0,8413 + 0,3085 = 0,5328$$

Appendix C

Statistical Tables

TABLE C.1 (concluded)

z	0	1	2	3	4	5	6	7	8	9
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

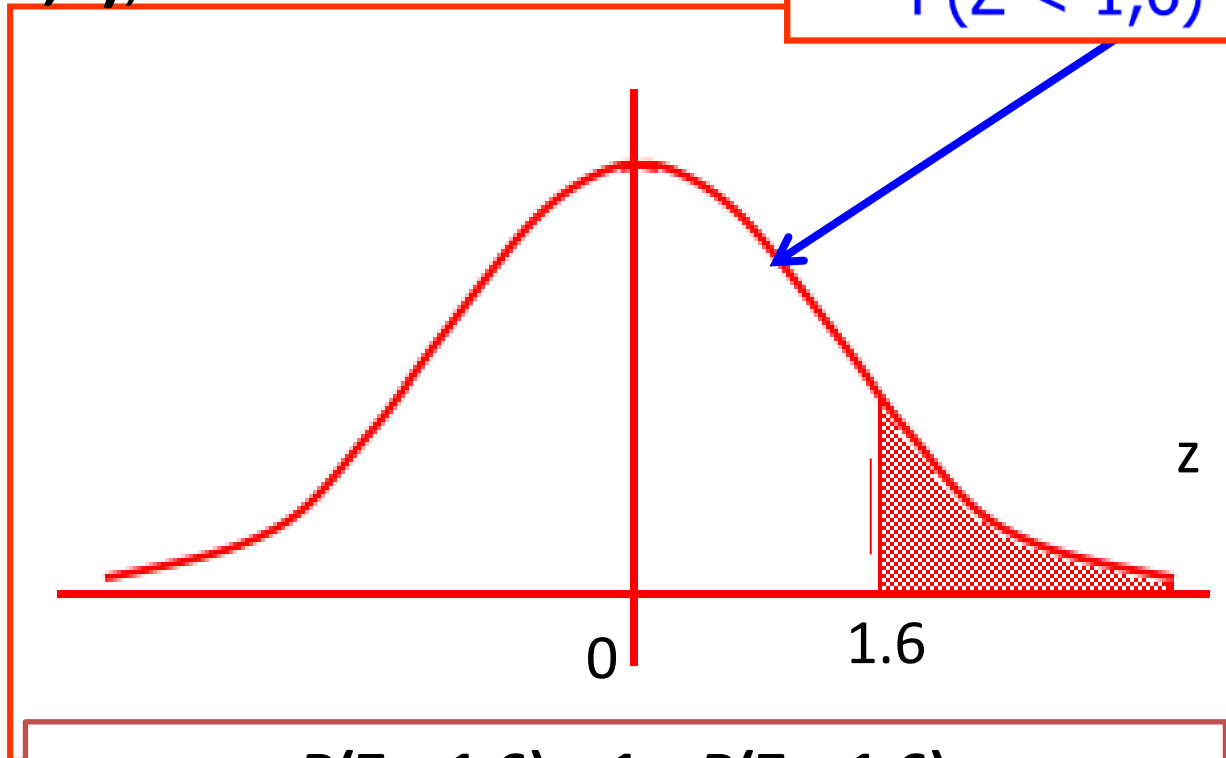
Examples: If $Z \sim \text{Normal}(0,1)$ then $P(Z \leq -1.32) = .0934$ and $P(Z \leq 1.84) = .9671$.
 Source: This table was generated using the Stata® function normd.



Χρήση Πίνακα (2)

$P(Z > 1,6)$;

$$P(Z < 1,6) = 0,9452$$



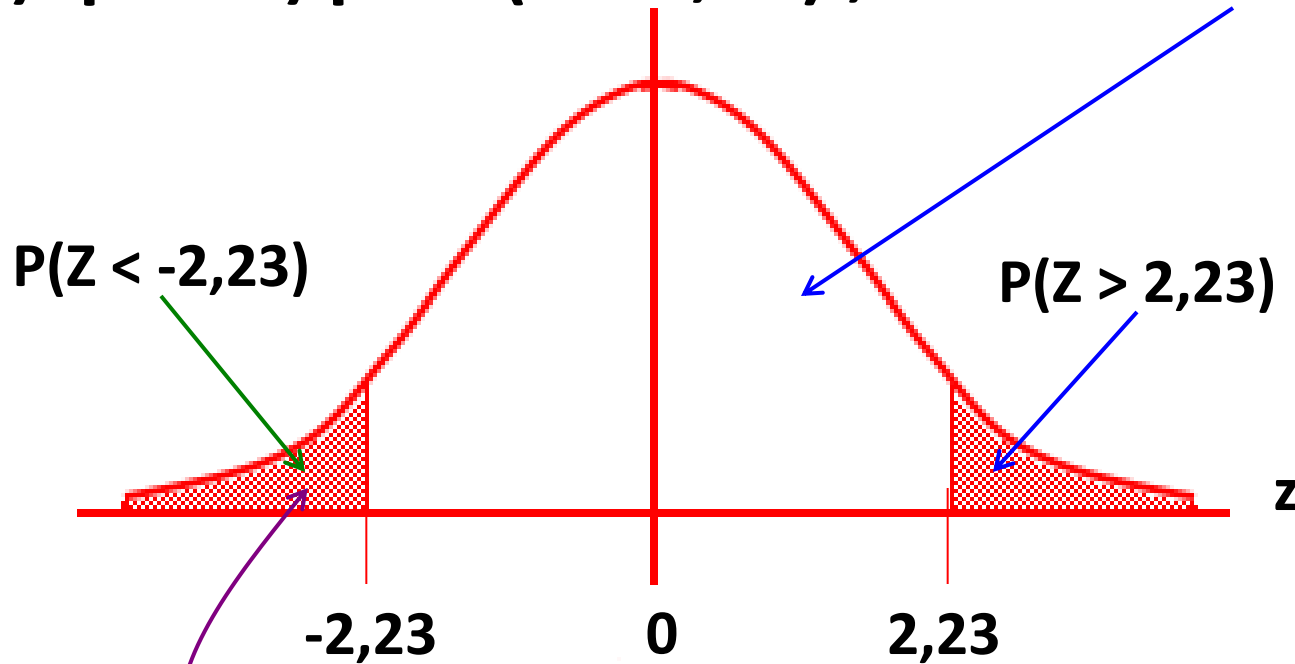
$$\begin{aligned} P(Z > 1,6) &= 1 - P(Z < 1,6) \\ &= 1 - 0,9452 \\ &= 0,0548 \end{aligned}$$



Χρήση Πίνακα (3)

Άλλος τρόπος για $P(Z > 2,23)$;

$P(0 < Z < 2,23)$

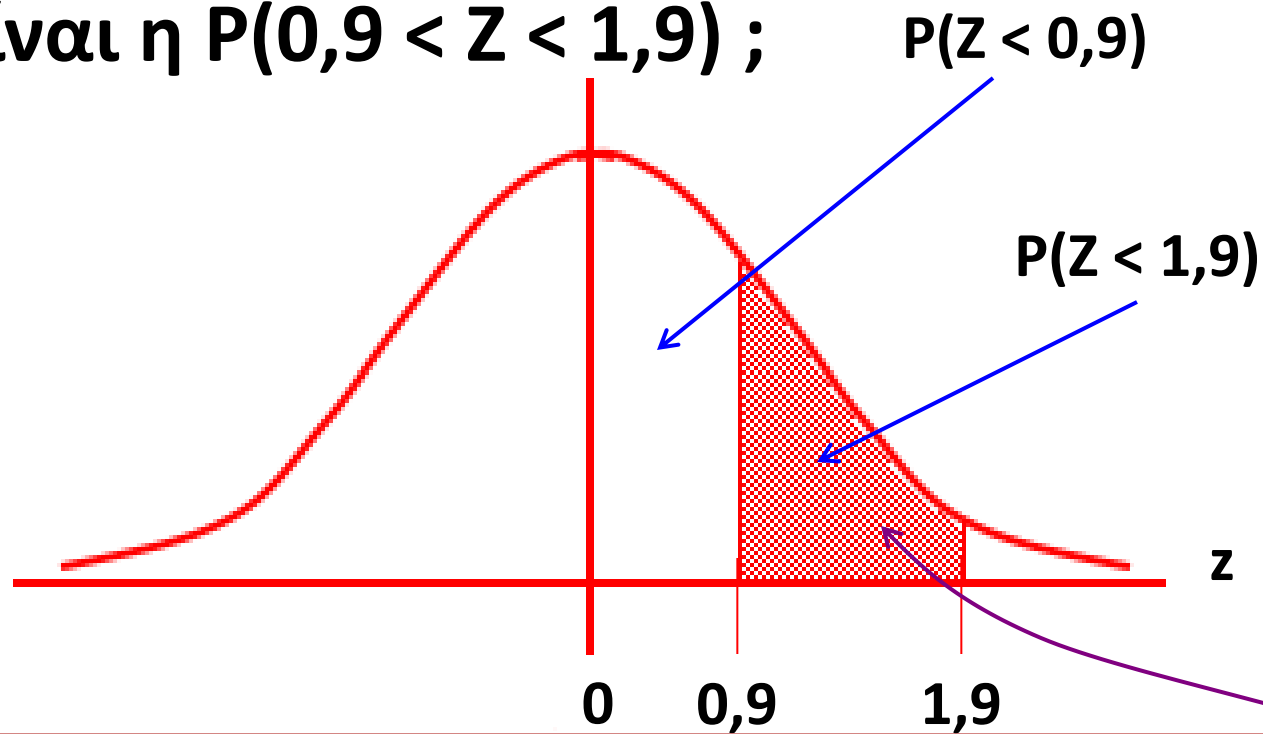


$$P(Z > 2,23) = P(Z < -2,23) \\ = 0,0129$$



Χρήση Πίνακα (4)

