

Περιεχόμενα

1	Πιθανότητες	1
1.1	Πειράματα τύχης, δειγματικός χώρος και ενδεχόμενα	1
1.2	Πιθανότητες: Ορισμοί και εφαρμογές	2
1.3	Πιθανότητες και πράξεις με ενδεχόμενα	2
1.4	Συνδυαστική & Πιθανότητες	3
2	Στατιστική	5
2.1	Πληθυσμός - Δείγμα – Μεταβλητές	5
2.2	Παρουσίαση στατιστικών δεδομένων	6
2.3	Μέτρα θέσης και μεταβλητότητας, θηκόγραμμα, συντελεστής μεταβλητότητας	7
2.4	Κανονική κατανομή και εφαρμογές	8

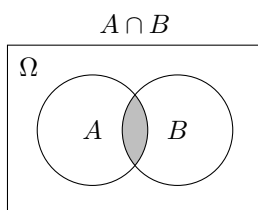
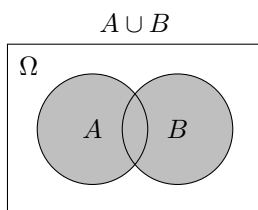
1 Πιθανότητες

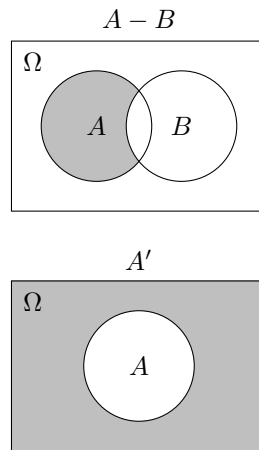
1.1 Πειράματα τύχης, δειγματικός χώρος και ενδεχόμενα

Δειγματικός χώρος. Αν $\omega_1, \omega_2, \dots, \omega_\kappa$ είναι τα δυνατά αποτελέσματα ενός πειράματος τύχης, τότε ο δειγματικός χώρος του πειράματος θα είναι το σύνολο:

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_\kappa\}.$$

Ενδεχόμενα. Οποιοδήποτε σύνολο δυνατών αποτελεσμάτων του πειράματος τύχης ονομάζεται ενδεχόμενο (event) ή γεγονός. Αν το αποτέλεσμα ενός πειράματος τύχης είναι το ω_ν , τότε όλα τα ενδεχόμενα που περιέχουν το ω_ν , λέμε ότι *πραγματοποιούνται*, ενώ τα ενδεχόμενα που δεν περιέχουν το ω_ν , λέμε ότι δεν πραγματοποιούνται.





1.2 Πιθανότητες: Ορισμοί και εφαρμογές

Κλασσικός ορισμός πιθανότητας. Σε ένα πείραμα τύχης με ν ισοπίθανα αποτελέσματα, η πιθανότητα ενός ενδεχομένου A που περιέχει κ τέτοια αποτελέσματα είναι

$$P(A) = \frac{\kappa}{\nu},$$

δηλαδή το πλήθος των ευνοϊκών αποτελεσμάτων για το A προς το πλήθος όλων των δυνατών αποτελεσμάτων. Άμεση συνέπεια του ορισμού είναι

$$0 \leq P(A) \leq 1 \quad \text{για κάθε ενδεχόμενο } A.$$

Αξιωματικός ορισμός πιθανότητας. Έστω $\Omega = \{\omega_1, \omega_2, \dots, \omega_\nu\}$ ένας δ.χ. ενός πειράματος τύχης. Σε κάθε ενδεχόμενο A του Ω αποδίδουμε έναν πραγματικό αριθμό που ονομάζουμε πιθανότητα του A και συμβολίζουμε με $P(A)$, έτσι ώστε:

- (α) $P(A) \geq 0$, για οποιοδήποτε A του Ω
- (β) $P(\Omega) = 1$
- (γ) $P(A \cup B) = P(A) + P(B)$ αν $A \cap B = \emptyset$.

