Δειγματική Κατανομή
Κεφάλαιο 9

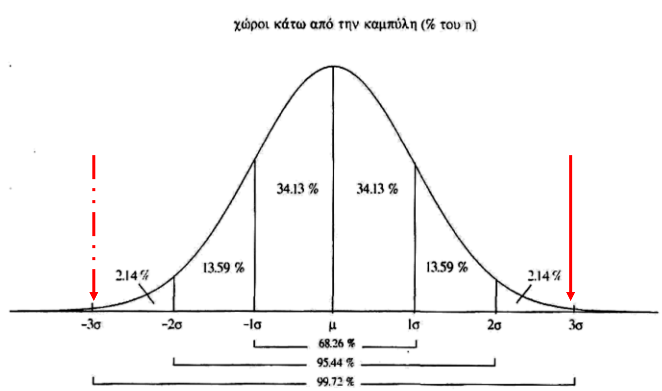
Αντώνης Κ. Τραυλός (B.A., M.A., Ph.D.)
Καθηγητής
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΛΟΠΟΝΝΗΣΟΥ
Σχολή Επιστημών Ανθρώπινης Κίνησης και Ποιότητας Ζωής
Τμήμα Οργάνωσης και Διαχείρισης Αθλητισμού

Η διάλεξη βασίζεται στο Κεφάλαιο 9 του συγγράμματος «Βαγενάς, Γ. Κ. (2019). *Στατιστικές εφαρμογές στην αθλητική επιστήμη με παραδείγματα στο SPSS* (7η Έκδοση). Αθήνα: Τζιόλας.

1

Στοιχεία θεωρίας Δείγματος

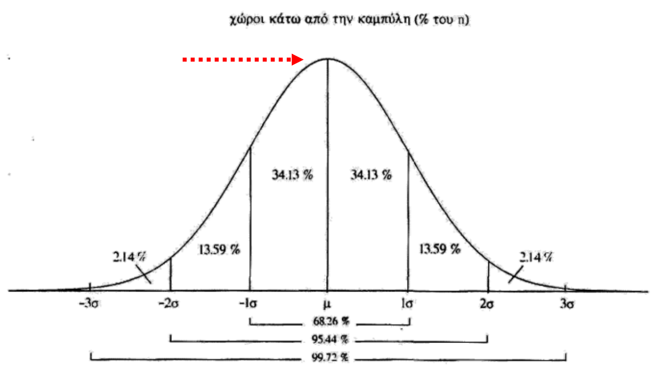
- Η έννοια της κανονικότητας της πληθυσμιακής κατανομής ορίζει ότι η πιθανότητα τυχαίας επιλογής μιας τιμής Χ ενός κανονικού πληθυσμού ***n*** εξαρτάται από την τυπική της απόσταση (z) από τον μέσο μ της κατανομής.
  - Υπενθυμίζεται ότι οι ακραίες σε μέγεθος τιμές, δηλαδή αυτές που απέχουν περισσότερο από 3 τυπικές αποκλίσεις από τον μ, έχουν πιθανότητα επιλογής σε ένα τυχαίο δείγμα περίπου 0.3% συνολικά.



Εικ. 8.3 - Κανονική καμπύλη : κατανομή %f κατά **σ**

2

Στοιχεία θεωρίας Δείγματος



Εικ. 8.3 - Κανονική καμπύλη : κατανομή %f κατά σ

- Ο συνδυασμός των δύο αυτών εννοιών υποδεικνύει ότι σε μια κανονικής μορφής πληθυσμιακή κατανομή η πιθανότητα συμπερίληψης σε ένα τυχαίο δείγμα Ν μιας τυχαίας τιμής Χ, της οποίας η πληθυσμιακή συχνότητα είναι f, είναι ίση **<u>με το ύψος της συνάρτησης F(x)=y</u>** στο σημείο z=(μ-Χ)/σ ή πιο απλά ίση με (f/n)N.

Στοιχεία θεωρίας Δείγματος π.χ.

- Για παράδειγμα, αν υποθέσουμε ότι σε έναν πληθυσμό n = 1000 τιμών μια τυχαία τιμή Χ έχει συχνότητα f = 20, η πιθανότητα επιλογής της σε ένα τυχαίο <u>δείγμα Ν = 50 τιμών</u> είναι
  - (f/n)N = (20/1000)50 = 100/100.
- Αν τώρα υποθέσουμε ότι η συχνότητα μιας άλλης τιμής Χ στον ίδιο πληθυσμό n = 1000 τιμών είναι f = 100, τότε η πιθανότητα επιλογής της σε ένα τυχαίο <u>δείγμα Ν = 100 τιμών</u> είναι
  - (f/n)N=(100/1000)100 = 1000/100.

Στοιχεία θεωρίας Δείγματος

- Η **δειγματική θεωρία ασχολείται γενικά με τις σχέσεις πληθυσμών και δειγμάτων** και παρέχει τη θεωρητική βάση για την ανάπτυξη των μεθόδων *εκτίμησης* των πληθυσμιακών *παραμέτρων* από τα αντίστοιχα δειγματικά μεγέθη, τα *στατιστικά.*
- Χωρίς τη θεωρία αυτή, μέρος της οποίας αποτελεί η δειγματική κατανομή που θα δούμε αναλυτικά στη συνέχεια, θα ήταν αδύνατος ο αντικειμενικός (στατιστικός) έλεγχος της σημαντικότητας της διαφοράς ή της σχέσης δύο δειγματικών μεταβλητών.
- Άλλωστε, όπως έχουμε τονίσει και σε προηγούμενες περιπτώσεις, η διαδικασία της μελέτης δειγμάτων για την εξαγωγή συμπερασμάτων για πληθυσμιακές ιδιότητες, συνυπολογίζοντας και τον έλεγχο της ακρίβειας των συμπερασμάτων αυτών, αποτελεί μέρος της **στατιστικής επαγωγής.**

5

5

Στοιχεία θεωρίας Δείγματος

- Για να είναι δυνατή όμως η αποδοχή της εγκυρότητας των συμπερασμάτων που προκύπτουν από αναλύσεις δειγμάτων πρέπει να πληρείται η αναγκαία **συνθήκη το δείγμα να είναι αντιπροσωπευτικό του πληθυσμού.**
- Στον τομέα αυτό υπεισέρχεται ο **πειραματικός σχεδιασμός** που μελετά τις δειγματοληπτικές τεχνικές καθώς και κάθε πρόβλημα ή συνθήκη που μπορεί να επηρεάσει άμεσα ή έμμεσα την αμεροληψία και την ακρίβεια της πειραματικής διαδικασίας.
- Ειδικά για την επιλογή του δείγματος χρησιμοποιείται η διαδικασία της **τυχαίας δειγματοληψίας** σύμφωνα με την οποία **όλα τα μέλη του υπό εξέταση πληθυσμού έχουν ίση πιθανότητα συμπερίληψης στο δείγμα.**
- Όπως είδαμε και στο 1^{ο} κεφάλαιο, ένας πληθυσμός μπορεί να είναι **άπειρος** (να αποτελείται από άπειρο αριθμό *μελών*), ή να είναι **πεπερασμένος** (να αποτελείται από ένα συγκεκριμένο αριθμό μελών).