Ποιά είναι η $P(0,9 < Z < 1,9)$;



$$\begin{aligned} P(0,9 < Z < 1,9) &= P(Z < 1,9) - P(Z < 0,9) \\ &= 0,9713 - 0,8159 \\ &= 0,1554 \end{aligned}$$



Παράδειγμα (1)

Η απόδοση μιας επένδυσης ακολουθεί κανονική κατανομή με μέση τιμή 10% και τυπική απόκλιση 5%. Ποια είναι η πιθανότητα να χάσουμε χρήματα; Θέλουμε να υπολογίσουμε $P(X < 0)$. Έτσι:

$$\begin{aligned} P(X < 0) &= P\left(\frac{X - \mu}{\sigma} < \frac{0 - 10}{5}\right) \\ &= P(Z < -2) \\ &= 0,0228 \end{aligned}$$



Παράδειγμα (2)

Η απόδοση μιας επένδυσης ακολουθεί κανονική κατανομή με μέση τιμή 10% και τυπική απόκλιση 5%. Ποια είναι η πιθανότητα να χάσουμε χρήματα; Θέλουμε να υπολογίσουμε $P(X < 0)$. Έτσι:

$$\begin{aligned} P(X < 0) &= P\left(\frac{X - \mu}{\sigma} < \frac{0 - 10}{5}\right) \\ &= P(Z < -2) \\ &= 0,0228 \end{aligned}$$



Παράδειγμα 2 (συνέχεια)

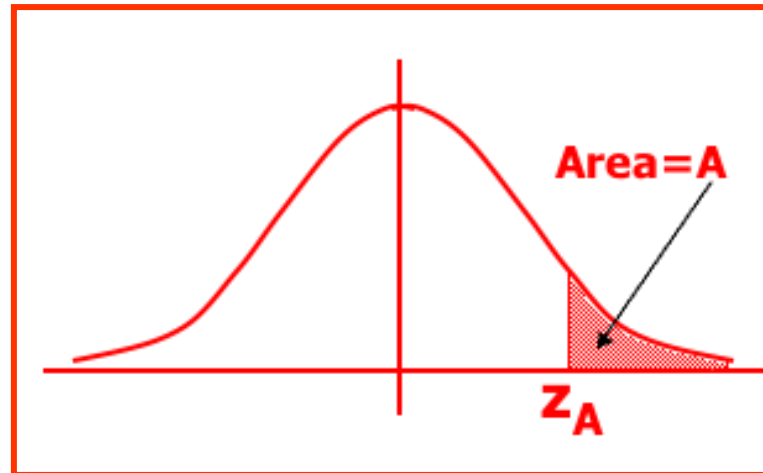
- Έτσι, όταν αυξάνεται η τυπική απόκλιση μεγαλώνει η πιθανότητα να χάσουμε λεφτά, από το οποίο φαίνεται ότι η τυπική απόκλιση (ή η διακύμανση) είναι ένα μέτρο ρίσκου.



Βρίσκοντας Τιμές Πιθανότητας για Δοσμένη Τιμή Z (1)

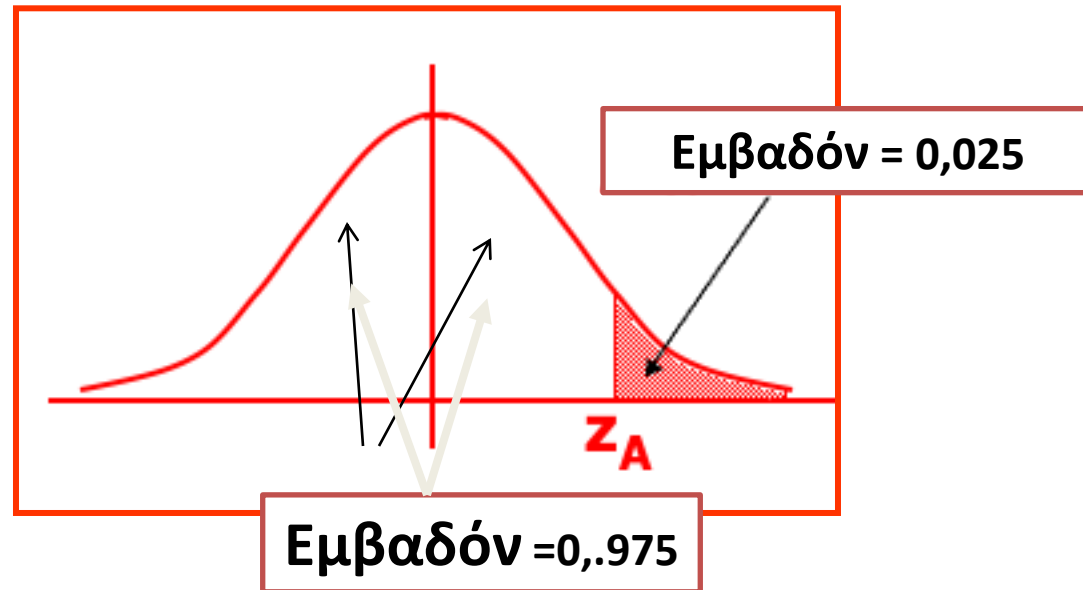
Συχνά ζητείται να βρούμε κάποια τιμή του Z για μία συγκεκριμένη τιμή πιθανότητας, δηλαδή δοθέντος ενός εμβαδού (A) κάτω από την καμπύλη, ποιά είναι η αντίστοιχη τιμή του z (z_A) στον οριζόντιο άξονα που δίνει αυτό το εμβαδόν; Δηλαδή:

$$P(Z > z_A) = A$$



Βρίσκοντας Τιμές Πιθανότητας για Δοσμένη Τιμή Z (2)

Ποια τιμή z αντιστοιχεί σε εμβαδόν, κάτω από την καμπύλη, ίσο με 2,5%; Δηλαδή, πόσο είναι το $z_{.025}$;



Εάν κάνουμε μία «αντίστροφη αναζήτηση» στο Πίνακα για 0,975, θα βρούμε την αντίστοιχη $z_A = 1,96$

Αφού $P(z > 1,96) = 0,025$, λέμε ότι $z_{.025} = 1,96$



Βρίσκοντας Τιμές Πιθανότητας για Δοσμένη Τιμή Z (3)

Άλλες Z τιμές είναι:

$$Z_{.05} = 1,645$$

$$Z_{.01} = 2,33$$

Αφού $z_{.025} = 1,96$ και $-z_{.025} = -1,96$, τότε:

$$P(-1,96 < Z < 1,96) = 0,95$$

Παρομοίως

$$P(-1,645 < Z < 1,645) = 0,90$$



Βρίσκοντας Τιμές Πιθανότητας για Δοσμένη Τιμή Z (4)

Αφού $z_{.025} = 1,96$ και $-z_{.025} = -1,96$, τότε:

$$P(-1,96 < Z < 1,96) = 0,95$$

Παρομοίως:

$$P(-1,645 < Z < 1,645) = 0,90$$



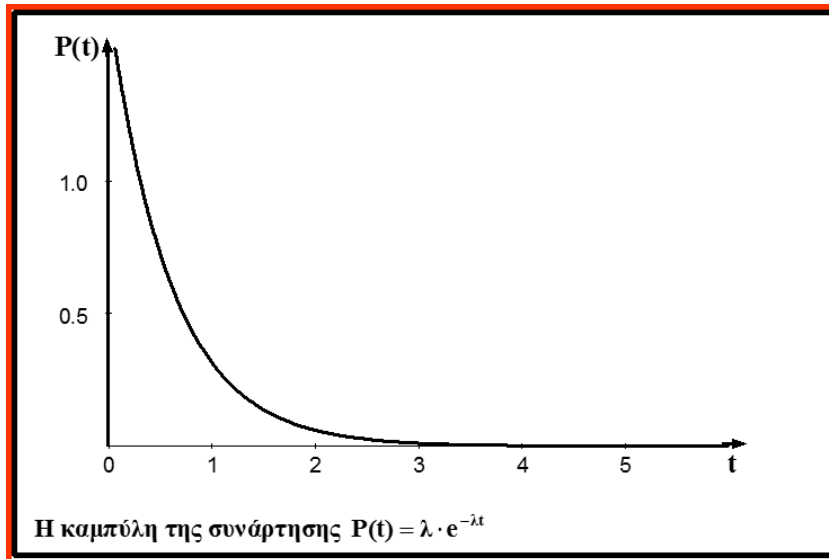
Εκθετική Κατανομή (1)

Όταν η συνεχής τυχαία μεταβλητή X έχει συνάρτηση πυκνότητας πιθανότητας την:

$$f(x) = \begin{cases} \frac{1}{\mu} e^{-\frac{x}{\mu}} & \text{για } x > 0 \\ 0 & \text{για } x \leq 0 \end{cases}$$

όπου $\mu > 0$

λέμε ότι αυτή ακολουθεί την εκθετική κανονική κατανομή.



$$P(t) = \lambda \cdot e^{-\lambda t}$$

Πηγή: Παπαδημητρίου (2001)



Εκθετική Κατανομή (2)