Η τελευταία ιδιότητα λέγεται **απλός προσθετικός νόμος**. Δύο ενδεχόμενα A, B λέγονται **ασυμβίβαστα** αν $A \cap B = \emptyset$. Προσοχή! ο απλός προσθετικός νόμος ισχύει μόνο για ασυμβίβαστα ενδεχόμενα.

1.3 Πιθανότητες και πράξεις με ενδεχόμενα

$$(Π1) \quad P(A') = 1 - P(A)$$

$$(Π2) \quad P(A) = P(A \cap B) + P(A - B)$$

(Π3) Αν $B \subset A$, τότε $P(B) \leq P(A)$

(Π4) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Άμεση συνέπεια των παραπάνω είναι

$$0 \leq P(A) \leq 1 \quad \text{για κάθε ενδεχόμενο } A. \quad (\text{Π5})$$

Πιθανότητες σύνθετων ενδεχομένων. Έστω A, B ενδεχόμενα ενός δ.χ. Ω .

- $P(A)$ είναι η πιθανότητα να πραγματοποιηθεί το A .
- $P(A')$ είναι η πιθανότητα να **μην** πραγματοποιηθεί το A .
- $P(A \cup B)$ είναι η πιθανότητα να πραγματοποιηθεί το A **ή** το B .
- $P(A \cup B)$ είναι η πιθανότητα να πραγματοποιηθεί **τουλάχιστον ένα** από τα A και B .
- $P(A \cap B)$ είναι η πιθανότητα να πραγματοποιηθεί το A **και** το B .
- $P((A \cup B)')$ είναι η πιθανότητα να μην πραγματοποιηθεί **κανένα** από τα A και B .
- $P(A - B)$ είναι η πιθανότητα να πραγματοποιηθεί το A **αλλά όχι το** B .
- $P((A - B) \cup (B - A))$ είναι η πιθανότητα να πραγματοποιηθεί **μόνο ένα** από τα A και B .

1.4 Συνδυαστική & Πιθανότητες

Βασική αρχή της απαρίθμησης. Έστω ότι μια διαδικασία μπορεί να πραγματοποιηθεί σε ν διαδοχικές φάσεις $\phi_1, \phi_2, \dots, \phi_\nu$. Αν κάθε φάση ϕ_ρ μπορεί να πραγματοποιηθεί με κ_ρ τρόπους, τότε η διαδικασία αυτή μπορεί να πραγματοποιηθεί με $\kappa_1 \cdot \kappa_2 \cdots \kappa_\nu$ τρόπους.

1. Διάταξη με επανάληψη. Διάταξη ν στοιχείων ενός συνόλου A ανά κ λέγεται καθέννας από τους διαφορετικούς τρόπους με τους οποίους μπορούμε να τοποθετήσουμε στοιχεία του A σε μια σειρά κ θέσεων επιτρέποντας επαναλήψεις.

- Το πλήθος των **διατάξεων των ν ανά κ** είναι ίσο με ν^κ .

Παράδειγμα. 5 ρίψεις ενός κέρματος. Κάθε διατεταγμένη πεντάδα (π.χ., $\{K, \Gamma, K, K, \Gamma\}$) ονομάζεται *διάταξη των 2 ανά 5*. Το πλήθος των διατάξεων των 2 ανά 5 είναι 2^5 .

2. Διάταξη χωρίς επαναλήψεις. Διάταξη ν στοιχείων ενός συνόλου A ανά κ χωρίς επαναλήψεις, με $\kappa \leq \nu$, λέγεται καθένας από τους διαφορετικούς τρόπους με τους οποίους μπορούμε να πάρουμε στοιχεία του A και να τα βάλουμε στην σειρά.

- Το πλήθος των διατάξεων των ν ανά κ χωρίς επαναλήψεις, με $\kappa \leq \nu$, συμβολίζεται με $(\nu)_\kappa$ και είναι ίσο με

$$(\nu)_\kappa = \nu \cdot (\nu - 1) \cdots (\nu - \kappa + 1) = \frac{\nu!}{(\nu - \kappa)!}$$

Παράδειγμα. Το πενταμελές ενός τμήματος συνεδριάζει για να εκλέξει πρόεδρο, γραμματέα και ταμία. Με πόσους διαφορετικούς τρόπους μπορεί να γίνει αυτή η εκλογή; Το πλήθος των δυνατών τριάδων είναι $5 \cdot 4 \cdot 3 = 60$. Κάθε τέτοια τριάδα ονομάζεται *διάταξη των 5 ανά 3 χωρίς επαναλήψεις*.