6

6

Οι γενικοί τρόποι επιλογής ενός τυχαίου δείγματος είναι δύο:

- (α) Κάθε μέλος του πληθυσμού που συμπεριλαμβάνεται σε μια τυχαία δειγματοληψία **να επανατίθεται** στον πληθυσμό.
  - Έτσι το μέλος αυτό έχει τη πιθανότητα τυχαίας συμπερίληψης του και σε κάθε επόμενη δειγματοληψία.
  - Η τεχνική αυτή μπορεί να μετατρέψει περιστασιακά έναν πεπερασμένο πληθυσμό σε άπειρο, λέγεται **τυχαία δειγματοληψία *με επανάθεση*** και παρέχει τη πιθανότητα επιλογής άπειρου αριθμού δειγμάτων.
- (β) Κάθε μέλος του πληθυσμού που συμπεριλαμβάνεται σε μια τυχαία δειγματοληψία **να μην επανατίθεται** στον πληθυσμό.
  - Έτσι το μέλος αυτό δεν έχει τη πιθανότητα τυχαίας συμπερίληψης του σε κάποια από τις επόμενες δειγματοληψίες.
  - Η τεχνική αυτή λέγεται **τυχαία δειγματοληψία *χωρίς επανάθεση*** και, θεωρητικά, έχει σαν αποτέλεσμα την παραγωγή ενός πεπερασμένου αριθμού δειγμάτων, όταν ο πληθυσμός είναι πεπερασμένος.

7

7

Τρόποι επιλογής ενός τυχαίου δείγματος

- Στη στατιστική πρακτική έχει επικρατήσει η άποψη ότι, όταν ο πληθυσμός **είναι πεπερασμένος** αλλά σχετικά πολύ μεγάλος (π.χ. n>10.000), τότε η διαδικασία της δειγματοληψίας, **ανεξάρτητα αν είναι με επανάθεση ή όχι**, μπορεί για πρακτικούς λόγους να θεωρηθεί ότι είναι δειγματοληψία από έναν **άπειρο πληθυσμό**.
  - Αυτό διευκολύνει τη διατύπωση ορισμένων στατιστικών παραδοχών που είναι απαραίτητες για την ανάπτυξη της επαγωγικής διαδικασίας.

- Ο αριθμός των τυχαίων ισοπληθών δειγμάτων N, που είναι πιθανό να επιλεγούν **με επανάθεση** από έναν **πεπερασμένο πληθυσμό n,** είναι ίσος με τον **αριθμό k** όλων των πιθανών συνδυασμών **πλήθους N** που μπορούν να προκύψουν με την επανάθεση και δίνεται από τη σχέση

$$\mathbf{k = n^N}$$

8

8

Τρόποι επιλογής ενός τυχαίου δείγματος

- Για **παράδειγμα,** τα πιθανά δείγματα πλήθους **N = 2** που μπορούν να επιλεγούν **με επανάθεση** από ένα πληθυσμό **n = 5** είναι **k = n^N = 5^2 = 25**.
- Έτσι, από ένα πληθυσμό αποτελούμενο από τα μέλη Α, Β, Γ, Δ, Ε μπορούμε να πάρουμε τα εξής **25 πιθανά δείγματα πλήθους** N = 2:

Α, Α	Α, Β	Α, Γ	Α, Δ	Α, Ε
Β, Α	Β, Β	Β, Γ	Β, Δ	Β, Ε
Γ, Α	Γ, Β	Γ, Γ	Γ, Δ	Γ, Ε
Δ, Α	Δ, Β	Δ, Γ	Δ, Δ	Δ, Ε
Ε, Α	Ε, Β	Ε, Γ	Ε, Δ	Ε, Ε

Τρόποι επιλογής ενός τυχαίου δείγματος

- Με τον ίδιο τρόπο, από έναν πληθυσμό n=10 μελών μπορούν να επιλεγούν *με επανάθεση* $k=n^N=10^2=100$ δείγματα πλήθους N=2,
  - $k=n^N=10^3=1000$ δείγματα πλήθους N=3,
  - $k=n^N=10^4=10000$ δείγματα πλήθους N=4,
  - $k= n^N=10^5=100000$ δείγματα πλήθους N=5.
- Ο αριθμός των τυχαίων ισοπληθών δειγμάτων N, που μπορούν να επιλεγούν **χωρίς επανάθεση** από έναν **πεπερασμένο πληθυσμό n,** είναι ίσος με **τον αριθμό k** όλων των πιθανών συνδυασμών **πλήθους N** *χωρίς επανάθεση* και δίνεται από τη σχέση

$$k={}_nC_N = \frac{n!}{(n-N)!N!}$$

Τρόποι επιλογής ενός τυχαίου δείγματος

- *Αξίζει να επισημανθεί στο σημείο αυτό ότι στην πλειονότητα των περιπτώσεων στη Φυσική Αγωγή και στον Αθλητισμό οι στατιστικοί πληθυσμοί είναι πεπερασμένοι, ενώ οι **περισσότερες δειγματοληψίες είναι χωρίς επανάθεση**.*
- *Αυτό συμβαίνει για τον λόγο ότι η επιστημονική έρευνα στους τομείς αυτούς στηρίζεται κατά κύριο λόγο σε πληθυσμούς ατόμων τα οποία πρακτικά δε γίνεται να συμπεριληφθούν στην ίδια δειγματοληψία δύο φορές.*
  - *(Το παραγοντικό ενός φυσικού αριθμού n συμβολίζεται με n!, διαβάζεται νι παραγοντικό, και είναι το γινόμενο όλων των θετικών ακεραίων μικρότερων ή ίσων με ν).*
    n! = 1 · 2 · 3 · ... · n)

$$k = {}_nC_N = \frac{n!}{(n-N)!N!}$$

11

11

Τρόποι επιλογής ενός τυχαίου δείγματος

- Έτσι από τον προηγούμενο πληθυσμό με τα 5 μέλη μπορούμε να έχουμε τους εξής 10 συνδυασμούς *χωρίς επανάθεση;*

Α, Β	Α, Γ	Α, Δ	Α, Ε
	Β, Γ	Β, Δ	Β, Ε
		Γ, Δ	Γ, Ε
			Δ, Ε

- πράγμα που αποδεικνύεται και από την εφαρμογή του τύπου

$$k = {}_nC_N = \frac{n!}{(n-N)!N!}$$

$$k = {}_5C_2 = \frac{5!}{(5-2)!\cdot 2!} = \frac{5\cdot4\cdot3\cdot2\cdot1}{(3\cdot2\cdot1)\cdot(2\cdot1)} = \frac{20}{2} = 10$$

- *Με την τεχνική αυτή ένα μέλος του πληθυσμού δεν μπορεί να συμπεριληφθεί δύο φορές στο ίδιο δείγμα και για το λόγο αυτό ο αριθμός των συνδυασμών που μπορούν να γίνουν χωρίς επανάθεση είναι μικρότερος από αυτόν με επανάθεση.*

12

12

- Σε ένα άλλο πιο πρακτικό **παράδειγμα,** από έναν **πληθυσμό n=8** περιπτώσεων μπορούν να γίνουν ***χωρίς επανάθεση*** **k = 56** συνδυασμοί πλήθους **N=3**:

$$k = {}_8C_3 = \frac{8!}{(8-3)!\cdot 3!} = \frac{8!}{5!\cdot 3!} = \frac{8\cdot 7\cdot 6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 1}{(5\cdot 4\cdot 3\cdot 2\cdot 1)\cdot(3\cdot 2\cdot 1)} = \frac{336}{6} = 56$$