- Η εκθετική κατανομή μπορούμε να πούμε, ότι έχει αντίστροφη χρησιμότητα από την κατανομή Poisson. Ενώ, δηλαδή, η κατανομή Poisson μας δίνει την πιθανότητα ένα ενδεχόμενο να πραγματοποιηθεί x φορές μέσα σ' ένα χρονικό διάστημα t , η εκθετική κατανομή ορίζει την πιθανότητα να μεσολαβήσει ένα χρονικό διάστημα t μεταξύ δύο διαδοχικών πραγματοποιήσεων ενός ενδεχομένου.
- Η παράμετρος μ της εκθετικής κατανομής συνδέεται με το λ της κατανομής του Poisson με τη σχέση: $\mu = 1/\lambda$

$$E(t) = \mu = \frac{1}{\lambda}$$
$$Var(t) = \mu^2 = \frac{1}{\lambda^2}$$

Αθροιστική
Πιθανότητα

$$P(t > \alpha) = e^{-\lambda \alpha}$$

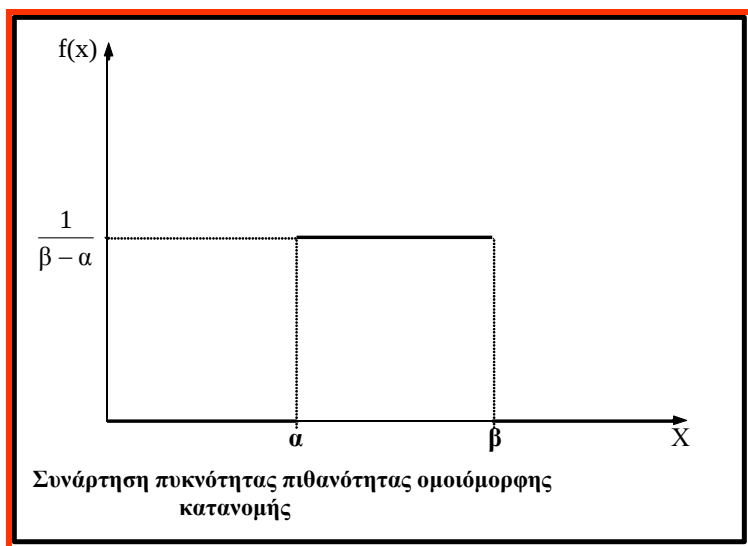


Ομοιόμορφη Κατανομή (1)

Αν η συνάρτηση πυκνότητας πιθανότητας της τυχαίας μεταβλητής X δίνεται από τη σχέση:

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{όταν } x \in [\alpha, \beta] \\ 0 & \text{όταν } x \notin [\alpha, \beta] \end{cases}$$

λέμε ότι η μεταβλητή X ακολουθεί την **ομοιόμορφη κατανομή** ή την **ορθογωνική κατανομή**. Παράμετροι αυτής της κατανομής είναι οι αριθμοί α και β ($\alpha < \beta$).



$$E(X) = \frac{\alpha + \beta}{2}$$

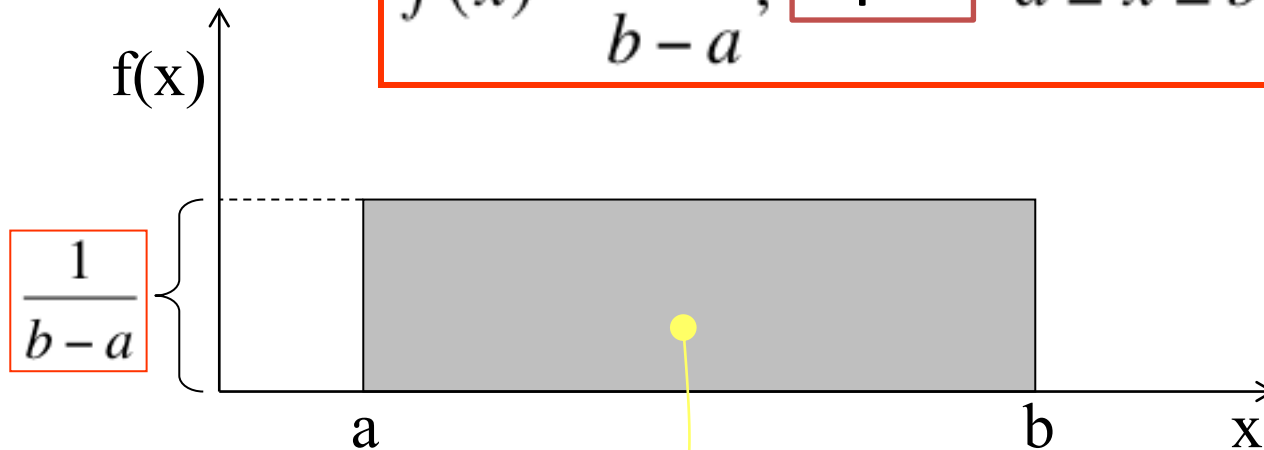
$$\text{Var}(X) = \frac{(\beta - \alpha)^2}{12}$$



Ομοιόμορφη Κατανομή (2)

Θεωρούμε την **ομοιόμορφη κατανομή πιθανότητας** (*uniform probability distribution*). Ορισμένες φορές καλείται ορθογώνια κατανομή πιθανότητας. Περιγράφεται από την συνάρτηση:

$$f(x) = \frac{1}{b-a}, \quad \text{με} \quad a \leq x \leq b$$

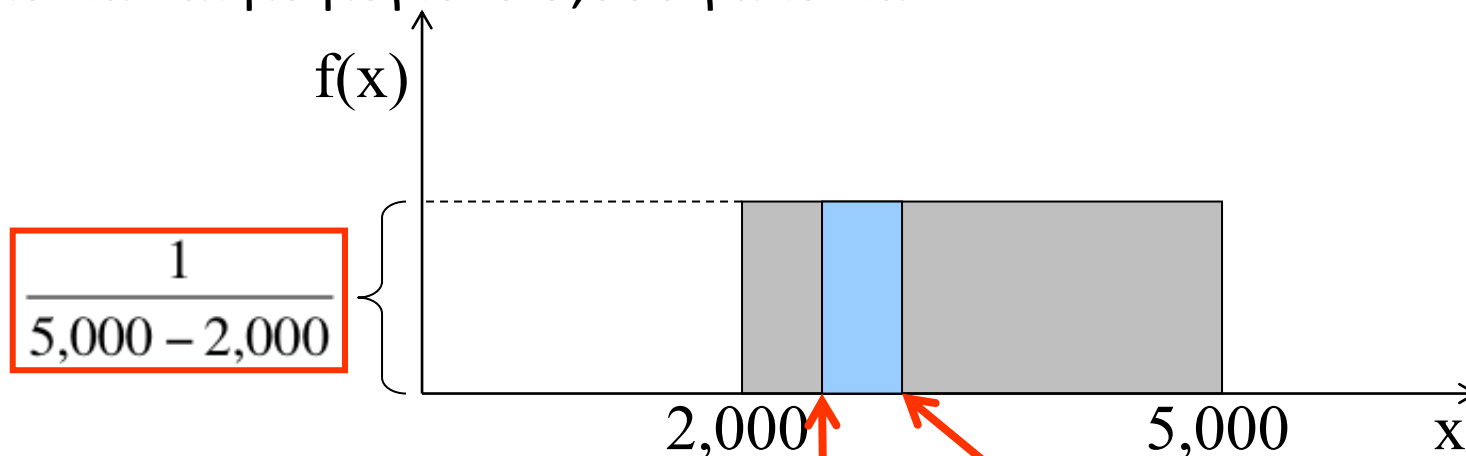


$$\text{εμβαδόν} = \text{μήκος} \times \text{ύψος} = (b-a) \times \frac{1}{b-a} = 1$$



Παράδειγμα (1)

Το ποσό της βενζίνης που πωλείται καθημερινώς σε ένα βενζινάδικο είναι ομοιόμορφα κατανομημένο με ελάχιστο 2,000 γαλόνια και με μέγιστο 5,000 γαλόνια.



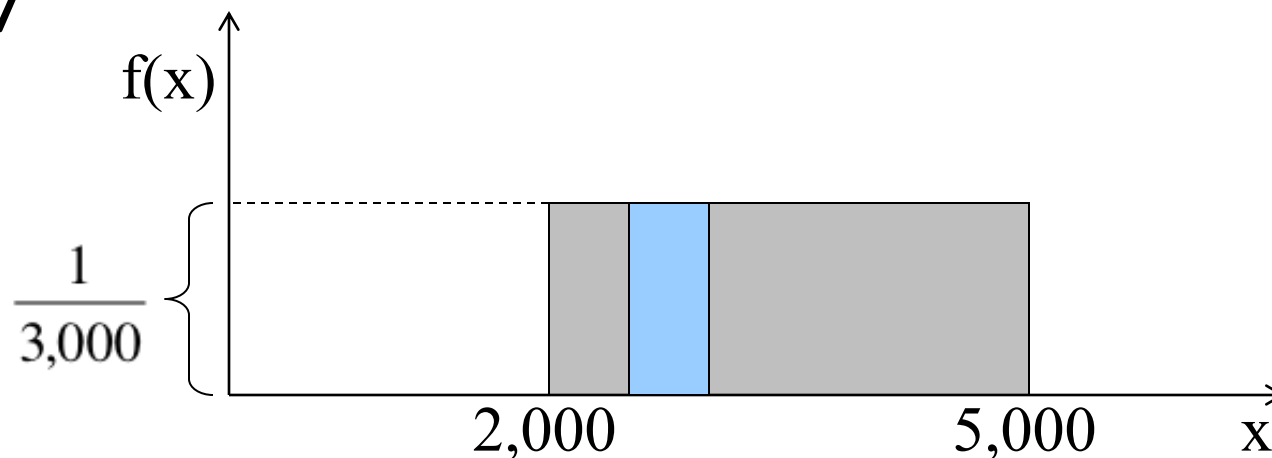
Βρείτε την πιθανότητα ότι οι ημερήσιες πωλήσεις θα είναι μεταξύ 2,500 και 3,000 γαλόνια.