3. Μεταθέσεις (ειδική περίπτωση διάταξη χωρίς επαναλήψεις). Το πλήθος των μεταθέσεων ν στοιχείων δίνεται από

$$\nu! = 1 \cdot 2 \cdots (\nu - 1) \cdot \nu.$$

Παράδειγμα. Το πλήθος των διαφορετικών τρόπων που μπορούν να καθίσουν 5 άτομα σε 5 καρέκλες είναι $5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$.

4. Συνδυασμοί. Συνδυασμός των ν στοιχείων ενός συνόλου A ανά κ λέγεται κάθε υποσύνολο του A με κ στοιχεία.

Το πλήθος των δυνατών τρόπων να επιλεγούν κ από τα ν στοιχεία ενός συνόλου:

$$\binom{\nu}{\kappa} = \frac{\nu!}{\kappa!(\nu - \kappa)!}$$

Παράδειγμα. Το πενταμελές ενός τμήματος συνεδριάζει για να εκλέξει τριμελή επιτροπή. Το πλήθος των δυνατών τρόπων να επιλεγούν 3 από τους 5 είναι

$$\binom{5}{3} = \frac{5!}{3!(5 - 3)!} = \frac{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5}{1 \cdot 2 \cdot 3 \cdot (1 \cdot 2)} = 10.$$

2 Στατιστική

2.1 Πληθυσμός - Δείγμα – Μεταβλητές

Η στατιστική είναι ο κλάδος των μαθηματικών ο οποίος βασίζεται σε ένα σύνολο αρχών και μεθοδολογιών για

- τον σχεδιασμό της διαδικασίας συλλογής δεδομένων,
- την συνοπτική και αποτελεσματική παρουσία τους,
- την ανάλυση και εξαγωγή αντίστοιχων συμπερασμάτων.

Έστω ότι θέλουμε να εξετάσουμε τα στοιχεία ενός συνόλου ως προς ένα ή περισσότερα χαρακτηριστικά του, π.χ.,

- (α) Τις προτιμήσεις των ψηφοφόρων εν όψει των προσεχών εκλογών.
- (β) Την μόλυνση των λαχανικών από τα φυτοφάρμακα.
- (γ) Από πόσα μέλη – κατά μέσο όρο – αποτελείτε η Ελληνική οικογένεια.

Σε καθένα από τα παραπάνω παραδείγματα το σύνολο υπό εξέταση ονομάζεται **πληθυσμός**. Π.χ., ο πληθυσμός από τα παραπάνω παραδείγματα είναι

- (α) Οι ψηφοφόροι.
- (β) Τα λαχανικά.
- (γ) Οι Ελληνικές οικογένειες.

Τα χαρακτηριστικά ως προς τα οποία εξετάζουμε έναν πληθυσμό λέγονται **μεταβλητές** και τις συμβολίζουμε συνήθως με τα κεφαλαία γράμματα $X, Y, Z, \dots$,

- (α) X ψήφος
- (β) Y ποσότητα φωσφόρου σε mg/lt .
- (γ) Z ο αριθμός των μελών μιας οικογένειας.

Οι δυνατές τιμές που μπορεί να πάρει μια μεταβλητή λέγονται **τιμές της μεταβλητής**. Οι μεταβλητές χωρίζονται σε κατηγορίες και υποκατηγορίες ανάλογα με τις τιμές που παίρνουν

Ποιοτικές ή κατηγορικές μεταβλητές. Είναι οι μεταβλητές των οποίων οι τιμές δεν είναι αριθμοί, π.χ., η ψήφος X είναι μια μεταβλητή που παίρνει τις τιμές Ν.Δ., ΠΑΣΟΚ, ΣΥΡΙΖΑ,...

