ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ
UNIVERSITY OF PATRAS

ΑΝΟΙΚΤΑ ακαδημαϊκά μαθήματα ΠΠ

ΤΙΤΛΟΣ ΜΑΘΗΜΑΤΟΣ: **ΣΤΑΤΙΣΤΙΚΗ**

ΕΝΟΤΗΤΑ: **Πιθανότητες - Κατανομές**

ΟΝΟΜΑ ΚΑΘΗΓΗΤΗ: **ΦΡ. ΚΟΥΤΕΛΙΕΡΗΣ**

ΤΜΗΜΑ**: Τμήμα Διαχείρισης Περιβάλλοντος και Φυσικών Πόρων**

ΑΓΡΙΝΙΟ

ΣΤΑΤΙΣΤΙΚΗ

Φραγκίσκος Κουτελιέρης

Αναπληρωτής Καθηγητής Παν/μίου Πατρών

Επικοινωνία: fcoutelieris@upatras.gr

Τυχαία γεγονότα

Δυνατά αποτελέσματα (**ΕΝΔΕΧΟΜΕΝΑ**)

S = {1,2,3,4,5,6}

Ως **τυχαία** ορίζεται η μεταβλητή που περιλαμβάνει τιμές, οι οποίες προέρχονται από μια **τυχαία διαδικασία**.

Ρίχνουμε ένα ζάρι

Ορισμοί

ΕΝΔΕΧΟΜΕΝΑ = όλα τα δυνατά αποτελέσματα

Στο παράδειγμα του **ζαριού**, S = {1,2,3,4,5,6}

ΓΕΓΟΝΟΣ = κάθε υποσύνολο δυνατών αποτελεσμάτων

Στο παράδειγμα του **ζαριού**, A = {2} π.χ.

Πιθανότητα

ΠΙΘΑΝΟΤΗΤΑ = ο λόγος των ευνοϊκών γεγονότων προς όλα τα δυνατά αποτελέσματα

$$P(A) = \frac{\text{Ευνοϊκά γεγονότα}}{\text{Όλα τα δυνατά γεγονότα}}$$

Στο παράδειγμα του **ζαριού**, P(A=2)=1/6

Ιδιότητες

- $P(A) \geq 0$
- $P(A) \leq 1$
- Άθροισμα των πιθανοτήτων όλων των ενδεχομένων = 1

Στο παράδειγμα του **ζαριού**,

$P(A=1) + P(A=2) + P(A=3) + P(A=4) + P(A=5) + P(A=6) = 1$

Σχέσεις μεταξύ τυχαίων γεγονότων

Από κοινού πιθανότητα των A και B είναι η πιθανότητα να συμβούν και τα δυο γεγονότα: $P(A,B) = P(A \cap B)$

Δεσμευμένη πιθανότητα των A και B είναι η πιθανότητα να συμβεί το A δεδομένου ότι συνέβη ήδη το B:

$$P(A|B) = \frac{P(A,B)}{P(B)}$$

Ανεξαρτησία

Τα γεγονότα A και B είναι **στατιστικώς ανεξάρτητα** όταν:

$$P(A,B) = P(A)P(B)$$

οπότε:

$$P(A \mid B) = P(A) \qquad P(B \mid A) = P(B)$$

Παράδειγμα

Έστω το παράδειγμα του ζαριού και έστω τα γεγονότα **A={3}** και **B={1,2,3,6}**

- $P(A)=1/3$
- $P(B)=4/6=2/3$
- $P(A,B)=1/6$ διότι $A \cap B=\{3\}$
- $P(A|B) = P(A,B)/P(B) = 1/4$ διότι $(1/6)/(2/3)= (1/6)/(4/6)=1/4$
- Είναι ανεξάρτητα? Όχι

Παράδειγμα

Έστω μια οικογένεια που έχει δυο παιδιά. Ποια η πιθανότητα να είναι **και τα δυο** αγόρια, δεδομένου ότι το ένα είναι αγόρι?

S={(α,α),(α,κ),(κ,α),(κ,κ)}
άρα
P(A=(α,α))=1/4

ΕΙΝΑΙ ΣΩΣΤΟ?
ΌΧΙ,
διότι ξέρουμε πως το ένα παιδί είναι αγόρι!