- Με τον ίδιο τρόπο, σε ένα πλήθος n=10 περιπτώσεων μπορούν να γίνουν k=120 συνδυασμοί των N=3 περιπτώσεων

$$k = {}_{10}C_3 = \frac{10!}{(10-3)!\cdot 3!} = \frac{10\cdot 9\cdot 8\cdot 7\cdot 6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 1}{(7\cdot 6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 1)\cdot(3\cdot 2\cdot 1)} = \frac{720}{6} = 120$$

- ενώ αν το πλήθος αυτό γίνει n=15 μπορούν να γίνουν k=455 συνδυασμοί των N=3 περιπτώσεων

$$k = {}_{15}C_3 = \frac{15!}{(15-3)!\cdot 3!} = \frac{15!}{12!\cdot 3!} = \frac{15\cdot 14\cdot 13\cdot\ldots\cdot 3\cdot 2\cdot 1}{(12\cdot 11\cdot 10\cdot\ldots\cdot 3\cdot 2\cdot 1)\cdot(3\cdot 2\cdot 1)}$$

$$= \frac{15\cdot 14\cdot 13}{3\cdot 2\cdot 1} = \frac{2730}{6} = 455$$

- Έτσι, γίνεται φανερό ότι, θεωρητικά, **για πολύ μεγάλα πληθυσμιακά σύνολα n**, ανάλογα βέβαια και με το **δειγματικό πλήθος N**, μπορεί να εξαχθεί ένας απροσδιόριστα μεγάλος αριθμός k τυχαίων ισοπληθών δειγμάτων.

13

Στον Πιν. 9.1 δίνονται απλά παραδείγματα όλων των πιθανών δειγμάτων επιλεγμένου μεγέθους N που μπορούν να επιλεγούν *χωρίς επανάθεση* από μικρούς πληθυσμούς (n=5 μέχρι n=25). Από τα δεδομένα του πίνακα είναι εμφανές ότι όσο το n μεγαλώνει τόσο αυξάνει ο αριθμός των πιθανών δειγμάτων, μικρού έστω πλήθους N.

Πίνακας 9.1 -Πιθανά (k) δείγματα
(N) από μικρούς πληθυσμούς (n) *χωρίς επανάθεση*

ΔΕΙΓΜΑ	ΜΙΚΡΟΙ ΠΛΗΘΥΣΜΟΙ (n)				
(N)	5	10	15	20	25
1	5	10	15	20	25
2	10	45	105	190	300
3	10	120	455	1140	**2300**
4	5	210	1365	4845	12650
5	-	252	3003	15504	53130

$k={}_nC_N$, π.χ. ο k=2300 σημαίνει ότι από ένα πληθυσμό n=25 μπορούμε να πάρουμε μέχρι 2300 τυχαία δείγματα N=3 χωρίς *επανάθεση*.

14

Στον Πιν. 9.2 δίνονται απλά παραδείγματα όλων των πιθανών δειγμάτων επιλεγμένου μεγέθους N που μπορούν να επιλεγούν *με επανάθεση* από μικρούς πληθυσμούς (n=5 μέχρι n=25). Από τα δεδομένα του πίνακα είναι εμφανές ότι όσο το n μεγαλώνει τόσο αυξάνει ο αριθμός των πιθανών δειγμάτων, μικρού έστω πλήθους N.

Πίνακας 9.2 -Πιθανά (k) δείγματα (N) από μικρούς πληθυσμούς (n) *με επανάθεση*

ΔΕΙΓΜΑ	ΜΙΚΡΟΙ ΠΛΗΘΥΣΜΟΙ (n)				
(N)	5	10	15	20	25
1	5	10	15	20	25
2	25	100	225	400	625
3	125	1000	3375	8000	**15625**
4	625	10000	50625	160000	390625
5	3125	100000	759375	3200000	9765625

$k=n^N$, π.χ. ο 15625 σημαίνει ότι από ένα πληθυσμό n=25 μπορούμε να πάρουμε μέχρι 15625 τυχαία δείγματα πλήθους N=3 με *επανάθεση*.

15

Δειγματική Κατανομή του Μέσου

- Αν από ένα πληθυσμιακό σύνολο **n** πάρουμε (χωρίς ή με επανάθεση) όλα τα k πιθανά δείγματα πλήθους N, λογαριάσουμε τους αριθμητικούς τους μέσους και παράγουμε την κατανομή τους τότε η κατανομή αυτή λέγεται **δειγματική κατανομή** και έχει μέση τιμή ίση με τη μέση τιμή της πληθυσμιακής κατανομής: $\mu = \mu_{\bar{x}}$.

16

16

Δειγματική Κατανομή του Μέσου

- Αυτό μπορεί να αποδειχθεί με τον εξής απλό τρόπο. Από έναν πληθυσμό n=5 αριθμών παίρνουμε *χωρίς επανάθεση* όλα τα πιθανά δείγματα πλήθους N=2. Τα δείγματα αυτά είναι

  k = (n!) / [(n-N)! N!] = (5!) / (3!) (2!) = 10.

Αν οι 5 αριθμοί παρασταθούν με X_1, X_2, X_3, X_4 & X_5, τότε οι 10 πιθανοί συνδυασμοί (δείγματα χωρίς επανάθεση) θα είναι

$X_1\ X_2$	$X_1\ X_3$	$X_1\ X_4$	$X_1\ X_5$
	$X_2\ X_3$	$X_2\ X_4$	$X_2\ X_5$
		$X_3\ X_4$	$X_3\ X_5$
			$X_4\ X_5$

και οι αντίστοιχες 10 μέσες τιμές θα είναι

$$\overline{X}_1=\frac{X_1+X_2}{2},\overline{X}_2=\frac{X_1+X_3}{2},\overline{X}_3=\frac{X_1+X_4}{2},\overline{X}_4=\frac{X_1+X_5}{2},\overline{X}_5=\frac{X_2+X_3}{2}$$

$$\overline{X}_6=\frac{X_2+X_4}{2},\overline{X}_7=\frac{X_2+X_5}{2},\overline{X}_8=\frac{X_3+X_4}{2},\overline{X}_9=\frac{X_3+X_5}{2},\overline{X}_{10}=\frac{X_4+X_5}{2}$$

Ο δειγματικός μέσος στην περίπτωση αυτή θα είναι

$$\mu_{\overline{X}}=\frac{1}{10}\left(\overline{X}_1+\overline{X}_2+\overline{X}_3+\overline{X}_4+\overline{X}_5+\overline{X}_6+\overline{X}_7+\overline{X}_8+\overline{X}_9+\overline{X}_{10}\right)=$$

$$=\frac{1}{10}\left(\frac{X_1+X_2}{2}+\frac{X_1+X_3}{2}+\ldots+\frac{X_3+X_5}{2}+\frac{X_4+X_5}{2}\right)$$

$$=\frac{1}{10}\left(\frac{4X_1+4X_2+4X_3+4X_4+4X_5}{2}\right)$$

$$=\frac{1}{5}\left(X_1+X_2+X_3+X_4+X_5\right)=\mu$$

Το ίδιο ακριβώς ισχύει και μπορεί εύκολα να αποδειχθεί, όταν τα δείγματα βγαίνουν *με επανάθεση*