Ποσοτικές. Είναι οι μεταβλητές των οποίων οι τιμές είναι αριθμοί και διακρίνονται σε δύο υποκατηγορίες:

- (i) **Διακριτές** μεταβλητές, που παίρνουν μόνο ακέραιες τιμές, π.χ., ο αριθμός των μελών μιας οικογένειας παίρνει τις τιμές $1, 2, \dots$.
- (ii) **Συνεχείς** μεταβλητές, που μπορούν να πάρουν οποιαδήποτε τιμή ενός διαστήματος πραγματικών αριθμών (α, β) , π.χ., η ποσότητα φωσφόρου Y σε ένα λάχανο είναι $0.02mg/lt$.

Δείγμα ονομάζεται κάθε υποσύνολο του πληθυσμού. *Μέγεθος του δείγματος* ονομάζεται το πλήθος των στοιχείων του δείγματος.

2.2 Παρουσίαση στατιστικών δεδομένων

Συμβολίζουμε με $x_1, x_2, \dots, x_k$ τις τιμές μιας μεταβλητής X .

Αν το μέγεθος του δείγματος είναι ν , τότε $k \leq \nu$. Δηλαδή, ο αριθμός που αντιστοιχεί στο πλήθος των τιμών που παίρνει μια μεταβλητή X δεν μπορεί να είναι μεγαλύτερος από τον αριθμό του μεγέθους του δείγματος.

Η συχνότητα ν_i είναι ο φυσικός αριθμός που δείχνει πόσες φορές εμφανίζεται η τιμή x_i της εξεταζόμενης μεταβλητής X στο σύνολο των παρατηρήσεων. Είναι φανερό ότι το άθροισμα όλων των συχνοτήτων είναι ίσο με το μέγεθος ν του δείγματος, δηλαδή

$$\nu_1 + \nu_2 + \dots + \nu_k = \nu. \quad (1)$$

Αν διαιρέσουμε τη συχνότητα ν_i με το μέγεθος ν του δείγματος, προκύπτει η **σχετική συχνότητα** f_i της τιμής x_i , δηλαδή

$$f_i = \frac{\nu_i}{\nu}, \quad i = 1, 2, \dots, k. \quad (2)$$

Για τη σχετική συχνότητα ισχύουν οι ιδιότητες:

(α) $0 \leq f_i \leq 1, i = 1, 2, \dots, k$, αφού $0 \leq \nu_i \leq \nu$.

(β) $f_1 + f_2 + \dots + f_k = 1$, αφού

$$f_1 + f_2 + \dots + f_k = \frac{\nu_1}{\nu} + \frac{\nu_2}{\nu} + \dots + \frac{\nu_k}{\nu} = \frac{\nu_1 + \nu_2 + \dots + \nu_k}{\nu} = \frac{\nu}{\nu} = 1.$$

Συνήθως, τις σχετικές συχνότητες f_i τις εκφράζουμε επί τοις εκατό, δηλαδή

$$f_i \% = 100f_i.$$

Γραφική παράσταση κατανομής συχνοτήτων

Το κυκλικό διάγραμμα είναι ένας κυκλικός δίσκος χωρισμένος σε κυκλικούς τομείς, τα εμβαδά ή ισοδύναμα, τα τόξα των οποίων είναι ανάλογα προς τις αντίστοιχες συχνότητες ν_i ή τις σχετικές συχνότητες f_i των τιμών x_i της μεταβλητής. Αν συμβολίσουμε με α_i το αντίστοιχο τόξο ενός κυκλικού τμήματος στο κυκλικό διάγραμμα συχνοτήτων, τότε

$$\alpha_i = \nu_i \frac{360^\circ}{\nu} = 360^\circ f_i. \quad (3)$$

Ομαδοποίηση των παρατηρήσεων

Οι πίνακες συχνοτήτων και κατ' αναλογία τα αντίστοιχα διαγράμματα είναι δύσκολο να κατασκευαστούν, όταν το πλήθος των τιμών μιας μεταβλητής είναι αρκετά μεγάλο. Σ' αυτές τις περιπτώσεις είναι απαραίτητο να ταξινομηθούν (ομαδοποιηθούν) τα δεδομένα σε μικρό πλήθος ομάδων, που ονομάζονται και κλάσεις, έτσι ώστε κάθε τιμή να ανήκει μόνο σε μία κλάση.