Παράδειγμα

Έστω μια οικογένεια που έχει δυο παιδιά. Ποια η πιθανότητα να είναι **και τα δυο** αγόρια, δεδομένου ότι το ένα είναι αγόρι?

$S=\{(α,α),(α,κ),(κ,α),(κ,κ)\}$

$A=$(δυο αγόρια)$=\{(α,α)\}$

$B=$(ένα τουλάχιστον αγόρι)$=\{(α,α),(α,κ),(κ,α)\}$

Άρα

$P(A)=P(A|B)/(P(B)= P(\{(α,α)\})/P(\{(α,α),(α,κ),(κ,α)\})=\mathbf{1/3}$

ΑΣΚΗΣΗ

Σε μια πόλη τρεις μάρκες αυτοκίνητων, A, B και C κατέχουν το 20%, 30% και 50% της αγοράς, αντίστοιχα.

Η πιθανότητα να χρειαστεί ένα αυτοκίνητο επισκευή τον πρώτο χρόνο κυκλοφορίας του είναι 5%, 10% και 15%, αντίστοιχα.

1. Ποια είναι η πιθανότητα επισκευής ενός αυτοκινήτου τον πρώτο χρόνο κυκλοφορίας του? 11.5%
2. Αν ένα αυτοκίνητο έχει ανάγκη επισκευής τον πρώτο χρόνο, ποια είναι η πιθανότητα να είναι μάρκας A? 8.7%

Κατανομή Πιθανότητας

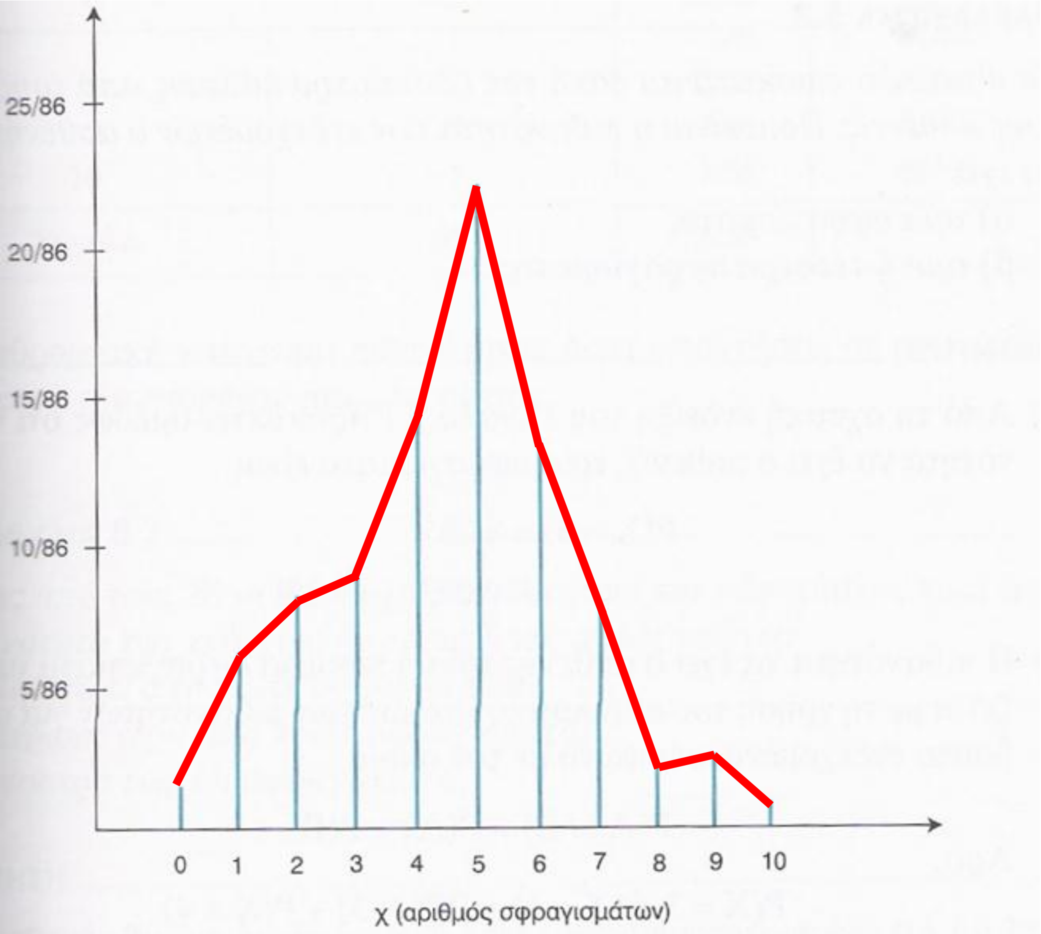
- Επίσκεψη σε οδοντίατρο 86 ασθενών.