Αν ο **πληθυσμός** από τον οποίο παίρνεται το δείγμα είναι **άπειρος** ή θεωρείται άπειρος λόγω του μεγάλου του μεγέθους ή αν το δείγμα παίρνεται από ένα **πεπερασμένο πληθυσμό** ***με επανάθεση,*** τότε

ο **δειγματικός μέσος** είναι

$$\mu_{\overline{X}} = \frac{1}{k}\sum_{j=1}^{k}\overline{X}_j = \frac{1}{k}\left(\overline{X}_1 + \overline{X}_2 + \overline{X}_3 + \ldots + \overline{X}_k\right)$$

η **δειγματική διασπορά** είναι

$$\sigma^2_{\overline{X}} = \frac{1}{k}\sum_{j=1}^{k}\left(\overline{X}_j - \mu\right)^2$$

και η **δειγματική τυπική απόκλιση** θα είναι

$$\sigma_{\overline{X}} = \sqrt{\frac{1}{k}\sum_{j=1}^{k}\left(\overline{X}_j - \mu\right)^2} = \sqrt{\sigma^2_{\overline{X}}}$$

- **όπου j = 1, 2,. .., k ο αριθμός των μέσων τιμών (αριθμός δειγμάτων).**

Δειγματική Κατανομή του Μέσου

- Η τυπική απόκλιση της δειγματικής κατανομής λέγεται **τυπικό σφάλμα** του **δειγματικού μέσου** και σε **δειγματοληψίες με επανάθεση** είναι

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{N}}$$

- όπου σ η τυπική απόκλιση του γεννήτορα πληθυσμού.

- Αν η **δειγματοληψία** γίνεται ***χωρίς επανάθεση,*** τότε το τυπικό σφάλμα του δειγματικού μέσου δίνεται από τη σχέση

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{N}}\sqrt{\frac{n-N}{n-1}}$$

- **Η τετραγωνική ρίζα του πηλίκου (n-N)/(n-1) αποτελεί διόρθωση για σχετικά μικρούς πληθυσμούς**.
- Για μεγάλους πεπερασμένους πληθυσμούς (π.χ. n>10000) η διόρθωση αυτή είναι κατά προσέγγιση 1, καθότι τα συνήθη δειγματικά πλήθη N είναι αμελητέα συγκριτικά με το n.

- **Για τον λόγο αυτό, και επειδή συνήθως οι πειραματικές έρευνες ασχολούνται, θεωρητικά, με μεγάλους στατιστικούς πληθυσμούς, η διόρθωση αυτή παραλείπεται και το τυπικό σφάλμα του μέσου της δειγματικής κατανομής υπολογίζεται από την απλούστερη σχέση**

Π.χ. 9.1 - Δείγμα *χωρίς επανάθεση*: μέσος & διασπορά.

- Έστω ότι ένας πληθυσμός n=5 αποτελείται από τους αριθμούς 1, 2, 3, 4, 5. Ο πληθυσμός αυτός έχει …

μέση τιμή $\mu = (1+2+3+4+5) / 5 = 15 / 5 = 3$

διασπορά $\sigma^2 = [(1-3)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2] / 5$

$= [4 + 1 + 0 + 1 + 4] / 5 = 10 / 5 = 2$

τυπική απόκλιση $\sigma = \sqrt{2} = 1.4142$

21

21

Οι πιθανοί συνδυασμοί πλήθους N=2 που μπορούν να γίνουν *χωρίς επανάθεση* στον πληθυσμό αυτό είναι $k={}_nC_N={}_5C_2=10$ (πίν. 9.1).

Έτσι, τα αντίστοιχα δείγματα πλήθους N=2 που μπορούν να βγουν είναι τα εξής 10:

1, 2	1, 3	1, 4	1, 5
	2, 3	2, 4	2, 5
		3, 4	3, 5
			4, 5

Οι μέσοι των 10 αυτών δειγμάτων είναι

(1+2)/2=1.5	(1+3)/2=2	(1+4)/2=2.5	(1+5)/2=3
	(2+3)/2=2.5	(2+4)/2=3	(2+5)/2=3.5
		(3+4)/2=3.5	(3+5)/2=4
			(4+5)/2=4.5

Ο δειγματικός μέσος είναι

$$\mu_X = [\ 1.5 + 2 + 2.5 + 3 + 2.5 + 3 + 3.5 + 3.5 + 4 + 4.5\ ] / 10 = 30 / 10 = 3$$

ίσος δηλαδή με τον πληθυσμιακό μέσο $\mu = 3$.

22

Η δειγματική διασπορά είναι

$$\sigma^2_{\bar{X}} = [\ (1.5-3)^2 + (2-3)^2 + (2.5-3)^2 + (3-3)^2 + (2.5-3)^2 + (3-3)^2 + (3.5-3)^2 + (3.5-3)^2 + (4-3)^2 + (4.5-3)^2\ ]\ /\ 10$$

$$= [(-1.5)^2 + (-1)^2 + (-0.5)^2 + (0)^2 + (-0.5)^2 + (0)^2 + (0.5)^2 + (0.5)^2 + (1)^2 + (1.5)^2]\ /10$$

$$= [\ 2.25 + 1 + 0.25 + 0 + 0.25 + 0 + 0.25 + 0.25 + 1 + 2.25\ ]\ /\ 10$$

$$= 7.5\ /\ 10 = \mathbf{0.75}$$

Η δειγματική τυπική απόκλιση είναι

$$\sigma_{\overline{X}} = \sqrt{0.75} = 0.87$$

Επαλήθευση - Χρησιμοποιώντας τον τύπο 9.7 του τυπικού σφάλματος του μέσου (χωρίς επανάθεση) και με δεδομένα μ=3, σ^2=2, n=5 & N=2 η δειγματική διασπορά είναι :

$$\sigma^2_{\bar{X}} = \frac{\sigma^2}{N} \cdot \frac{n-N}{n-1} = \frac{2}{2} \cdot \frac{5-2}{5-1} = \frac{3}{4} = 0.75$$

και η δειγματική τυπική απόκλιση (τυπικό σφάλμα του μέσου), είναι:

$$\sigma_{\overline{X}} = \sqrt{0.75} = 0.87$$

Αυτό επαληθεύει την ισχύ του τύπου 9.7 για τον υπολογισμό του τυπικού σφάλματος του μέσου για δειγματοληψίες *χωρίς επανάθεση*.

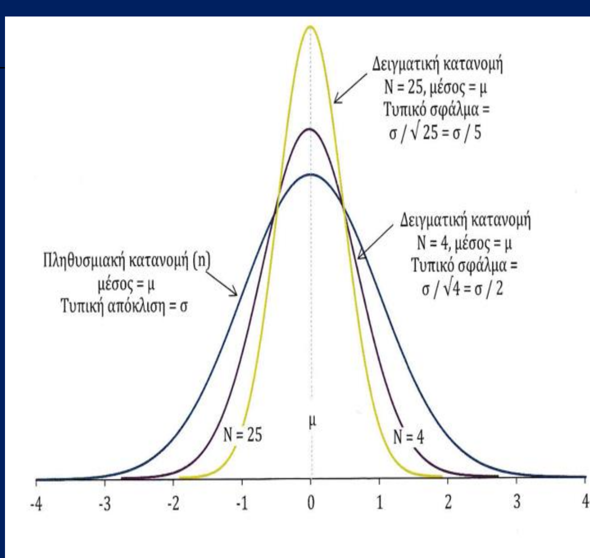
23

Καμπύλη Δειγματικής Κατανομής

- Η καμπύλη της δειγματικής κατανομής, δηλαδή των μέσων τιμών των πιθανών δειγμάτων, είναι συμμετρική και ορίζεται από τις σχέσεις

$$Y = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \qquad\qquad z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}}$$

- Η κυρτότητα της καμπύλης της δειγματικής κατανομής εξαρτάται από το μέγεθος του δειγματικού πλήθους N. Όσο μεγαλύτερο είναι το δειγματικό πλήθος N τόσο πιο λεπτόκυρτη βγαίνει η συμμετρική καμπύλη της κατανομής των μέσων τιμών των δειγμάτων.



