

Εισαγωγή στη Θεωρία Πιθανοτήτων και την Στατιστική

Ε01Π03

Βασικοί Όροι

Πιθανοκρατικές μέθοδοι δειγματοληψίας

- Η λύση για την αμερόληπτη επιλογή δείγματος είναι η εφαρμογή *τυχαίων διαδικασιών* , ώστε κάθε υποκείμενο του πληθυσμού να έχει *a priori* γνωστή και μη μηδενική πιθανότητα να συμπεριληφθεί στο δείγμα.
- Οι δειγματοληπτικές μέθοδοι που χρησιμοποιούν τυχαίους μηχανισμούς για την επιλογή ενός δείγματος ονομάζονται **πιθανολογικές ή πιθανοκρατικές**

Απλή τυχαία δειγματοληψία

- Το δείγμα επιλέγεται με τέτοιο τρόπο, ώστε κάθε μέλος του πληθυσμού να έχει την ίδια και ανεξάρτητη πιθανότητα να επιλεγεί.
- **Ανεξάρτητη πιθανότητα επιλογής** σημαίνει ότι η επιλογή ενός μέλους δεν επηρεάζει την πιθανότητα επιλογής κάποιου άλλου. Προϋποτίθεται ότι έχουμε έναν πλήρη κατάλογο με όλα τα μέλη του πληθυσμού.
- Η επιλογή γίνεται συνήθως με **πίνακες τυχαίων αριθμών**, ή στον υπολογιστή με κατάλληλους αλγόριθμους τυχαίων αριθμών ή με λοταρία.

Οι πίνακες τυχαίων αριθμών κατασκευάζονται έτσι, ώστε κάθε δεκαδικό ψηφίο (0 έως 9) να έχει *ίδια και ανεξάρτητη* πιθανότητα εμφάνισης .

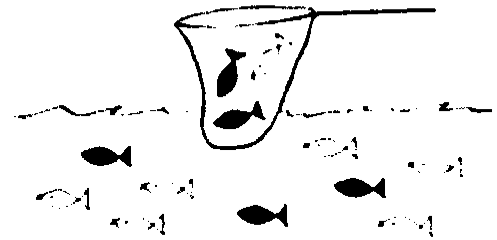
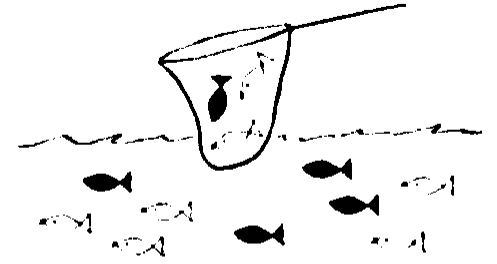
Δηλαδή, σε οποιαδήποτε θέση του πίνακα το κάθε δεκαδικό ψηφίο έχει πιθανότητα εμφάνισης ίση με $1/10$, και η εμφάνιση κάποιου ψηφίου δεν επηρεάζει την πιθανότητα εμφάνισης του επόμενου.

Πίνακας τυχαίων αριθμών

	1	2	3	4	5	6	7	8	9	10
1	57245	39666	18545	50534	57654	25519	35477	71309	12212	98911
2	42726	58321	59267	72742	53968	63679	54095	56563	09820	86291
3	82768	32694	62828	19097	09877	32093	23518	08654	64815	19894
4	97742	58918	33317	34192	06286	39824	74264	01941	95810	26247
5	48332	38634	20510	09198	56256	04431	22753	20944	95319	29515
6	26700	40484	28341	25428	08806	98858	04816	16317	94928	05512
7	66156	16407	57395	86230	47495	13908	97015	58225	82255	01956
8	64062	10061	01923	29260	32771	71002	58132	58646	69089	63694
9	24713	95591	26970	37647	26282	89759	69034	55281	64853	50837
10	90417	18344	22436	77006	87841	94322	45526	38145	86554	42733
11	78886	86557	11295	07253	29289	44814	58898	36929	66839	81250
12	39681	54696	38482	48217	73598	93649	92705	34912	18981	74299
13	38265	45196	31143	82190	27279	79883	20219	38823	84543	22119
14	34270	41885	00079	63600	59152	10670	27951	77830	05368	58315
15	73869	34748	75787	88844	89522	71436	04166	06246	20952	56808
16	21732	36017	69149	70330	90.500	73110	92908	55789	73450	68282
17	72583	49811	67519	98476	97889	37112	94963	91140	24571	23446
18	72.678	49483	57.039	18420	74773	16869	72077	27720	14058	66743
19	88572	01294	14117	56884	77107	53023	02243	26415	52233	12818
20	82868	59988	42323	96542	96733	00056	74887	21914	48300	96404