- Έστω X η τυχαία μεταβλητή που εκφράζει το **πλήθος των σφραγισμάτων** του κάθε ασθενή

Κατανομή Πιθανότητας

Αριθμός Σφραγισμάτων (x_i)	Συχνότητα Εμφάνισης της τιμής x_i	$P(X = x_i)$
0	1	1/86
1	6	6/86
2	8	8/86
3	9	9/86
4	15	15/86
5	22	22/86
6	13	13/86
7	7	7/86
8	2	2/86
9	2	2/86
10	1	1/86
Σύνολο	86	86/86 (=1)

Κατανομή Πιθανότητας



Κατανομή Πιθανότητας

• **εμπειρικές κατανομές πιθανότητας:** οι κατανομές πιθανότητας που καταρτίζονται βάσει των παρατηρήσεων που προκύπτουν από πραγματικές μετρήσεις

• **θεωρητικές κατανομές πιθανότητας:** οι κατανομές πιθανότητας που ακολουθούν κάποιο γνωστό θεωρητικό υπόδειγμα

Βασικές Κατανομές Πιθανότητας

- Διωνυμική κατανομή
- Κατανομή Poisson
- Κανονική κατανομή
- Τυχαία κατανομή
- …

Χαρακτηριστικά (1)

Μέσος όρος (mean): Το άθροισμα των τιμών διά το πλήθος τους

Πλεονέκτημα: εύκολα αντιληπτός και ευρέως γνωστός

Μειονέκτημα: ευαισθησία στις ακραίες τιμές

Χαρακτηριστικά (2)

Διάμεσος (median): Αν παρατάξουμε τις τιμές κατά αύξουσα σειρά, η μεσαία τιμή

Πλεονέκτημα: δεν επηρεάζεται από τις μεσαίες τιμές

Μειονέκτημα: -

Μέτρα διασποράς

- Εύρος τιμών (range)
- Διακύμανση (variance)
- Τυπική απόκλιση (standard deviation)
- Τυπικό σφάλμα (standard error)
- Συντελεστής μεταβλητότητας ή συντελεστής διακύμανσης (coefficient of variation)

Εύρος τιμών (range)

Η διαφορά μεγίστου ελαχίστου

Διακύμανση (variance)

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

Τυπική απόκλιση (standard deviation)

$$s = \sqrt{s^2}$$

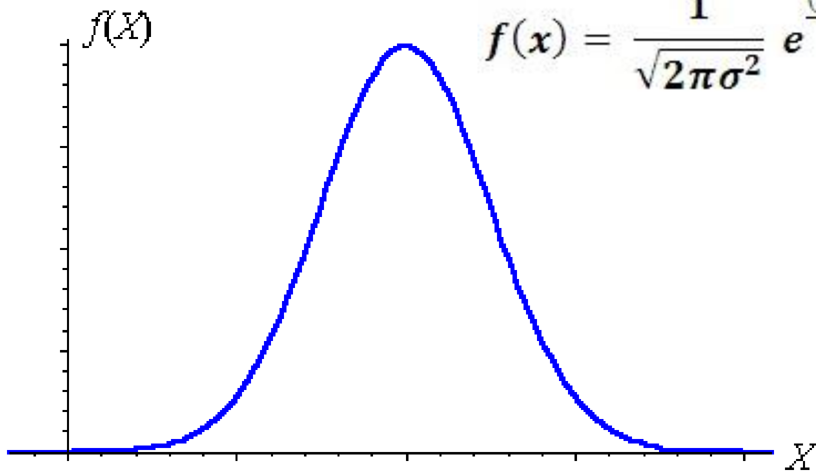
Τυπικό σφάλμα (standard error)