Αλγεβρικά: ποια είναι η $P(2,500 \leq X \leq 3,000)$;,,

Παράδειγμα (2)

$$P(2,500 \leq X \leq 3,000) = (3,000 - 2,500) \times .1667$$

$$\frac{1}{3,000}$$

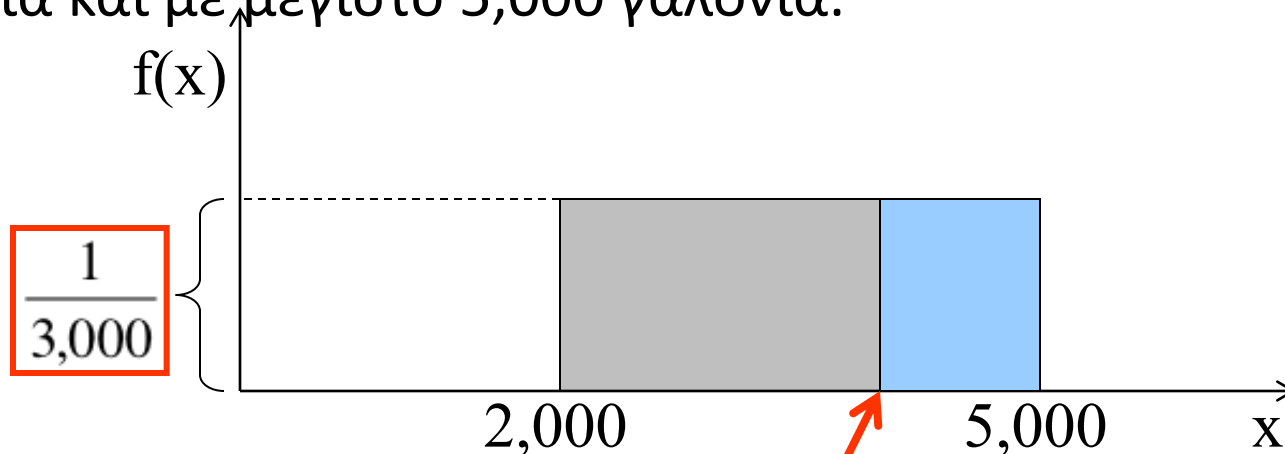


Υπάρχει περίπου 17% πιθανότητα ότι σε μία μέρα θα πουληθούν 2,500-3,000 γαλόνια



Παράδειγμα (3)

Το ποσό της βενζίνης που πωλείται καθημερινώς σε ένα βενζινάδικο είναι ομοιόμορφα κατανεμημένο με ελάχιστο 2,000 γαλόνια και με μέγιστο 5,000 γαλόνια.



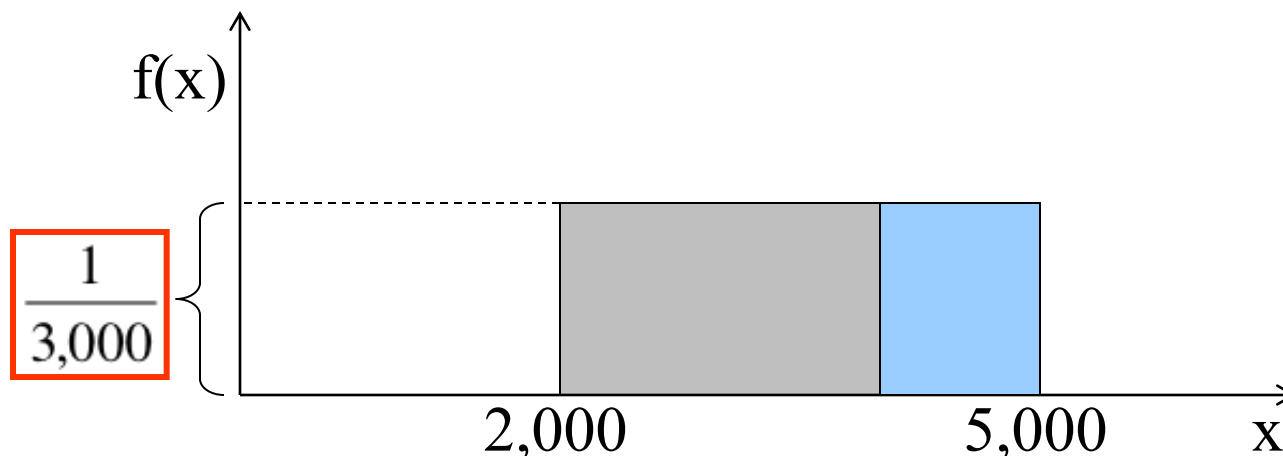
Ποια είναι η πιθανότητα ότι το βενζινάδικο θα πουλήσει τουλάχιστον 4,000 γαλόνια;

Αλγεβρικά: ποια είναι η $P(X \geq 4,000)$;



Παράδειγμα (4)

$$P(X \geq 4,000) = (5,000 - 4,000) \times \frac{1}{3,000} = .3333$$

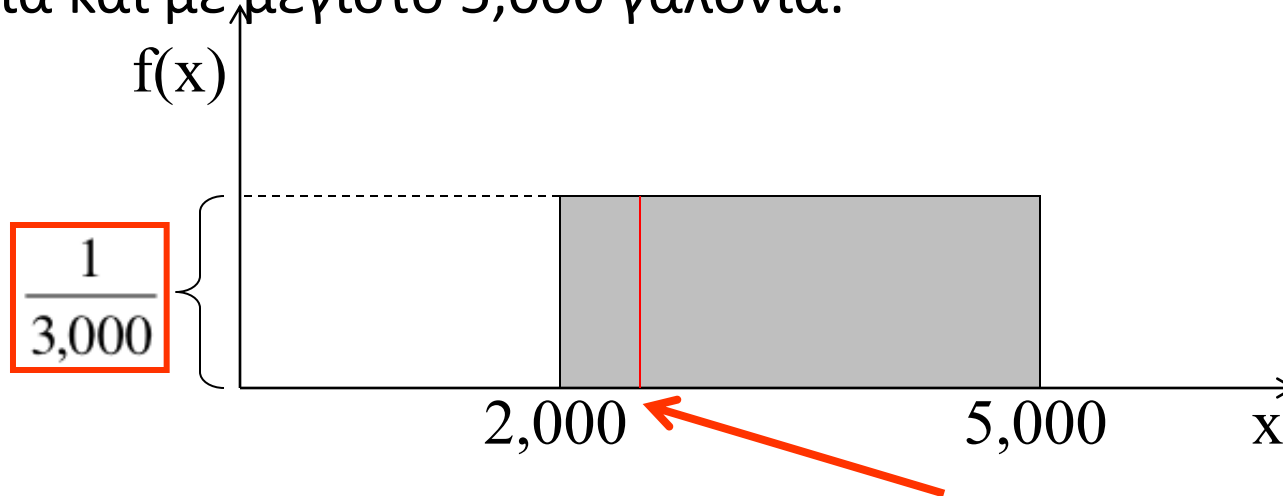


Υπάρχει μία στις τρεις περιπτώσεις το βενζινάδικο να πουλήσει περισσότερο από 4,000 γαλόνια σε μια οποιαδήποτε ημέρα.



Παράδειγμα (5)

Το ποσό της βενζίνης που πωλείται καθημερινώς σε ένα βενζινάδικο είναι ομοιόμορφα κατανεμημένο με ελάχιστο 2,000 γαλόνια και με μέγιστο 5,000 γαλόνια.



Ποια είναι η πιθανότητα ότι το βενζινάδικο θα πουλήσει ακριβώς 2,500 γαλόνια;

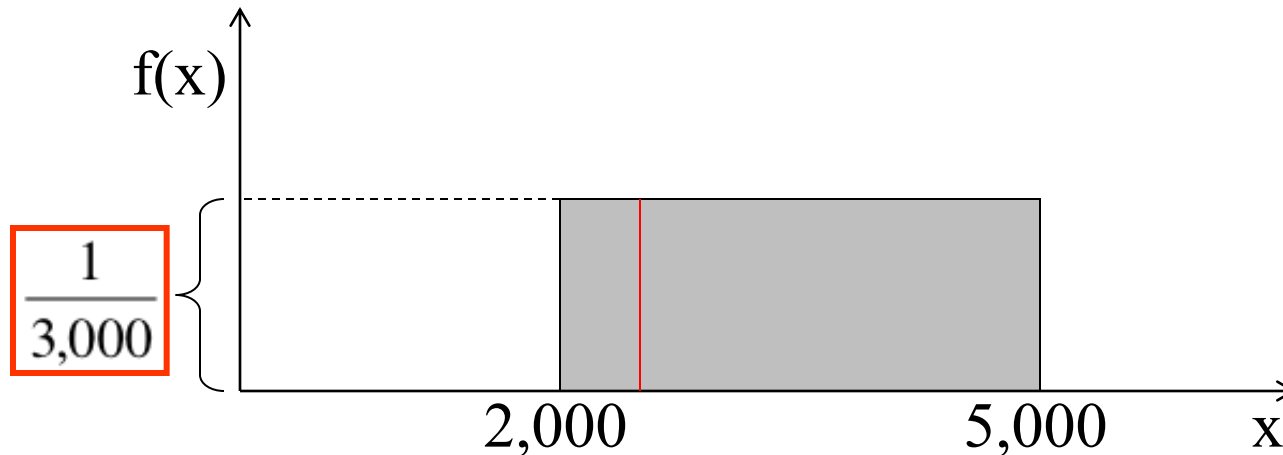
Αλγεβρικά: ποια είναι η $P(X = 2,500)$;



Παράδειγμα (6)

$$P(X = 2,500) = (2,500 - 2,500) \times$$

$$\frac{1}{3,000}$$



Η πιθανότητα ότι το βενζινάδικο θα πουλήσει ακριβώς 2,500 γαλόνια είναι ίση με 0.