Απλή τυχαία δειγματοληψία

- Η τυχαία δειγματοληψία μπορεί να εξαλείψει τη μεροληψία που οφείλεται στη λάθος επιλογή δείγματος, αλλά δεν εξαλείφει τη μεταβλητότητα από δείγμα σε δείγμα.



1-ανά- n συστηματική δειγματοληψία

- Στη συστηματική τυχαία δειγματοληψία, επιλέγουμε πρώτα τυχαία ένα μέλος από τις πρώτες n μονάδες της λίστας των στοιχείων, όπου n είναι ο αριθμός που προκύπτει διαιρώντας το μέγεθος του πληθυσμού με το επιδιωκόμενο μέγεθος δείγματος.
- Στη συνέχεια, κάθε n -οστό μέλος, ξεκινώντας από το πρώτο επιλεγμένο μέλος, συμπεριλαμβάνεται στο δείγμα.

1-ανά- n συστηματική δειγματοληψία

- Για παράδειγμα, αν θέλουμε ένα συστηματικό **δείγμα 5 ατόμων** από έναν **πληθυσμό 100 ατόμων**, επιλέγουμε τυχαία έναν αριθμό από το 1 ως το 20 ($100:5$), έστω το 11, και από τη διαθέσιμη λίστα επιλέγουμε τα υποκείμενα με αριθμό 11, 31, 51, 71 και 91

1-ανά-ν συστηματική δειγματοληψία

Αν χρειαστεί να επιλέξουμε 150 νοικοκυριά από μια λίστα 45.000...

1. Θα διατάξουμε και τα 45.000 νοικοκυριά αλφαβητικά (ή με βάση κάποιο άλλο χαρακτηριστικό).
2. Αφού το μέγεθος του δείγματος θα πρέπει να ισούται με 150, η αναλογία πληθυσμού προς μέγεθος δείγματος είναι $45.000/150 = 300$.
3. Χρησιμοποιώντας αυτήν την αναλογία, επιλέγουμε τυχαία ένα νοικοκυριό από τα πρώτα 300 νοικοκυριά της λίστας χρησιμοποιώντας οποιαδήποτε μέθοδο. Έστω το 210^ο νοικοκυριό.
4. Στη συνέχεια επιλέγουμε το 210^ο νοικοκυριό από κάθε 300 νοικοκυριά στη λίστα.
5. Με άλλα λόγια, το δείγμα μας περιλαμβάνει τα νοικοκυριά με αριθμούς 210, 510, 810, 1110, 1410, 1710 κ.ο.κ.

Κατά στρώματα δειγματοληψία

- Πρόκειται για απλή τυχαία δειγματοληψία από ξεχωριστές υποομάδες του πληθυσμού.
- Σε ένα στρωματοποιημένο τυχαίο δείγμα, πρώτα διαιρούμε τον πληθυσμό σε υποπληθυσμούς, οι οποίοι ονομάζονται στρώματα.
- Στη συνέχεια, επιλέγεται ένα απλό τυχαίο δείγμα από καθένα από αυτά τα στρώματα.
- Η συλλογή όλων των δειγμάτων από όλα τα στρώματα δίνει το στρωματοποιημένο τυχαίο δείγμα.