- Στην ουσία **κλάσεις** είναι διαστήματα της μορφής $[a, b)$. Τα άκρα των κλάσεων a, b καλούνται **όρια των κλάσεων**.
- Στην ομαδοποίηση δεδομένων, ο αριθμός κ των κλάσεων θα δίνετε. Ο προσδιορισμός του **πλάτους των κλάσεων** c , γίνεται ως εξής: βρίσκουμε το **εύρος** R του δείγματος

$$R = \text{Μεγαλύτερη παρατήρηση} - \text{Μικρότερη παρατήρηση},$$

και μετά υπολογίζουμε το πλάτος c των κλάσεων διαιρώντας το εύρος R διά του αριθμού των κλάσεων κ , δηλαδή

$$c = \frac{R}{\kappa},$$

στρογγυλεύοντας, αν χρειαστεί για λόγους διευκόλυνσης, πάντα προς τα πάνω.

2.3 Μέτρα θέσης και μεταβλητότητας, θηκόγραμμα, συντελεστής μεταβλητότητας

Μέτρα θέσης

Μέση τιμή είναι το άθροισμα των παρατηρήσεων διά του πλήθους των παρατηρήσεων

$$\bar{x} = \frac{t_1 + t_2 + \dots + t_\nu}{\nu}. \quad (4)$$

Σε ένα πρόβλημα εάν δίνονται οι τιμές x_i της μεταβλητής X και οι αντίστοιχες συχνότητες τους ν_i , τότε για να βρω τη μέση τιμή μπορώ να χρησιμοποιήσω τον παρακάτω ισοδύναμο τύπο

$$\bar{x} = \frac{x_1\nu_1 + x_2\nu_2 + \dots + x_\kappa\nu_\kappa}{\nu}. \quad (5)$$

Διάμεσος δ (ή δεύτερο τεταρτημόριο Q_2) ενός δείγματος ν παρατηρήσεων οι οποίες έχουν διαταχθεί σε αύξουσα σειρά ορίζεται ως η μεσαία παρατήρηση, όταν το ν είναι περιττός αριθμός, ή ο μέσος όρος (ημιάθροισμα) των δύο μεσαίων παρατηρήσεων όταν το ν είναι άρτιος αριθμός.

$$\delta = \begin{cases} t_{\frac{\nu+1}{2}} & \text{όταν } \nu \text{ περιττός,} \\ \frac{1}{2}(t_{\frac{\nu}{2}} + t_{\frac{\nu}{2}+1}) & \text{όταν } \nu \text{ άρτιος.} \end{cases} \quad (6)$$

Το πρώτο (Q_1) και τρίτο (Q_3) τεταρτημόριο είναι η διάμεσος του πρώτου και του δεύτερου μισού των διατεταγμένων παρατηρήσεων, αντίστοιχα.

Επικρατούσα τιμή M_0 είναι η παρατήρηση με τη μεγαλύτερη συχνότητα.

Μέτρα διασποράς

Εύρος R ορίζεται η διαφορά της ελάχιστης παρατήρησης από τη μέγιστη παρατήρηση, δηλαδή

$$R = \text{Μεγαλύτερη παρατήρηση} - \text{Μικρότερη παρατήρηση}. \quad (7)$$

Το ενδοτεταρτημοριακό εύρος Q είναι η διαφορά τρίτου τεταρτημόριου Q_3 από το πρώτο τεταρτημόριο Q_1 . Δηλαδή:

$$Q = Q_3 - Q_1.$$

Όσες παρατηρήσεις βρίσκονται έξω από το διάστημα $[Q_1 - 1,5 \cdot Q, Q_3 + 1,5 \cdot Q]$ ονομάζονται **ακραίες τιμές**. Η **διάμεσος** είναι ένα μέτρο θέσης το οποίο δεν επηρεάζεται από το μέγεθος των ακραίων τιμών.