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Συντελεστής μεταβλητότητας ή συντελεστής διακύμανσης (coefficient of variation)

$$CV\ (\%) = \frac{s}{\bar{x}} 100$$

Κανονική κατανομή



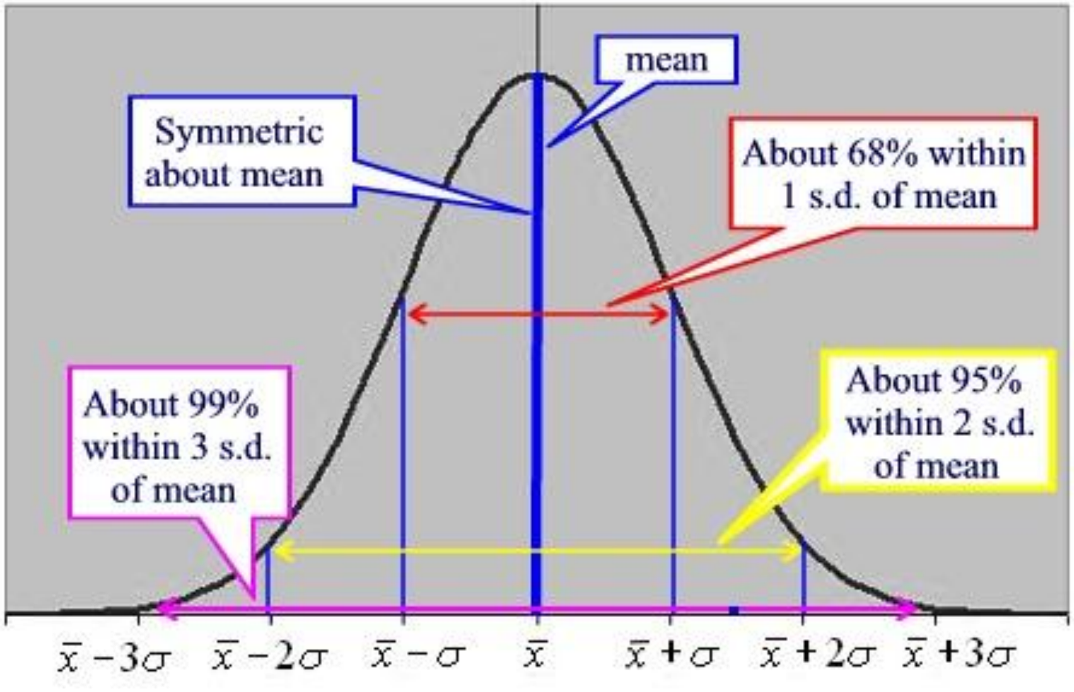
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{\frac{(x-\mu)^2}{2\sigma^2}}$$