Κατά στρώματα δειγματοληψία

- Κάθε φορά που παρατηρούμε ότι ένας πληθυσμός διαφέρει σε μεγάλο βαθμό ως προς την κατοχή ενός χαρακτηριστικού, μπορεί να προτιμήσουμε να τον χωρίσουμε σε διαφορετικά στρώματα και στη συνέχεια να επιλέξουμε ένα δείγμα από κάθε στρώμα.
- Μπορούμε να διαιρέσουμε τον πληθυσμό με βάση οποιοδήποτε χαρακτηριστικό, όπως το εισόδημα, τις δαπάνες, το φύλο, την εκπαίδευση, τη φυλή, την απασχόληση ή το μέγεθος της οικογένειας.
- Για παράδειγμα, αντί να πάρουμε ένα απλό τυχαίο δείγμα 100 ατόμων, παίρνουμε χωριστά ένα απλό τυχαίο δείγμα 50 ανδρών και ένα απλό τυχαίο δείγμα 50 γυναικών και τα ενώνουμε.

Κατά στρώματα δειγματοληψία – Αναλογικό δείγμα

- Αν διατηρήσουμε και στο δείγμα την αναλογία που έχουν τα στρώματα στον πληθυσμό, τότε το δείγμα λέγεται **αναλογικό** (proportional).
- Προϋπόθεση είναι να γνωρίζουμε την ποσοστιαία αναλογία κάθε κατηγορίας στον πληθυσμό, ώστε να απαιτήσουμε αυτή να διατηρηθεί στο δείγμα.

Κατά στρώματα δειγματοληψία – Αναλογικό δείγμα

- Σε ένα πανεπιστημιακό τμήμα το 35% των φοιτητών είναι στο πρώτο έτος, το 30% στο δεύτερο, το 25% στο τρίτο και το 10% στο τέταρτο. Για ένα αναλογικό δείγμα 60 φοιτητών θα επιλέξουμε τυχαία 21 φοιτητές (το 35% του 60) από το πρώτο έτος, 18 (το 30% του 60) από το δεύτερο, 15 (το 25% του 60) από το τρίτο και 6 (το 10% του 60) από το τέταρτο.

Κατά δεσμίδες δειγματοληψία

1. Στη δειγματοληψία κατά δεσμίδες, ολόκληρος ο πληθυσμός αρχικά χωρίζεται σε (γεωγραφικές) ομάδες που ονομάζονται δεσμίδες ή συστάδες.
 - Κάθε συστάδα πρέπει να είναι αντιπροσωπευτική του πληθυσμού.
2. Στη συνέχεια επιλέγεται ένα τυχαίο δείγμα συστάδων.
3. Τέλος, επιλέγεται ένα τυχαίο δείγμα στοιχείων από κάθε ένα από τα επιλεγμένα συμπλέγματα.

Κατά δεσμίδες δειγματοληψία

Ένας ερευνητής θέλει να μελετήσει τις συνήθειες κατανάλωσης φρούτων και λαχανικών των νοικοκυριών σε μια μεγάλη πόλη.

- Ο ερευνητής χωρίζει την πόλη σε γειτονιές (π.χ. κάθε γειτονιά αποτελεί μια συστάδα).
- Υποθέτοντας ότι υπάρχουν 20 γειτονιές στην πόλη, κάθε γειτονιά θα είναι μια δεσμίδα.
- Από αυτές τις 20 γειτονιές, ο ερευνητής επιλέγει τυχαία 5 γειτονιές (συστάδες) για να συμπεριληφθούν στη μελέτη.
- Στη συνέχεια, ο ερευνητής επιλέγει τυχαία 10 νοικοκυριά από κάθε μία από τις 5 επιλεγμένες γειτονιές.
- Έτσι, ο συνολικός αριθμός των νοικοκυριών που θα συμμετάσχουν στη μελέτη είναι 50 (5 γειτονιές $\times$ 10 νοικοκυριά).

Εφαρμογή 1

Μια σχολική επιτροπή θέλει να διεξάγει μια έρευνα για τις απόψεις των μαθητών σχετικά με την ποιότητα της σίτισης στο σχολείο. Υπάρχουν 300 μαθητές στο σχολείο.

Οι μαθητές αριθμούνται από 1 έως 300 (π.χ., 1, 2, 3, ..., 300).