Λογαριθμική Κανονική Κατανομή

Μια τυχαία μεταβλητή X θα λέμε ότι ακολουθεί την λογαριθμική κανονική κατανομή, όταν η τυχαία μεταβλητή Y , που ορίζεται από τη σχέση :

$$Y = \alpha \cdot \log X + \beta \quad 0 < X < \infty$$

ή από τη σχέση:

$$Y = \alpha \cdot \log(X - X_0) + \beta \quad X_0 < X < \infty$$

ακολουθεί την κανονική κατανομή.

$$\log[E(X) - X_0] = \frac{1}{2\alpha^2} - \frac{\beta}{\alpha}$$

$$\log\sqrt{\text{Var}(X)} = -\frac{\beta}{\alpha} + \frac{1}{2\alpha^2} + \frac{1}{2}\log(e^{\frac{1}{\alpha^2}} - 1)$$

Πηγή: Παπαδημητρίου (2001)



Εμπειρικές Κατανομές



Κατανομή χ^2

Έστω ότι έχουμε τις τυχαίες μεταβλητές $x_1, x_2, x_3, \dots, x_v$ που είναι ανεξάρτητες μεταξύ τους και που η κάθε μία από αυτές κατανέμεται κανονικά με μέση τιμή 0 και διακύμανση 1. Έστω, δηλαδή, ότι οι ανεξάρτητες μεταξύ τους τυχαίες μεταβλητές ακολουθούν την $N(0,1)$.

Το άθροισμα όμως των τετραγώνων τους είναι μια άλλη τυχαία μεταβλητή :

$$\chi^2 = x_1^2 + x_2^2 + \dots + x_v^2$$

που η συνάρτηση πυκνότητας πιθανότητας της δεν ακολουθεί την κανονική κατανομή αλλά μια άλλη κατανομή, που ονομάζεται **χ τετράγωνο κατανομή** και συμβολίζεται με “ χ^2 κατανομή”, με συνάρτηση πυκνότητας πιθανότητας την:

$$f(\chi^2) = \begin{cases} \frac{1}{2^{\frac{v}{2}} \Gamma\left(\frac{v}{2}\right)} \cdot e^{-\frac{\chi^2}{2}} \cdot \chi^{\frac{v}{2}-1} & \text{όταν } \chi > 0 \\ 0 & \text{όταν } \chi < 0 \end{cases}$$

με v βαθμούς ελευθερίας.

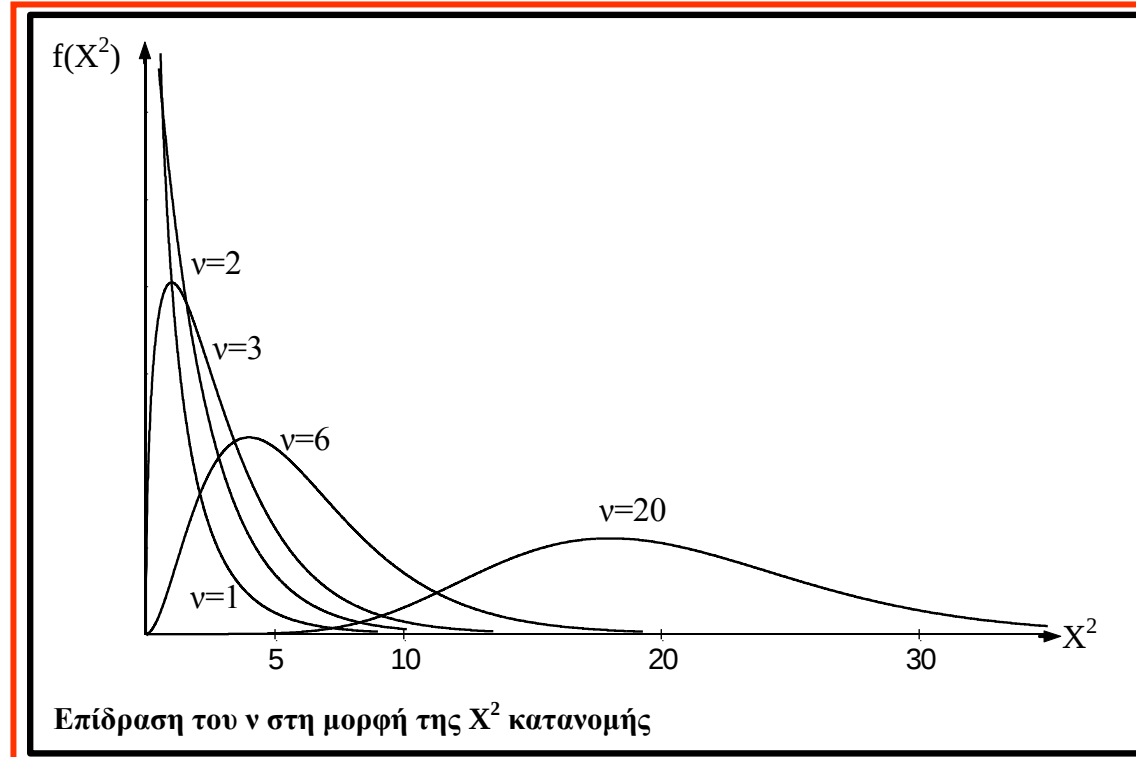
Πηγή: Παπαδημητρίου (2001)



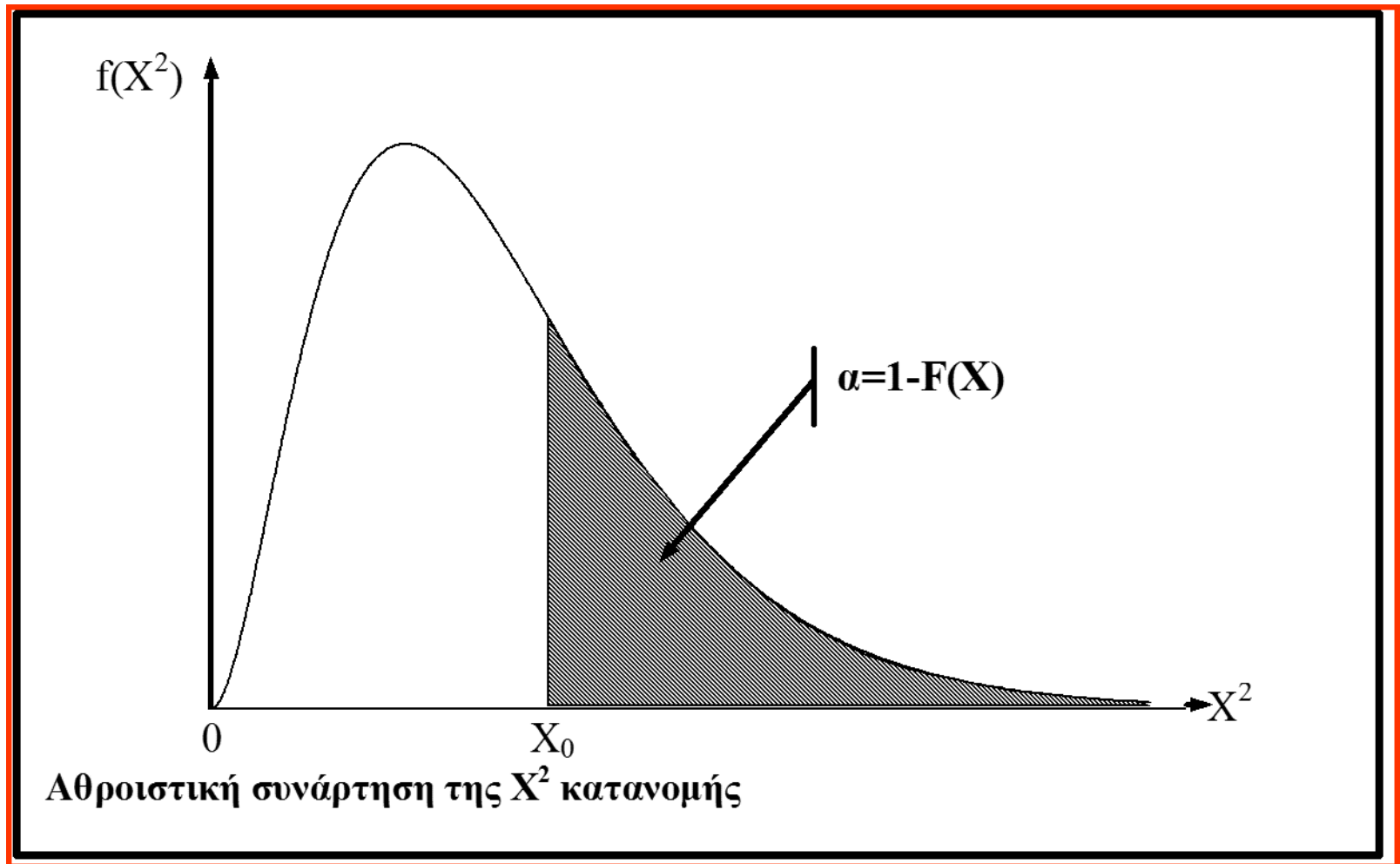
Κατανομή X^2 (συνέχεια)

$$E(X^2)=v$$

$$\text{Var}(X^2)=2v$$



Κατανομή X^2 (συνέχεια)