Πληθυσμιακή κανονική και 2 δειγματικές κατανομές

24

Τι γίνεται όμως στην περίπτωση που ο γεννήτορας πληθυσμός δεν παρουσιάζει κανονική κατανομή: Η απάντηση δίνεται από το θεώρημα των κεντρικών ορίων (central limit theorem).

- Το **θεώρημα των κεντρικών ορίων** ορίζει ότι σε μια διαδικασία εξαγωγής όλων των πιθανών ισοπληθών τυχαίων δειγμάτων N από ένα πληθυσμό ***n*** μιας μεταβλητής **X**, μέσου **μ** και τυπικής απόκλισης **σ**, *όσο **το μέγεθος N του δείγματος μεγαλώνει τόσο η κατανομή**, με μέσο **μ** και τυπική απόκλιση **σ-**, που σχηματίζεται από τους δειγματικούς μέσους (δειγματική κατανομή) **παίρνει όλο και πιο κανονική μορφή**, ανεξάρτητα από την κανονικότητα ή μη της αρχικής πληθυσμιακής κατανομής.*
  - Αυτό το θεώρημα έχει τεράστια πρακτική αξία**, καθότι επιτρέπει τη χρήση της τυπικής κατανομής z** για τον υπολογισμό της πιθανότητας εμφάνισης της μέσης τιμής ενός δείγματος σε κάποια συγκεκριμένη τυπική απόσταση από το μ ή την πιθανότητα εμφάνισης της μέσα σε ένα συγκεκριμένο εύρος τιμών πριν ή μετά το μ.
- *Με απλά λόγια το θεώρημα των κεντρικών ορίων επιτρέπει τη χρήση της κανονικής κατανομής ως στατιστικού προτύπου για τη δειγματική κατανομή του μέσου σε πληθώρα προβλημάτων στατιστικής επαγωγής.*

25

25

Π.χ. 9.3 - Πιθανότητα μέσου - *5 χαρακτηριστικές περιπτώσεις.*

- Δίνεται η μεταβλητή *κοιλιακές αναδιπλώσεις σε 30" σπουδαστών ηλικίας 20-25 ετών.*
- Η μεταβλητή αυτή έχει κανονική πληθυσμιακή κατανομή με μέσο **μ=26** και τυπική απόκλιση **σ=3**.
- **Ζητείται να υπολογιστεί η πιθανότητα (p) η μέση τιμή ενός τυχαίου δείγματος της μεταβλητής αυτής να βγει ίση ή μεγαλύτερη του 27**.
  - **Δηλαδή αν πάρουμε από τον γεννήτορα πληθυσμό ένα τυχαίο δείγμα, ποια είναι η πιθανότητα η μέση τιμή του δείγματος αυτού να βγει 27**.
- Για να επιλυθεί το πρόβλημα αυτό, χρειάζεται να είναι γνωστό το μέγεθος N του δείγματος.
- Προκειμένου λοιπόν να γίνει μια αναλυτική παρουσίαση της επίδρασης του μεγέθους του δείγματος στην περίπτωση αυτή, επιλέχτηκαν και παρουσιάζονται 5 χαρακτηριστικές περιπτώσεις δειγματικού πλήθους N (4, 9, 25, 49 & 100).

26

26

Π.χ. 9.3 - Πιθανότητα μέσου - *5 χαρακτηριστικές περιπτώσεις.*

Περίπτωση 1η – πολύ μικρό πλήθος δείγματος : N=4

Λογαριάζουμε το τυπικό σφάλμα του μέσου

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{N}} = \frac{3}{\sqrt{4}} = \frac{3}{2} = 1.5$$

Μετασχηματίζουμε την κρίσιμη τιμή 27 σε τυπική τιμή z

$$z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} = \frac{27-26}{1.5} = \frac{1}{1.5} = 0.67$$

Από το παράρτημα Γ διαπιστώνουμε ότι ο χώρος κάτω από την τυπική κανονική καμπύλη από το z=0.67 και μετά είναι ίσος με 0.2514 ή 25.14%.

Αυτό σημαίνει ότι, αν από το πληθυσμιακό σύνολο αυτής της μεταβλητής πάρουμε ένα τυχαίο δείγμα πλήθους N=4, υπάρχει πιθανότητα περίπου 25% η μέση τιμή του δείγματος αυτού να βγει ≥27, δηλαδή ≥μ+1 μονάδα της κλίμακας μέτρησης της X.

27

Περίπτωση 2η – μικρό πλήθος δείγματος : N=9

Λογαριάζουμε το τυπικό σφάλμα του μέσου

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{N}} = \frac{3}{\sqrt{9}} = \frac{3}{3} = 1$$

Μετασχηματίζουμε την κρίσιμη τιμή 27 σε τυπική τιμή z

$$z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} = \frac{27-26}{1} = \frac{1}{1} = 1$$

Από το παράρτημα Γ διαπιστώνουμε ότι ο χώρος κάτω από την τυπική κανονική καμπύλη από το z=1 και μετά είναι ίσος με 0.1587 ή 15.87%.

Αυτό σημαίνει ότι, αν από το πληθυσμιακό σύνολο αυτής της μεταβλητής πάρουμε ένα τυχαίο δείγμα πλήθους N=9, υπάρχει πιθανότητα περίπου 16% η μέση τιμή του δείγματος αυτού να βγει ≥27, δηλαδή ≥μ+1 μονάδα της κλίμακας μέτρησης της X.

28

Περίπτωση 3η – μέτριο πλήθος δείγματος : N=25

Λογαριάζουμε το τυπικό σφάλμα του μέσου

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{N}} = \frac{3}{\sqrt{25}} = \frac{3}{5} = 0.60$$

Μετασχηματίζουμε την κρίσιμη τιμή 27 σε τυπική τιμή z

$$z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} = \frac{27 - 26}{0.60} = \frac{1}{0.60} = 1.67$$

Από το παράρτημα Γ διαπιστώνουμε ότι, ο χώρος κάτω από την τυπική κανονική καμπύλη από το z=1.67 και μετά είναι ίσος με 0.0475 ή 4.75%.

Αυτό σημαίνει ότι αν από το πληθυσμιακό σύνολο αυτής της μεταβλητής πάρουμε ένα τυχαίο δείγμα πλήθους N=25, υπάρχει πιθανότητα περίπου 5% η μέση τιμή του δείγματος αυτού να βγει ≥27, δηλαδή ≥μ+1 μονάδα της κλίμακας μέτρησης της X.