Χρησιμοποιώντας μια γεννήτρια τυχαίων αριθμών ή άλλες μεθόδους τυχαίας επιλογής (όπως την κλήρωση από μια κούτα με αριθμούς), επιλέγουν 30 τυχαίους αριθμούς από το 1 έως το 300.

ΑΠΛΗ ΤΥΧΑΙΑ

Εφαρμογή 2

Μια ερευνητική ομάδα θέλει να μελετήσει τις συνήθειες ανάγνωσης των φοιτητών σε ένα πανεπιστήμιο που έχει 1,000 φοιτητές. Η ομάδα επιλέγει 100 φοιτητές για την έρευνα με τον ακόλουθο τρόπο.

1. Οι φοιτητές αριθμούνται από 1 έως 1,000.

2. Υπολογίζει τον βήμα (k) ως εξής: $k = \frac{\text{Συνολικός πληθυσμός}}{\text{Μέγεθος δείγματος}} = \frac{1000}{100} = 10$

3. Επιλέγουν έναν τυχαίο αριθμό μεταξύ 1 και 10 (π.χ., το 3).

4. Ξεκινώντας από τον αριθμό 3, οι ερευνητές επιλέγουν κάθε 10ο φοιτητή.

5. Έτσι, οι επιλεγμένοι φοιτητές θα είναι οι 3, 13, 23, 33, 43, ..., 993.

**1-ανά-10
ΣΥΣΤΗΜΑΤΙΚΗ**

Εφαρμογή 3

Ένα πανεπιστήμιο θέλει να διεξάγει μια έρευνα για την ικανοποίηση των φοιτητών από τις υπηρεσίες του. Στο πανεπιστήμιο υπάρχουν τρεις σχολές με τους εξής αριθμούς φοιτητών:

Σχολή Α: 200 φοιτητές

Σχολή Β: 300 φοιτητές

Σχολή Γ: 500 φοιτητές

Συνολικά, υπάρχουν 1,000 φοιτητές στο πανεπιστήμιο. Η ερευνητική ομάδα αποφασίζει να επιλέξει δείγμα 100 φοιτητών για την έρευνα.

**ΚΑΤΑ ΣΤΡΩΜΑΤΑ
ΑΝΑΛΟΓΙΚΗ**

Εφαρμογή 4

Μια καθηγήτρια στατιστικής ήθελε να μάθει τον μέσο όρο (ΜΟ) βαθμολογίας για όλους τους φοιτητές στο πανεπιστήμιό της. Χρησιμοποίησε όλους τους φοιτητές που ήταν εγγεγραμμένοι στην τάξη της στα στατιστικά ως δείγμα και συνέλεξε πληροφορίες για τον ΜΟ τους για να βρει τον ΜΟ βαθμολογίας για όλους τους φοιτητές.

- Τι είδους δειγματοληψία χρησιμοποίησε; (τυχαία, συστηματική, στρωματοποιημένη, κατά δεσμίδες, βολική, εθελοντική.....)
- Εντοπίζετε κάποιο συστηματικό σφάλμα; Και αν ναι, τι είδους;

Εφαρμογή 5

Μια καθηγήτρια στατιστικής ήθελε να μάθει τον μέσο όρο (ΜΟ) βαθμολογίας για όλους τους φοιτητές στο πανεπιστήμιό της. Πήρε μια λίστα με όλους τους φοιτητές που είναι εγγεγραμμένοι στο πανεπιστήμιο από τη γραμματεία και στη συνέχεια επιλέγει 150 φοιτητές τυχαία από αυτήν τη λίστα χρησιμοποιώντας ένα στατιστικό πακέτο λογισμικού.

- Τι είδους δείγμα χρησιμοποίησε; (τυχαίο, συστηματικό, στρωματοποιημένο, κατά δεσμίδες, βολικό, εθελοντικό.....)
- Εντοπίζεται κάποιο συστηματικό σφάλμα; Και αν ναι, τι είδους;

Εφαρμογή 6

Στον πίνακα που ακολουθεί φαίνεται ο φοιτητικός πληθυσμός ενός πανεπιστημίου ανά πρόγραμμα σπουδών και ανά φύλο: 2.100 προπτυχιακοί φοιτητές (1.100 κορίτσια και 1.000 αγόρια) και 700 μεταπτυχιακοί (300 κορίτσια και 400 αγόρια). Θέλουμε ένα απλό τυχαίο δείγμα 10 φοιτητών από το πανεπιστήμιο αυτό.