Κατανομή t (Student)

Αν έχουμε δύο ανεξάρτητες μεταξύ τους τυχαίες μεταβλητές X και Y , από τις οποίες η X έχει κατανομή πιθανότητας που ακολουθεί την κανονική κατανομή με μέση τιμή μηδέν και διακύμανση μονάδα ($N(0,1)$) και η Y ακολουθεί την χ^2 κατανομή με ν βαθμούς ελευθερίας, τότε η τυχαία μεταβλητή

$$T = \frac{X}{\sqrt{Y/\nu}}$$

ακολουθεί την t κατανομή με ν βαθμούς ελευθερίας.

$$f(t_\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi} \cdot \Gamma\left(\frac{\nu}{2}\right)} \cdot \frac{1}{\left(1 + \frac{t^2}{\nu}\right)^{\frac{\nu+1}{2}}} \quad \text{για } -\infty < t < +\infty$$

Πηγή: Παπαδημητρίου (2001)



Κατανομή t (συνέχεια)

Αν πάρουμε ένα τυχαίο δείγμα $\{x_1, x_2, x_3, \dots, x_n\}$ από έναν κανονικό πληθυσμό τότε, η τυχαία μεταβλητή $X = \bar{X}$ ακολουθεί την κανονική κατανομή, επομένως η $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ ακολουθεί την κατανομή $N(0,1)$, ενώ η $Y = \frac{S^2}{\sigma^2}$ ακολουθεί την χ^2 κατανομή με $n-1$ βαθμούς ελευθερίας.

Είναι φανερό, από τον ορισμό που δώσαμε στην t κατανομή ότι η τυχαία μεταβλητή T

$$T = \frac{Z}{\sqrt{Y}} = \frac{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{S^2}{\sigma^2}}} = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

ακολουθεί την t κατανομή με $n-1$ βαθμούς ελευθερίας.

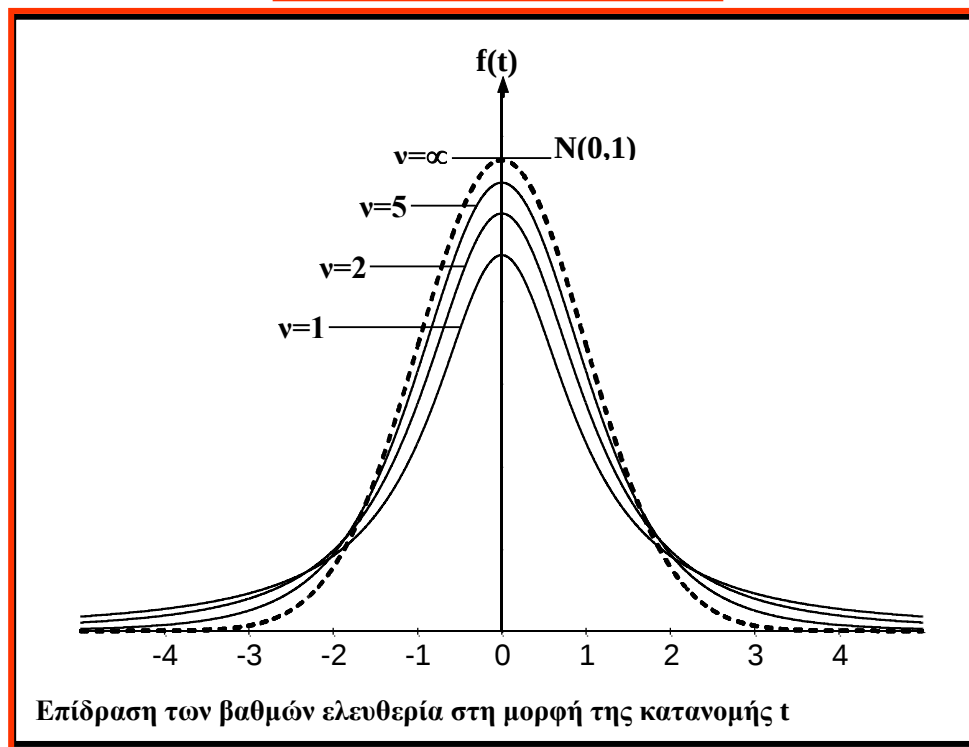
Πηγή: Παπαδημητρίου (2001)



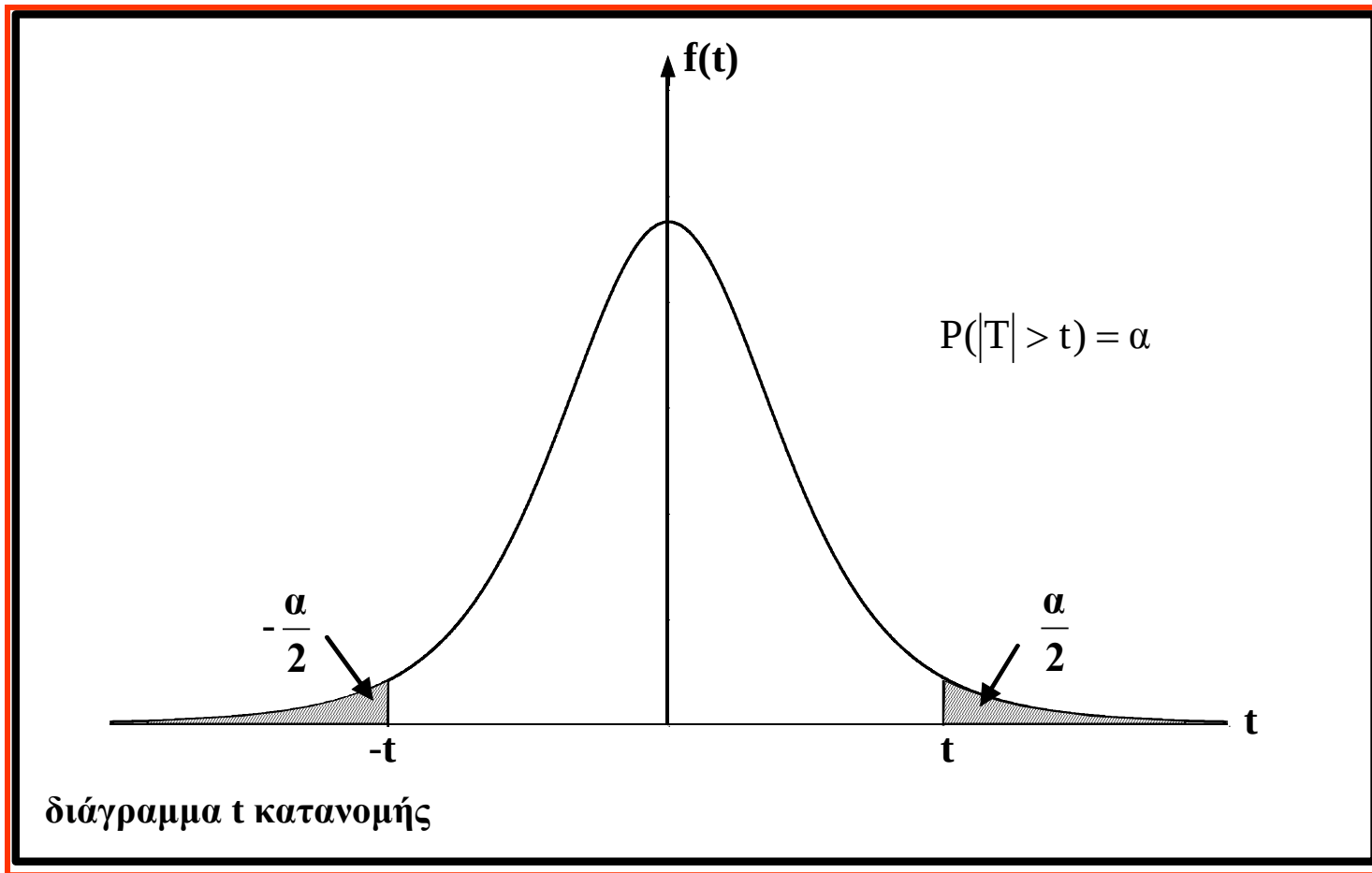
Κατανομή t (συνέχεια)

$$E(t) = 0$$

$$\text{Var}(t) = \frac{v}{v-2}$$



Κατανομή t (συνέχεια)