Διακύμανση ή διασπορά s^2 ονομάζεται ο μέσος όρος των τετραγώνων της διαφοράς των παρατηρήσεων από τη μέση τιμή:

$$s^2 = \frac{(t_1 - \bar{x})^2 + (t_2 - \bar{x})^2 + \dots + (t_\nu - \bar{x})^2}{\nu} \quad (8)$$

$$= \frac{(x_1 - \bar{x})^2 \cdot \nu_1 + (x_2 - \bar{x})^2 \cdot \nu_2 + \dots + (x_\kappa - \bar{x})^2 \cdot \nu_\kappa}{\nu}. \quad (9)$$

Η **τυπική απόκλιση** s δίνεται από $s = \sqrt{s^2}$.

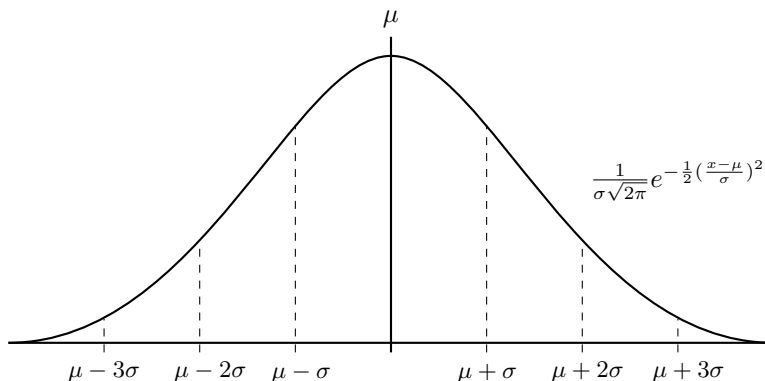
Συντελεστής μεταβολής ή συντελεστής μεταβλητότητας (CV)

$$CV = \frac{s}{|\bar{x}|} 100\%. \quad (10)$$

Ένα δείγμα τιμών μιας μεταβλητής είναι **ομοιογενές**, εάν ο συντελεστής μεταβολής δεν ξεπερνά το 10%.

2.4 Κανονική κατανομή και εφαρμογές

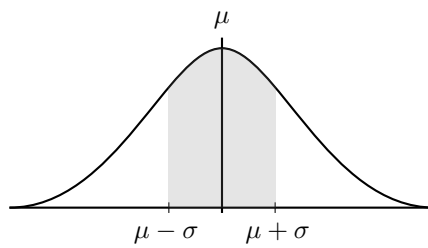
Αν η καμπύλη συχνοτήτων για το χαρακτηριστικό που εξετάζουμε είναι κανονική ή περίπου κανονική, δηλαδή έχει την ακόλουθη μορφή



τότε μπορούμε να συμπεράνουμε τα ακόλουθα, όπου μ η μέση τιμή και σ η τυπική απόκλιση:

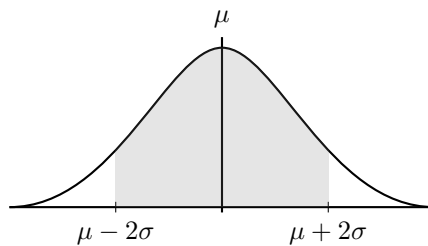
Το 68% περίπου των παρατηρήσεων βρίσκεται στο διάστημα

$$(\mu - \sigma, \mu + \sigma)$$



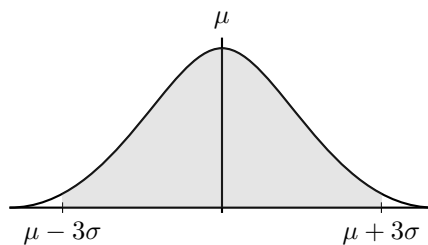
Το 95% περίπου των παρατηρήσεων βρίσκεται στο διάστημα

$$(\mu - 2\sigma, \mu + 2\sigma)$$



Το 99,7% περίπου των παρατηρήσεων βρίσκεται στο διάστημα

$$(\mu - 3\sigma, \mu + 3\sigma)$$



Το εύρος ισούται περίπου με έξι τυπικές αποκλίσεις, δηλαδή $R = 6\sigma$.