Πίνακας 1.1. Αριθμός φοιτητών ανά πρόγραμμα σπουδών και ανά φύλο.

Πρόγραμμα σπουδών	Κορίτσια	Αγόρια	Σύνολο
Προπτυχιακό	1.100	1.000	2.100
Μεταπτυχιακό	300	400	700
Σύνολο	1.400	1.400	2.800

Εφαρμογή 6

Βήμα 1

Αριθμούμε

- από το 1 ως το 1.100 τα κορίτσια που είναι στο προπτυχιακό πρόγραμμα,
- από το 1.101 ως το 2.100 τα αγόρια του προπτυχιακού προγράμματος,
- από το 2.101 ως το 2.400 τα κορίτσια του μεταπτυχιακού προγράμματος
- από το 2.401 ως το 2.800 τα αγόρια του μεταπτυχιακού προγράμματος.

Εφαρμογή 6

Βήμα 2

Αρχίζοντας από μια τυχαία γραμμή και στήλη του πίνακα διαβάζουμε τετράδες ψηφίων (γιατί ο πληθυσμός μας εδώ έχει τετραψήφιο αριθμό μελών) και συνεχίζουμε προς τα κάτω (ή οριζόντια ή διαγώνια) αγνοώντας αριθμούς μεγαλύτερους από το 2.800.

Αν εξαντλήσουμε μια στήλη πριν συμπληρωθεί το απαιτούμενο μέγεθος του δείγματος, συνεχίζουμε με την επόμενη.

Εφαρμογή 6

Πίνακας τυχαίων αριθμών

	1	2	3	4	5	6	7	8	9	10
1	57245	39666	18545	50534	57654	25519	35477	71309	12212	98911
2	42726	58321	59267	72742	53968	63679	54095	56563	09820	86291
3	82768	32694	62828	19097	09877	32093	23518	08654	64815	19894
4	97742	58918	33317	34192	06286	39824	74264	01941	95810	26247
5	48332	38634	20510	09198	56256	04431	22753	20944	95319	29515
6	26700	40484	28341	25428	08806	98858	04816	16317	94928	05512
7	66156	16407	57395	86230	47495	13908	97015	58225	82255	01956
8	64062	10061	01923	29260	32771	71002	58132	58646	69089	63694
9	24713	95591	26970	37647	26282	89759	69034	55281	64853	50837
10	90417	18344	22436	77006	87841	94322	45526	38145	86554	42733
11	78886	86557	11295	07253	29289	44814	58898	36929	66839	81250
12	39681	54696	38482	48217	73598	93649	92705	34912	18981	74299
13	38265	45196	31143	82190	27279	79883	20219	38823	84543	22119
14	34270	41885	00079	63600	59152	10670	27951	77830	05368	58315
15	73869	34748	75787	88844	89522	71436	04166	06246	20952	56808
16	21732	36017	69149	70330	90.500	73110	92908	55789	73450	68282
17	72583	49811	67519	98476	97889	37112	94963	91140	24571	23446
18	72.678	49483	57.039	18420	74773	16869	72077	27720	14058	66743
19	88572	01294	14117	56884	77107	53023	02243	26415	52233	12818
20	82868	59988	42323	96542	96733	00056	74887	21914	48300	96404

Εφαρμογή 7

Βήμα 1

Διαιρούμε τον αριθμό των μελών του πληθυσμού με το μέγεθος του δείγματος, δηλαδή $2.800:10=280$.

Άρα, εδώ θα έχουμε 1-ανά-280 συστηματική δειγματοληψία.

Βήμα 2

Επιλέγουμε από τον πίνακα τυχαίων αριθμών έναν αριθμό **x** από το 1 ως το 280.

Βήμα 3

Επιλέγουμε τον $x+280$, τον $x+560$, τον $x+840$ κ.ο.κ

Εφαρμογή 7

Οι πίνακες τυχαίων αριθμών κατασκευάζονται έτσι, ώστε κάθε δεκαδικό ψηφίο (0 έως 9) να έχει *ίδια και ανεξάρτητη* πιθανότητα εμφάνισης.