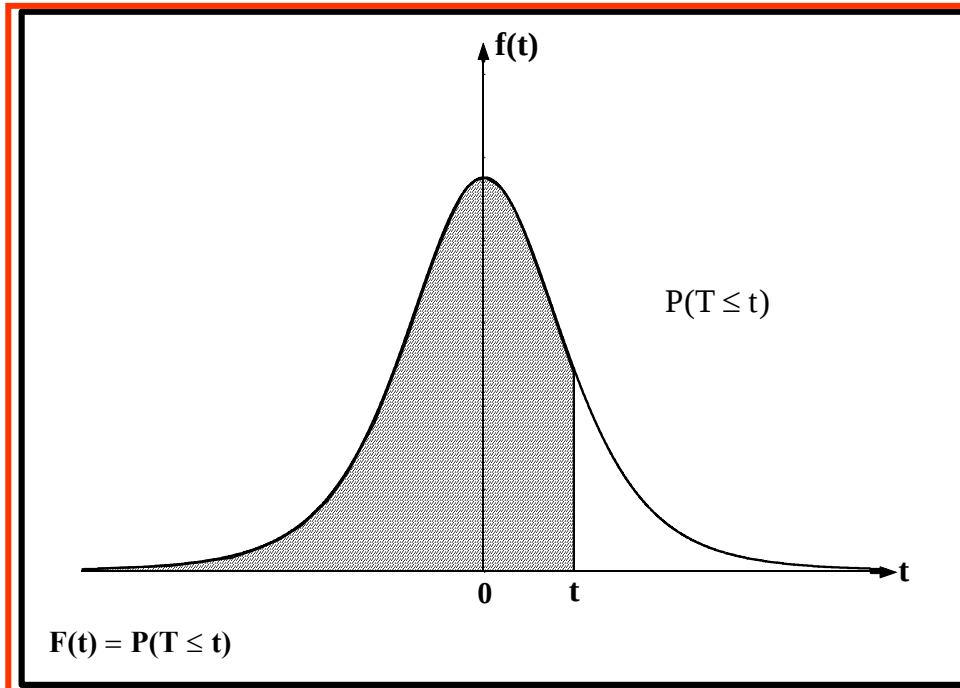


Πηγή: Παπαδημητρίου (2001)

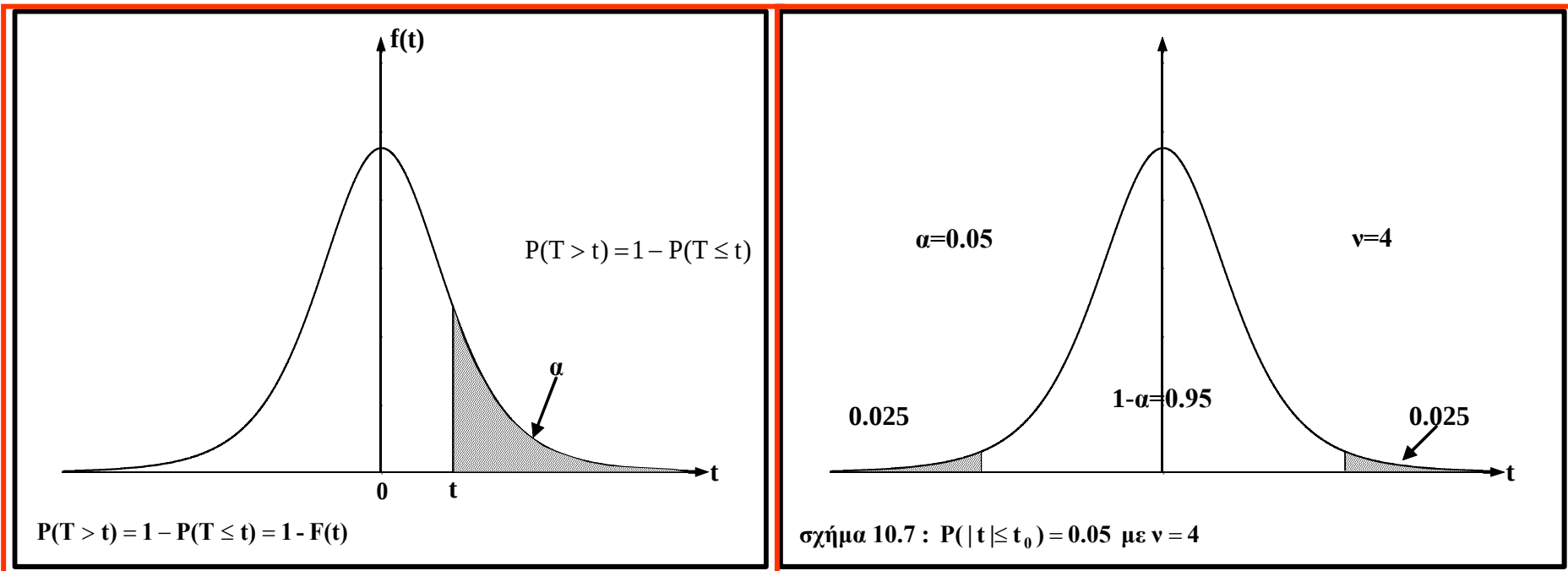


Κατανομή t (συνέχεια)

$$F(t) = P(T \leq t) = \int_{-\infty}^t \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi} \cdot \Gamma\left(\frac{v}{2}\right)} \cdot \frac{1}{\left(1 + \frac{t^2}{v}\right)^{\frac{v+1}{2}}} dt$$



Κατανομή t (συνέχεια)



Πηγή: Παπαδημητρίου (2001)



F Κατανομή

Αν έχουμε δύο ανεξάρτητες μεταξύ τους τυχαίες μεταβλητές X και Y , των οποίων οι κατανομές πυκνότητας πιθανότητας ακολουθούν την X^2 κατανομή με ν_1 και ν_2 βαθμούς ελευθερίας αντίστοιχα, τότε η τυχαία μεταβλητή :

$$F = \frac{X/\nu_1}{Y/\nu_2}$$

έχει κατανομή πυκνότητας πιθανότητας που ακολουθεί την F_{ν_1, ν_2} κατανομή με ν_1 και ν_2 βαθμούς ελευθερίας. Η κατανομή F , που είναι λόγος δύο τυχαίων μεταβλητών που ακολουθούν την X^2 κατανομή, ονομάζεται και **λόγος των διακυμάνσεων**.

$$E(F) = \frac{\nu_2}{\nu_2 - 2}$$

για $\nu_2 > 2$

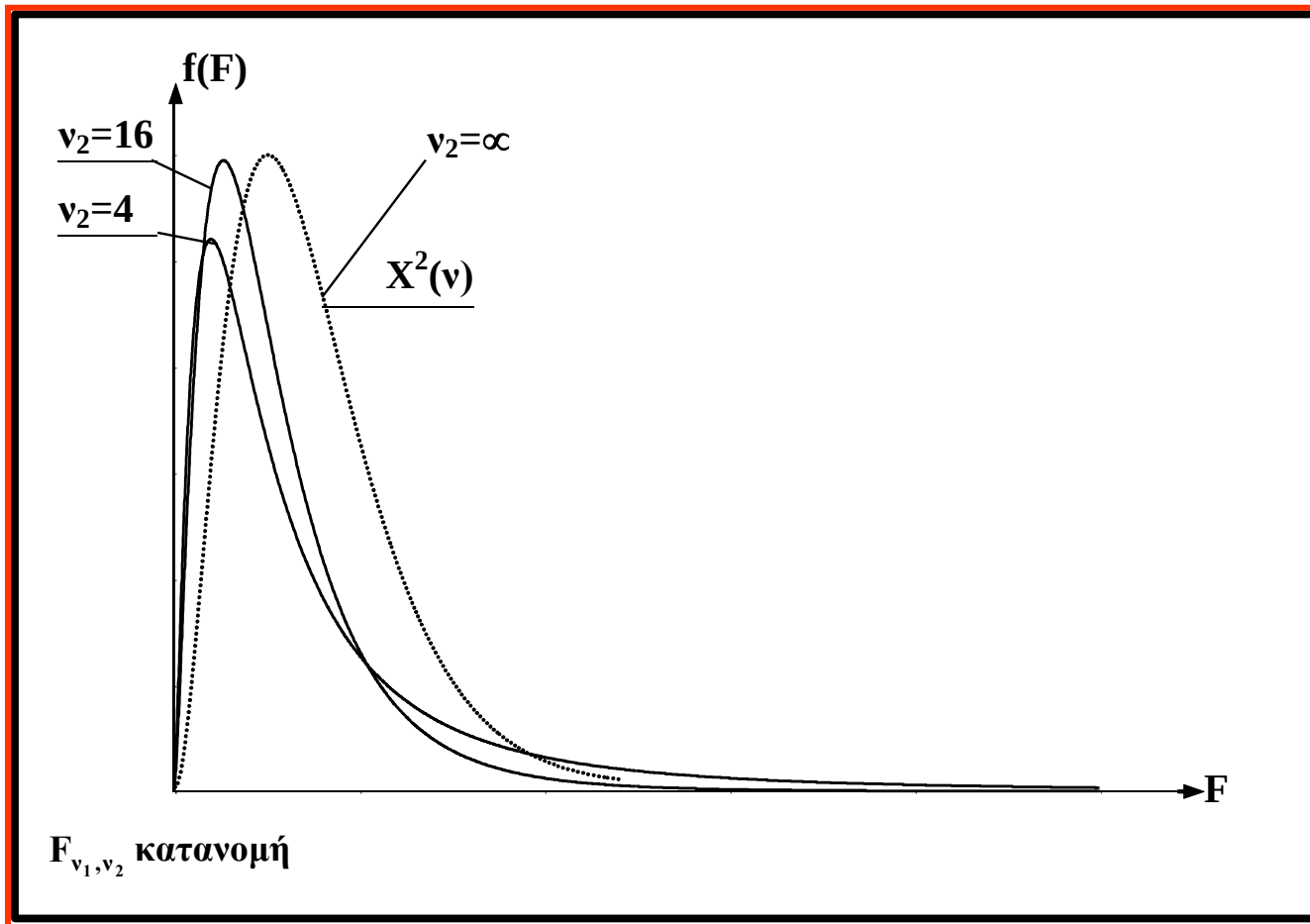
$$\text{Var}(F) = \frac{2\nu_2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)}$$