29

Περίπτωση 4η – μεγάλο πλήθος δείγματος : N=49

Λογαριάζουμε το τυπικό σφάλμα του μέσου

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{N}} = \frac{3}{\sqrt{49}} = \frac{3}{7} = 0.43$$

Μετασχηματίζουμε την κρίσιμη τιμή 27 σε τυπική τιμή z

$$z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} = \frac{27 - 26}{0.43} = \frac{1}{0.43} = 2.33$$

Από το παράρτημα Γ διαπιστώνουμε ότι ο χώρος κάτω από την τυπική κανονική καμπύλη από το z=2.33 και μετά είναι ίσος με 0.0099 ή 0.99%.

Αυτό σημαίνει ότι, αν από το πληθυσμιακό σύνολο αυτής της μεταβλητής πάρουμε ένα τυχαίο δείγμα πλήθους N=49, υπάρχει πιθανότητα περίπου 1% η μέση τιμή του δείγματος αυτού να βγει ≥27, δηλαδή ≥μ+1 μονάδα της κλίμακας μέτρησης της X.

30

Περίπτωση 5η – πολύ μεγάλο πλήθος δείγματος : N=100

Λογαριάζουμε το τυπικό σφάλμα του μέσου

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{N}} = \frac{3}{\sqrt{100}} = \frac{3}{10} = 0.30$$

Μετασχηματίζουμε την κρίσιμη τιμή 27 σε τυπική τιμή z

$$z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} = \frac{27-26}{0.30} = \frac{1}{0.30} = 3.33$$

Από το παράρτημα Γ διαπιστώνουμε ότι ο χώρος κάτω από την τυπική κανονική καμπύλη από το z=3 και μετά είναι ίσος με 0.0013 ή 0.13%.

Αυτό σημαίνει ότι, αν από το πληθυσμιακό σύνολο αυτής της μεταβλητής πάρουμε ένα τυχαίο δείγμα πλήθους N=100, υπάρχει πιθανότητα περίπου 0.13% η μέση τιμή του δείγματος αυτού να βγει ≥27, δηλαδή ≥μ+1 μονάδα της κλίμακας μέτρησης της X.

31

Τα συνοπτικά αποτελέσματα της ανάλυσης των 5 περιπτώσεων του παραδείγματος 9.3 παρουσιάζονται στον πίνακα 9.3, από τον οποίο διαπιστώνεται ότι όσο το N αυξάνει, τόσο μειώνεται το τυπικό σφάλμα του μέσου, τόσο αυξάνει η τιμή z της εξεταζόμενης μέσης τιμής 27 και **τόσο μειώνεται αντίστοιχα η πιθανότητα εμφάνισης μιας μέσης τιμής ίσης ή μεγαλύτερης του 27**.

Πίν. 9.3 – Πλήθος N και πιθανότητα εμφάνισης μέσου ≥27

Δείγμα N	ΤυπικόΣφάλμα σ_X	ΤυπικήΤιμή z	Πιθανότητα p	%
4	1.5	0.67	0.2514	25%
9	1	1	0.1587	16%
25	0.6	1.67	0.0475	5%
49	0.43	2.33	0.0099	1%
100	0.333	3	0.0013	0.1%

Με βάση την κανονική δειγματική κατανομή με μ=26 και σ=3 (κοιλιακές αναδιπλώσεις σε 30″).

32

Παράδειγμα 9.4 - Εύρεση πιθανότητας μέσου - *Διάστημα τιμών: 2 περιπτώσεις.*

- Από την πληθυσμιακή κατανομή του προηγούμενου παραδείγματος (κοιλιακές αναδιπλώσεις με μ = 26 & σ = 3) παίρνουμε ένα τυχαίο δείγμα N = 100 ατόμων.
  - Διατηρούμε το μέγεθος του δείγματος αυτού σταθερό και **εξετάζουμε τον υπολογισμό της πιθανότητας (p) η μέση τιμή του τυχαίου δείγματος να ανήκει σε 2 χαρακτηριστικές περιπτώσεις διαστήματος τιμών της κλίμακας μέτρησης της μεταβλητής** : **μ+1 μονάδα & μ+0.5 μονάδες**.
- **Για να επιλυθεί το πρόβλημα αυτό**, πρέπει πρώτα να μετασχηματιστούν τα 2 όρια κάθε διαστήματος τιμών σε τιμές z και στη συνέχεια να υπολογιστεί το εμβαδόν του χώρου που περιλαμβάνεται μεταξύ της (κανονικής) καμπύλης και των τεταγμένων που αντιστοιχούν στις 2 τιμές z κάθε διαστήματος.

33

Περίπτωση 1η – Πιθανότητα εμφάνισης ενός μέσου στο συμμετρικό διάστημα μ±1 μονάδα.

Ποια είναι η πιθανότητα ο μέσος του τυχαίου δείγματος N=100 να πέσει στο διάστημα 25–27, δηλαδή να έχει τιμή $25 \leq \bar{X} \leq 27$;

Η πιθανότητα αυτή είναι $p(z_1 - z_2) = p(z_1) - p(z_2)$ και επειδή ο χώρος κάτω από την κανονική δειγματική καμπύλη είναι στην περίπτωση αυτή συμμετρικός $|z_1| = z_2$ και $p = 2p(z_1)$.

Το τυπικό σφάλμα του μέσου είναι

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}} = \frac{3}{\sqrt{100}} = \frac{3}{10} = 0.30$$

Οι κρίσιμες τιμές 25 & 27 έχουν τυπικές τιμές z

$$z_1 = \frac{\bar{X}_1 - \mu}{\sigma_{\bar{X}}} = \frac{25 - 26}{0.30} = \frac{-1}{0.30} = -3.33$$

$$z_2 = \frac{\bar{X}_2 - \mu}{\sigma_{\bar{X}}} = \frac{27 - 26}{0.30} = \frac{1}{0.30} = +3.33$$

Από το παράρτημα Γ διαπιστώνουμε ότι ο χώρος κάτω από την τυπική κανονική καμπύλη μεταξύ z=3.33 και 0 είναι ίσος με 0.49957.
Έτσι, η πιθανότητα που ζητάμε είναι $p = 2\,p(z_1) = 2(0.49957) = 0.9991$.
Αυτό σημαίνει ότι με μέγεθος δείγματος N=100 για τη συγκεκριμένη κατανομή (μ=26, σ=3) υπάρχει σχεδόν 100% πιθανότητα ο μέσος του τυχαίου δείγματος να πέσει μέσα στα όρια του διαστήματος 25–27.

34

Περίπτωση 2η – Πιθανότητα εμφάνισης ενός μέσου στο συμμετρικό διάστημα μ±0.5 μονάδες.

Συγκεκριμένα ποια είναι η πιθανότητα ο μέσος του δείγματος N=100 να πέσει στο διάστημα 25.5–26.5, δηλαδή να έχει τιμή $25.5 \leq X \leq 26.5$;

Η πιθανότητα αυτή είναι $p(z_1 - z_2) = p(z_1) - p(z_2)$

και επειδή ο χώρος κάτω από την κανονική δειγματική καμπύλη είναι στην περίπτωση αυτή συμμετρικός $|z_1| = z_2$ και $p = 2p(z_1)$.