Δηλαδή, σε οποιαδήποτε θέση του πίνακα το κάθε δεκαδικό ψηφίο έχει πιθανότητα εμφάνισης ίση με $1/10$, και η εμφάνιση κάποιου ψηφίου δεν επηρεάζει την πιθανότητα εμφάνισης του επόμενου.

Πίνακας τυχαίων αριθμών

	1	2	3	4	5	6	7	8	9	10
1	57245	39666	18545	50534	57654	25519	35477	71309	12212	98911
2	42726	58321	59267	72742	53968	63679	54095	56563	09820	86291
3	82768	32694	62828	19097	09877	32093	23518	08654	64815	19894
4	97742	58918	33317	34192	06286	39824	74264	01941	95810	26247
5	48332	38634	20510	09198	56256	04431	22753	20944	95319	29515
6	26700	40484	28341	25428	08806	98858	04816	16317	94928	05512
7	66156	16407	57395	86230	47495	13908	97015	58225	82255	01956
8	64062	10061	01923	29260	32771	71002	58132	58646	69089	63694
9	24713	95591	26970	37647	26282	89759	69034	55281	64853	50837
10	90417	18344	22436	77006	87841	94322	45526	38145	86554	42733
11	78886	86557	11295	07253	29289	44814	58898	36929	66839	81250
12	39681	54696	38482	48217	73598	93649	92705	34912	18981	74299
13	38265	45196	31143	82190	27279	79883	20219	38823	84543	22119
14	34270	41885	00079	63600	59152	10670	27951	77830	05368	58315
15	73869	34748	75787	88844	89522	71436	04166	06246	20952	56808
16	21732	36017	69149	70330	90.500	73110	92908	55789	73450	68282
17	72583	49811	67519	98476	97889	37112	94963	91140	24571	23446
18	72.678	49483	57.039	18420	74773	16869	72077	27720	14058	66743
19	88572	01294	14117	56884	77107	53023	02243	26415	52233	12818
20	82868	59988	42323	96542	96733	00056	74887	21914	48300	96404

Εφαρμογή 8

Σε ένα νηπιαγωγείο υπάρχουν 12 παιδάκια, τα οποία έχουμε χωρίσει σε στρώματα (ανάλογα με το φύλο τους) και σε δεσμίδες (ανάλογα με το πού κάθονται), όπως δείχνει ο πίνακας.

Με ποια ή ποιες από τις αναγραφόμενες δειγματοληπτικές μεθόδους είναι δυνατό να έχουν προέλθει τα παρακάτω δείγματα 4 νηπίων;

		Στρώμα 1: αγόρια		Στρώμα 2: κορίτσια	
Δεσμίδες	3η γραμμή	Γιάννης (9)	Αντώνης (10)	Λένα (11)	Φιλία (12)
	2η γραμμή	Μανόλης (8)	Γιώργος (7)	Ζωή (6)	Μαρία (5)
	1η γραμμή	Κώστας (1)	Νίκος (2)	Γιάννα (3)	Άννα (4)

Εφαρμογή 8

		Στρώμα 1: αγόρια		Στρώμα 2: κορίτσια	
Δεσμίδες	3η γραμμή	Γιάννης (9)	Αντώνης (10)	Λένα (11)	Φιλία (12)
	2η γραμμή	Μανόλης (8)	Γιώργος (7)	Ζωή (6)	Μαρία (5)
	1η γραμμή	Κώστας (1)	Νίκος (2)	Γιάννα (3)	Άννα (4)

Δείγμα	Απλή τυχαία	Κατά στρώματα	Κατά δεσμίδες
Γιώργος, Μαρία, Μανόλης, Ζωή			
Γιώργος, Μανόλης, Αντώνης, Νίκος			
Μανόλης, Νίκος, Φιλία, Λένα			

- Αξιοποιώντας τον πίνακα τυχαίων αριθμών να πάρετε 1-ανά-3 συστηματικό δείγμα από τα παιδιά αυτά.