mean = μ

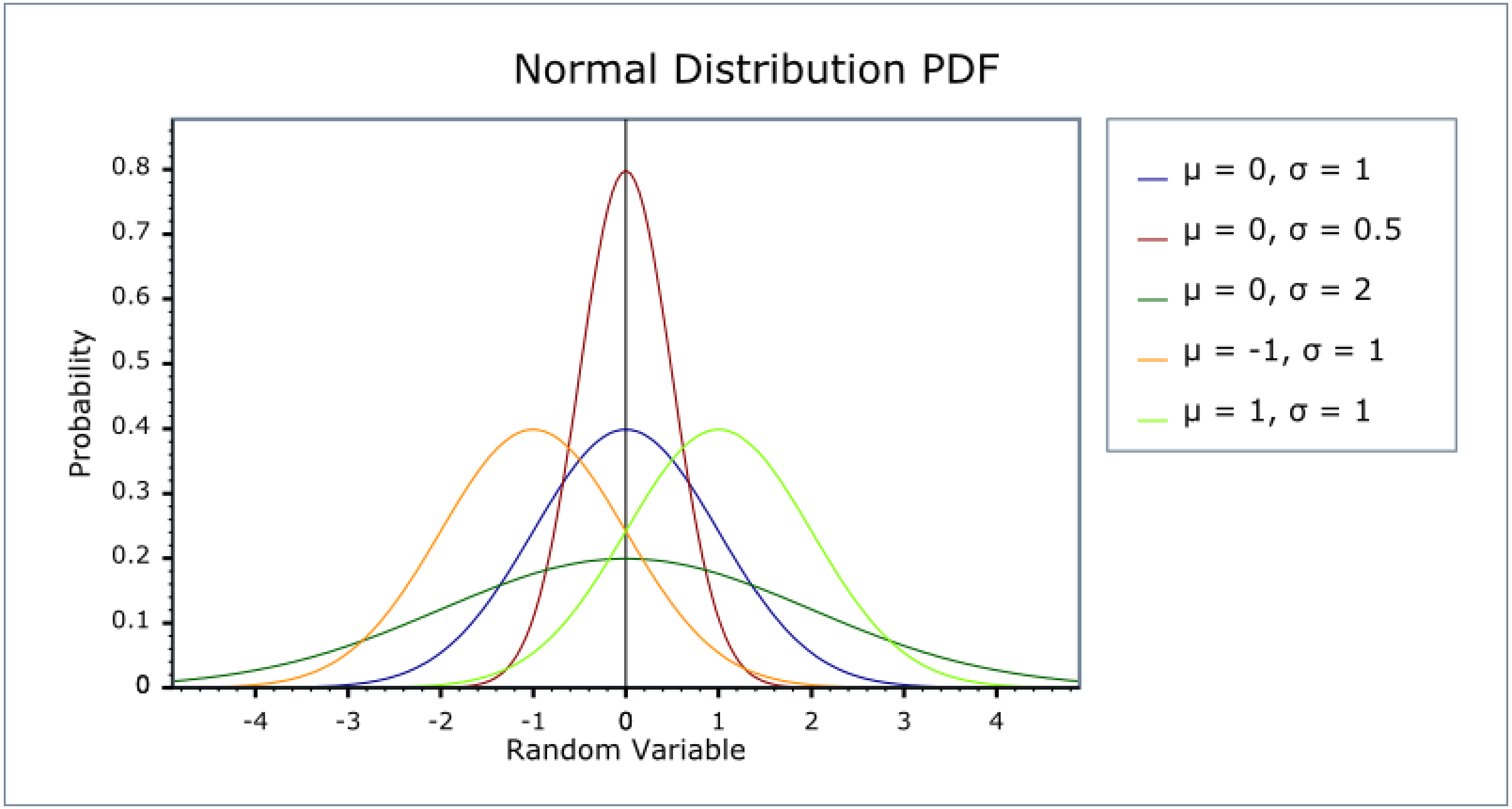
$$\underbrace{X\sim N}_{\text{x has normal distribution}}(M, \sigma^2)$$

variance = σ^2

Κανονική κατανομή



Κανονική κατανομή



Γιατί η κανονική κατανομή?

Κεντρικό Οριακό Θεώρημα (**Central Limit Theorem**)

Το άθροισμα ενός μεγάλου αριθμού ανεξάρτητων τυχαίων μεταβλητών ακολουθεί μία κατανομή η οποία προσεγγίζει την κανονική κατανομή **ανεξαρτήτως από την κατανομή που ακολουθούν οι τυχαίες μεταβλητές**!!!

Αν από έναν πληθυσμό που ακολουθεί οποιαδήποτε κατανομή με μέση τιμή μ και διασπορά $σ^2$, επιλέξουμε τυχαία δείγματα μεγέθους n και υπολογίσουμε τους μέσους τους, τότε, για μεγάλα n (θεωρητικά $n \to \infty$) η κατανομή αυτών των μέσων (των δειγματικών) είναι κατά προσέγγιση κανονική κατανομή με μέση τιμή επίσης μ και διασπορά $σ^2/n$.