για $\nu_2 > 4$

Πηγή: Παπαδημητρίου (2001)



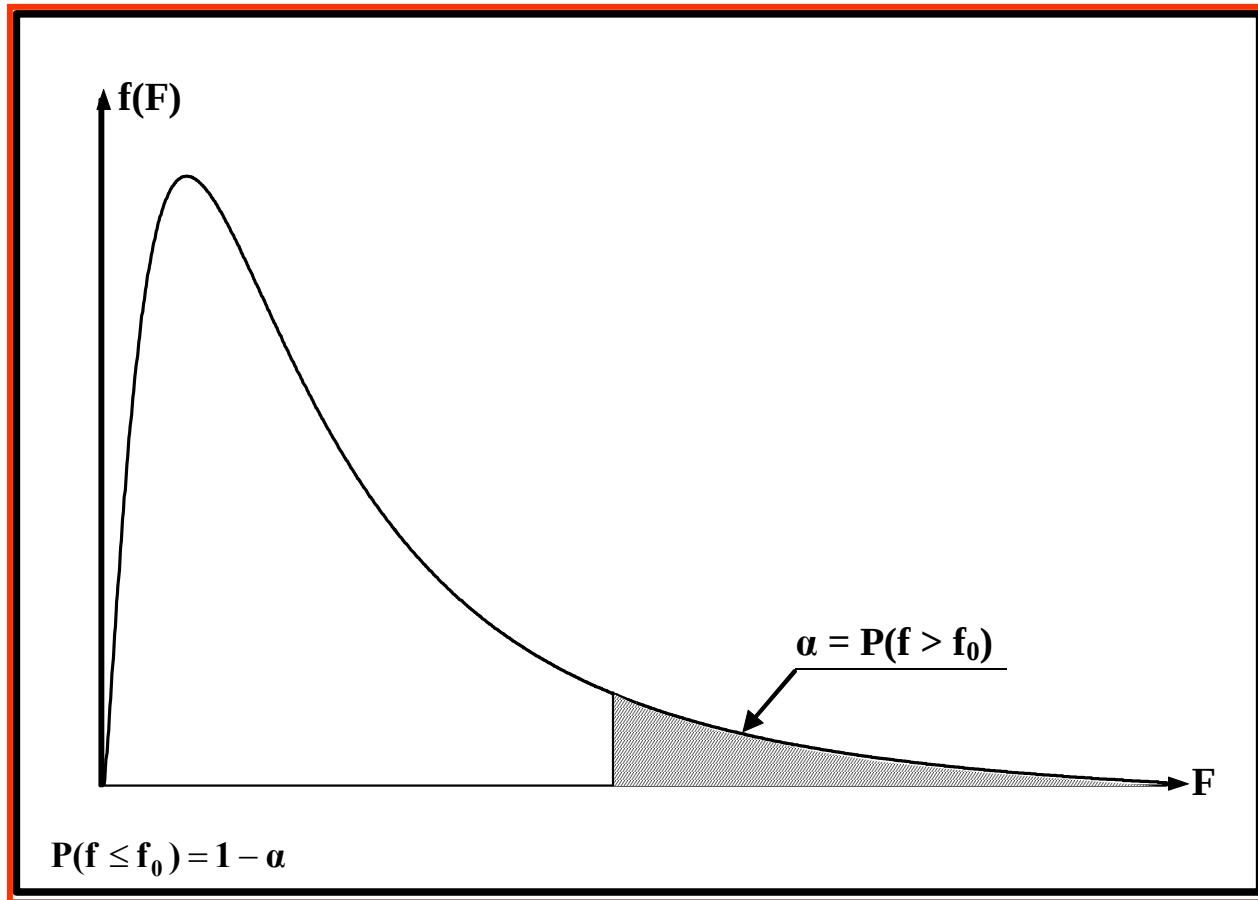
F Κατανομή (συνέχεια)



Πηγή: Παπαδημητρίου (2001)



F Κατανομή (συνέχεια)



Πηγή: Παπαδημητρίου (2001)



F Κατανομή (συνέχεια)

Η F_{v_1, v_2} κατανομή είναι πολύ χρήσιμη, διότι μ' αυτή μπορούμε να εκφράσουμε την κατανομή δειγματοληψίας του λόγου των διακυμάνσεων, που αναφέρονται σε δύο τυχαία δείγματα που προέρχονται από κανονικούς πληθυσμούς. Αν, δηλαδή, έχουμε τα δύο τυχαία δείγματα $\{X_1, X_2, \dots, X_{v_1}\}$ και $\{Y_1, Y_2, \dots, Y_{v_2}\}$ που προέρχονται από πληθυσμούς που ακολουθούν τους νόμους $N(\mu_1, \sigma_1^2)$, $N(\mu_2, \sigma_2^2)$ αντίστοιχα και S_1^2 , S_2^2 οι διακυμάνσεις των δύο αυτών δειγμάτων, όταν $\sigma_1^2 = \sigma_2^2$, η τυχαία μεταβλητή :

$$F = \frac{s_1^2}{s_2^2}$$

έχει κατανομή πυκνότητας πιθανότητας που ακολουθεί την F κατανομή με $v_1 - 1$ και $v_2 - 1$ βαθμούς ελευθερίας.

Πηγή: Παπαδημητρίου (2001)



Κεντρικό Οριακό Θεώρημα

Η δειγματική κατανομή της μέσης τιμής $\bar{X}$ ενός τυχαίου δείγματος, το οποίο εξάγεται από οποιοδήποτε πληθυσμό, είναι προσεγγιστικά κανονική για επαρκώς **μεγάλο μέγεθος δείγματος**.

Όσο πιο μεγάλο είναι το μέγεθος του δείγματος, τόσο πιο πολύ η δειγματική κατανομή της μέσης τιμής $\bar{X}$ θα μοιάζει με κανονική κατανομή.



Κεντρικό Οριακό Θεώρημα (συνέχεια)

Εάν ο πληθυσμός είναι κανονικός, τότε η δειγματική μέση τιμή $\bar{X}$ είναι κανονικά κατανεμημένη για όλες τις τιμές του n .

Εάν ο πληθυσμός είναι μη-κανονικός, τότε η δειγματική μέση τιμή είναι προσεγγιστικά κανονική μόνο για μεγάλες του n .

Σε πολλές πρακτικές περιπτώσεις, ένα μέγεθος δείγματος ίσο με 30 ενδέχεται να είναι επαρκώς μεγάλο ώστε να επιτρέπει την χρήση της κανονικής κατανομής ως προσέγγιση για την δειγματική κατανομή της μέσης τιμής της τ.μ. X .



Βιβλιογραφία

- Παπαδημητρίου, Γ. (2001). *Περιγραφική Στατιστική*. Θεσσαλονίκη: Παρατηρητής.
- Φωτιάδης, Ν. (1995). *Εισαγωγή στη Στατιστική για Βιολογικές Επιστήμες*. Θεσσαλονίκη: University Studio Press.
- Κουνιάς, Σ. & Μωυσιάδης, Χ. (1985). *Πιθανότητες Ι: Θεωρία και Ασκήσεις*. Θεσσαλονίκη: Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, Έκδοση: Υπηρεσία Δημοσιευμάτων.
- Κουνιάς, Σ., Κολυβά-Μαχαίρα, Φ., Μπαγιάτης, Κ. & Μπόρας-Σέντα, Ε. (1985). *Εισαγωγή στη Στατιστική*. Θεσσαλονίκη.
- Ελληνικό Ανοικτό Πανεπιστήμιο (2003). *Σημειώσεις Στατιστικής*.



Σημείωμα Αναφοράς

Copyright Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, Γεώργιος Μενεξές.
«Στατιστική. Θεωρητικές και Εμπειρικές - Δειγματοληπτικές Κατανομές».
Έκδοση: 1.0. Θεσσαλονίκη 2014. Διαθέσιμο από τη δικτυακή διεύθυνση:
<https://opencourses.auth.gr/courses/OCRS484/>



Σημείωμα Αδειοδότησης

Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά - Μη Εμπορική Χρήση - Όχι Παράγωγα Έργα 4.0 [1] ή μεταγενέστερη, Διεθνής Έκδοση. Εξαιρούνται τα αυτοτελή έργα τρίτων π.χ. φωτογραφίες, διαγράμματα κ.λ.π., τα οποία εμπεριέχονται σε αυτό και τα οποία αναφέρονται μαζί με τους όρους χρήσης τους στο «Σημείωμα Χρήσης Έργων Τρίτων».



Ο δικαιούχος μπορεί να παρέχει στον αδειοδόχο ξεχωριστή άδεια να χρησιμοποιεί το έργο για εμπορική χρήση, εφόσον αυτό του ζητηθεί.

Ως **Μη Εμπορική** ορίζεται η χρήση:

- που δεν περιλαμβάνει άμεσο ή έμμεσο οικονομικό όφελος από την χρήση του έργου, για το διανομέα του έργου και αδειοδόχο
- που δεν περιλαμβάνει οικονομική συναλλαγή ως προϋπόθεση για τη χρήση ή πρόσβαση στο έργο
- που δεν προσπορίζει στο διανομέα του έργου και αδειοδόχο έμμεσο οικονομικό όφελος (π.χ. διαφημίσεις) από την προβολή του έργου σε διαδικτυακό τόπο

[1] <http://creativecommons.org/licenses/by-nc-nd/4.0/>





Τέλος Ενότητας

Επεξεργασία: Μαρία Αλεμπάκη
Θεσσαλονίκη, Εαρινό Εξάμηνο 2014-2015