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{N}} = \frac{3}{\sqrt{100}} = \frac{3}{10} = 0.30$$

Οι κρίσιμες τιμές 25.5 & 26.5 έχουν τυπικές τιμές z

$$z_1 = \frac{\overline{X}_1 - \mu}{\sigma_{\overline{X}}} = \frac{25.5 - 26}{0.30} = \frac{-0.5}{0.30} = -1.67$$

$$z_2 = \frac{\overline{X}_2 - \mu}{\sigma_{\overline{X}}} = \frac{26.5 - 26}{0.30} = \frac{0.5}{0.30} = +1.67$$

Από το παράρτημα Γ διαπιστώνουμε ότι ο χώρος κάτω από την τυπική κανονική καμπύλη μεταξύ z=1.67 και 0 είναι ίσος με 0.45254.

Έτσι, η πιθανότητα που ζητάμε είναι

$$p = 2p(z_1) = 2(0.45254) = 0.90508.$$

Αυτό σημαίνει ότι με μέγεθος δείγματος N=100 για τη συγκεκριμένη κατανομή (μ=26, σ=3) υπάρχει σχεδόν 91% πιθανότητα ο μέσος του τυχαίου δείγματος να πέσει μέσα στα όρια του διαστήματος 25.5–26.5.

35

Παράδειγμα 9.5 - Εύρεση μέσου με πιθανότητες p < 0.05 & p < 0.01.

- Από την ίδια πληθυσμιακή κατανομή (μ=26, σ=3) παίρνουμε ένα τυχαίο δείγμα N=50 ατόμων.
- Διατηρούμε και πάλι το δείγμα αυτό σταθερό και εξετάζουμε τον υπολογισμό της μέσης τιμής δείγματος που έχει πιθανότητα εμφάνισης **ίση** ή μικρότερη από 2 κλασικές περιπτώσεις πιθανότητας τις οποίες συναντάμε κατά κανόνα στον έλεγχο της στατιστικής σημαντικότητας: $p<0.05$ & $p<0.01$.
- Για να επιλυθεί το πρόβλημα αυτό, υπολογίζουμε το τυπικό σφάλμα του μέσου, βρίσκουμε τις τιμές z που αντιστοιχούν στις πιθανότητες αυτές και μετά υπολογίζουμε τις αντίστοιχες τιμές των δειγμάτων με το γνωστό τύπο του z.

36

Περίπτωση 1η - Ποιος μέσος έχει πιθανότητα εμφάνισης $p \leq 0.05$;

Το τυπικό σφάλμα του μέσου είναι

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{N}} = \frac{3}{\sqrt{50}} = \frac{3}{7.07} = 0.424$$

Από το παράρτημα Γ βρίσκουμε ότι η τιμή z που αντιστοιχεί στον χώρο κάτω από την καμπύλη 0.05, δηλαδή στην πιθανότητα $p \leq 0.05$, είναι z=1.645.

Έτσι, από τον τύπο του z έχουμε

$$\overline{X} = z\,\sigma_{\overline{X}} + \mu = (1.645)(0.424) + 26 = 0.697 + 26 = 26.697 \approx 26.70$$

Περίπτωση 2η - Πιθανότητα $p \leq 0.01$

Το τυπικό σφάλμα του μέσου υπολογίστηκε στην 1η περίπτωση 0.424.

Από το παράρτημα Γ βρίσκουμε ότι η τιμή z που αντιστοιχεί στον χώρο 0.01 της καμπύλης, δηλαδή στην πιθανότητα $p \leq 0.01$, είναι z=2.325.

Έτσι, από τον τύπο του z έχουμε

$$\overline{X} = z\,\sigma_{\overline{X}} + \mu = (2.325)(0.424) + 26 = 0.986 + 26 = 26.986 \approx 26.99$$

37

Επίπεδο Σημαντικότητας

- **Οι δύο αυτές κλασικές περιπτώσεις πιθανότητας ($p < 0.05$ & $p < 0.01$) χρησιμοποιούνται πολύ για τον έλεγχο της σημαντικότητας των στατιστικών δεικτών και κριτηρίων**.
- Προς το παρόν αρκεί να δοθεί η απλή εξήγηση ότι στις πειραματικές έρευνες τίθεται εξαρχής η οριακή πιθανότητα *5%* και μερικές φορές 1% η μέση τιμή του τυχαίου δείγματος που παίρνεται από ένα γεννήτορα πληθυσμό να μην οφείλεται στην πειραματική αγωγή ή συνθήκη αλλά μόνο σε τυχαία σύμπτωση.

38

38

Επίπεδο Σημαντικότητας

- Κατά συνέπεια, όταν σε ένα πείραμα προκύψει μια μέση τιμή της οποίας η αντίστοιχη τιμή z απέχει πολύ από τον πληθυσμιακό μέσο (μ), όπως για παράδειγμα στον ακραίο χώρο πιθανότητας <0.05, τότε, επειδή ακριβώς η πιθανότητα εμφάνισης μιας τέτοιας μέσης τιμής από σύμπτωση και μόνο είναι πολύ μικρή, θεωρούμε ότι τελικά η εμφάνισή της οφείλεται στην πειραματική αγωγή ή στην επιβληθείσα πειραματική συνθήκη κ.λπ.

- Με τον τρόπο αυτό, το πειραματικό μας αποτέλεσμα λαμβάνεται ως στατιστικώς σημαντικό για το επίπεδο σημαντικότητας, που θέσαμε πριν τη διεξαγωγή του πειράματος, σύμφωνα και με τη διατύπωση της αντίστοιχης στατιστικής υπόθεσης.

39

39

Δειγματική Κατανομή Διαφορών

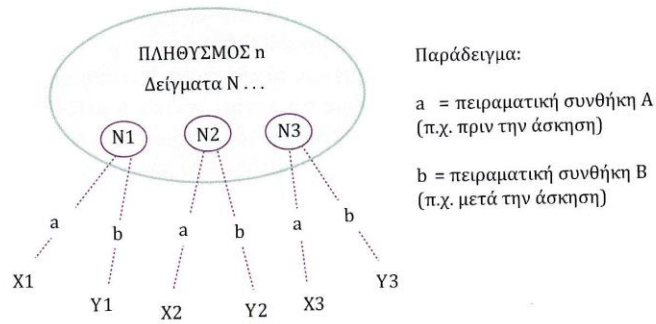
- Όπως εξετάζουμε την κατανομή των μέσων ισοπληθών δειγμάτων ενός πληθυσμού έτσι μπορούμε να εξετάσουμε και δειγματικές κατανομές διαφορών ή αθροισμάτων οποιουδήποτε άλλου δείκτη θέσης ή διασποράς.
- Ας υποθέσουμε ότι έχουμε τους πληθυσμούς n_x και n_Y των **ομοιογενών** (συγκρίσιμων) μεταβλητών X & Y, αντίστοιχα. Ο πληθυσμός της μεταβλητής X έχει μέσο μ_X και τυπική απόκλιση σ_χ και ο πληθυσμός της μεταβλητής Y έχει μέσο μ_Y και τυπική απόκλιση σ_Y.
- Από κάθε πληθυσμό παίρνουμε k συνεχή τυχαία δείγματα πλήθους N_X και N_Y αντίστοιχα και λογαριάζουμε τις μέσες τιμές και τις διασπορές τους.

40

40

Εξαρτημένοι Πληθυσμοί Χ & Υ

- Στην κατηγορία αυτή ανήκουν πληθυσμιακά σύνολα αποτελούμενα από ίδια άτομα (ή γενικά υποκείμενα) που εξετάζονται κάτω από δύο διαφορετικές πειραματικές συνθήκες ή καταστάσεις (εικ. 9.2).



Σχήμα 9.2 - Συμβολική απεικόνιση εξαρτημένων δειγμάτων ($\rho \neq 0$)

41

41

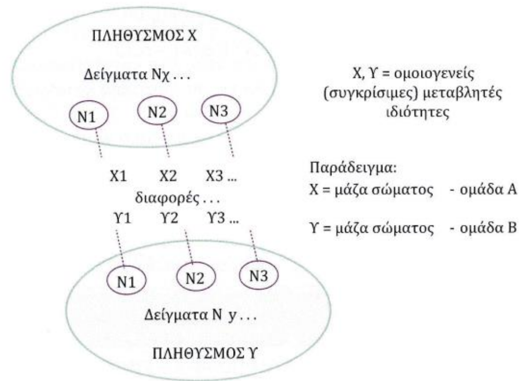
Εξαρτημένοι Πληθυσμοί Χ & Υ … συνέχεια

- Στην ερευνητική διαδικασία εφαρμόζεται συχνά η εξέταση της ίδιας ομάδας ατόμων κάτω από 2 ή περισσότερες συνθήκες, όπως π.χ. *πριν* και *μετά* από κάποια πειραματική αγωγή (π.χ. προπονητικό πρόγραμμα).
  - Είναι εύλογο, λοιπόν, να θεωρούμε ότι μεταξύ της μεταβλητής Χ, που μετρήθηκε στην **1η** , ας πούμε, πειραματική συνθήκη, και της ομοιογενούς (στην ουσία ίδιας) μεταβλητής Υ, που μετρήθηκε στη **2η πειραματική συνθήκη**, θα υπάρχει κάποιος βαθμός σχέσης. Με απλά λόγια αναμένουμε οι τιμές Υ να είναι υψηλές, μέτριες, χαμηλές ανάλογα ή αντιστρόφως ανάλογα με τις αντίστοιχες ατομικές τιμές Χ.
  - Αυτή η τάση θεωρείται δεδομένη για εξαρτημένους πληθυσμούς και φανερώνει την ύπαρξη μεταξύ των μεταβλητών Χ & Υ **συσχέτισης**, που σε πληθυσμιακό επίπεδο συμβολίζεται με το ρ.

42

Ανεξάρτητοι Πληθυσμοί X & Y

- Στην κατηγορία αυτή ανήκουν πληθυσμοί που αποτελούνται από διαφορετικά μέλη και εξετάζονται κάτω από ίδιες πειραματικές συνθήκες (εικ. 9.3).
- Τα δείγματα δύο ανεξάρτητων πληθυσμών X & Y δεν είναι απαραίτητο να είναι ισοπληθή, όπως γίνεται με τους εξαρτημένους πληθυσμούς.



Σχήμα 9.3 - Συμβολική απεικόνιση ανεξάρτητων δειγμάτων (ρ = 0)

43

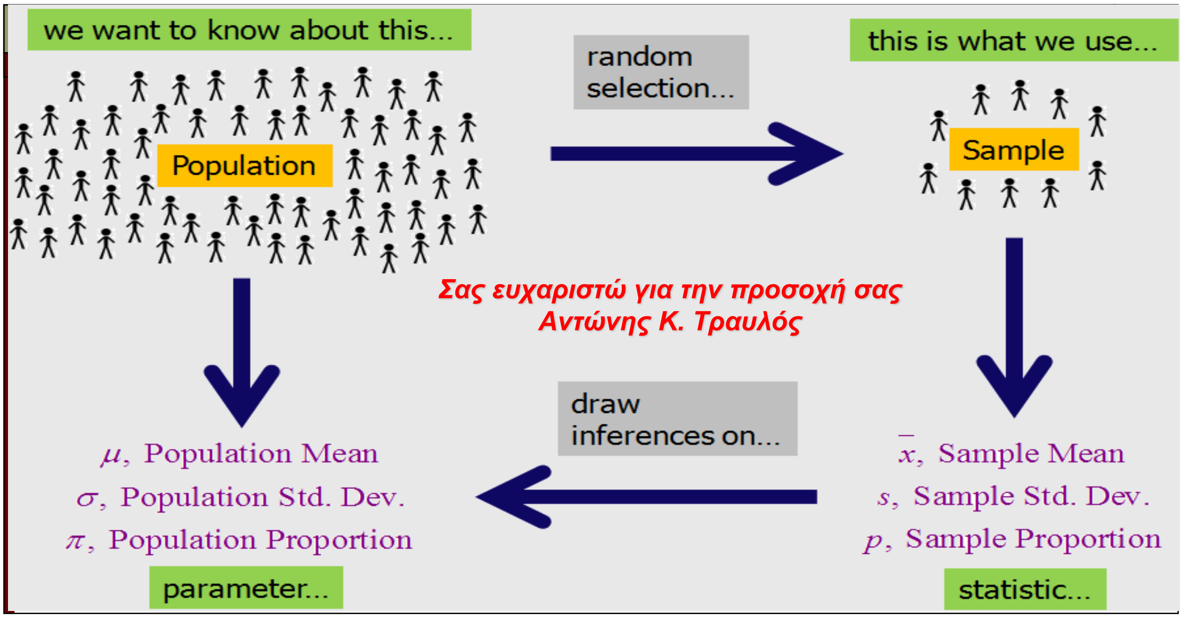
43

Π.χ. 9.8 - Πιθανότητα διαφοράς μέσων: ευκαμψία Α-ετών και Γ-ετών σπουδαστών

- Η μέση τιμή και η τυπική απόκλιση της ευκαμψίας κορμού των Α-ετών σπουδαστών μιας παραγωγικής σχολής είναι μ_X=25 cm και σ_χ=8 cm, αντίστοιχα, ενώ των Γ-ετών σπουδαστών είναι μ_Y=27 και σ_Y=7, αντίστοιχα.
  - Οι κατανομές των επιδόσεων στην ευκαμψία κορμού είναι κανονικές και για τους δύο υποπληθυσμούς σπουδαστών.
  - Υποθέστε λοιπόν ότι από τους δύο αυτούς υποπληθυσμούς σπουδαστών σκοπεύουμε να πάρουμε με τυχαία δειγματοληψία N_X=30 Α-ετείς και N_Y=30 Γ-ετείς σπουδαστές.
- Ένα πρακτικά χρήσιμο και ενδιαφέρον ερώτημα που μπορεί να τεθεί στην περίπτωση αυτή είναι ***ποια είναι η πιθανότητα το δείγμα, των Α-ετών σπουδαστών (X) να παρουσιάσει μέσο ίσο ή μεγαλύτερο από το μέσο του δείγματος των Γ-ετών σπουδαστών (Y);***

44

44


we want to know about this...
Population
random selection...
this is what we use...
Sample
Σας ευχαριστώ για την προσοχή σας
Αντώνης Κ. Τραυλός
μ, Population Mean
σ, Population Std. Dev.
π, Population Proportion
parameter...
draw inferences on...
$\bar{x}$, Sample Mean
s, Sample Std. Dev.
p, Sample Proportion
statistic...
45

45