Ε Λ Λ Η Ν Ι Κ Η Δ Η Μ Ο Κ Ρ Α Τ Ι Α
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ
UNIVERSITY OF PATRAS
Καθηγητής Ι. Μητρόπουλος
ΤΜΗΜΑ ΔΙΟΙΚΗΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ & ΤΕΧΝΟΛΟΓΙΑΣ
Τηλ.: +030 2610 369213, email: imitro@upatras.gr
Διεύθυνση: Μεγάλου Αλεξάνδρου 1, 263 34 ΠΑΤΡΑ

Ποσοτικές Μέθοδοι στην Οικονομία και Διοίκηση (Ι)

ΕΝΟΤΗΤΑ 8η: Κατανομές Δειγματοληψίας

Στόχοι

Σε αυτό το κεφάλαιο μαθαίνετε:

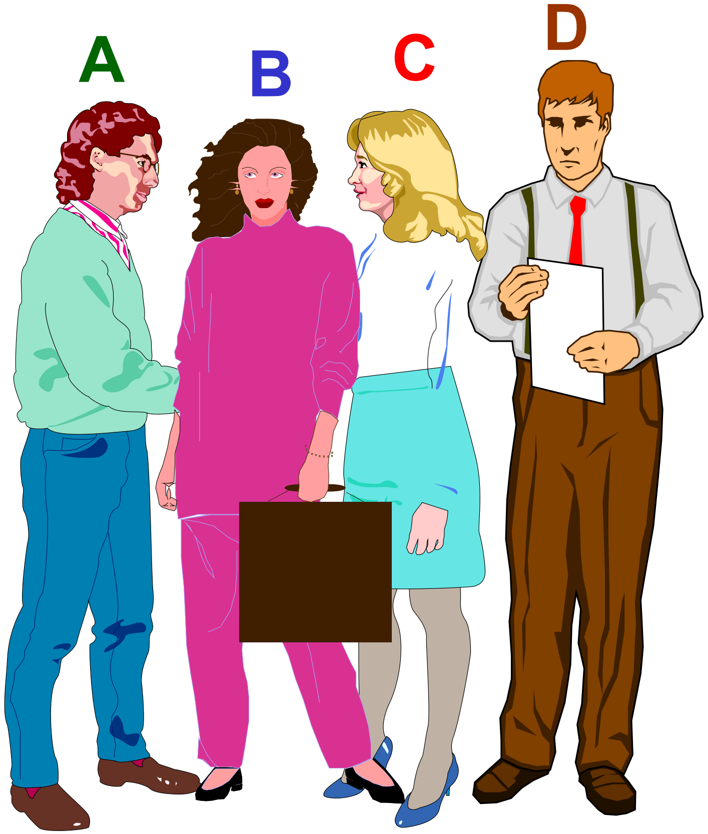
- Την έννοια της κατανομής δειγματοληψίας
- Τον υπολογισμό πιθανοτήτων που σχετίζονται με τον δειγματικό μέσο και το δειγματικό ποσοστό
- Την σπουδαιότητα του Κεντρικού Οριακού Θεωρήματος

Κατανομές Δειγματοληψίας

- Μια κατανομή δειγματοληψίας είναι μια κατανομή όλων των πιθανών τιμών ενός δειγματικού στατιστικού μέτρου για ένα δεδομένο μέγεθος δείγματος που επιλέγεται από έναν πληθυσμό.
- Για παράδειγμα, υποθέστε οτι πραγματοποιείται δειγματοληψία 50 φοιτητών από το κολέγιο σας όσον αφορά το μέσο όρο της βαθμολογίας τους (GPA). Αν λάβατε πολλά διαφορετικά δείγματα μεγέθους 50, υπολογίζετε διαφορετικό μέσο όρο για κάθε δείγμα. Μας ενδιαφέρει η κατανομή όλων των πιθανών μέσων όρων GPA που μπορούμε να υπολογίσουμε για οποιοδήποτε δείγμα 50 φοιτητών.

Ανάπτυξη Κατανομής Δειγματοληψίας

- **Υποθέστε οτι υπάρχει ένας πληθυσμός …**
- Μέγεθος πληθυσμού N=4
- Τυχαία μεταβλητή, X, είναι η ηλικία των ατόμων
- Τιμές του X: 18, 20, 22, 24 (έτη)



Ανάπτυξη Κατανομής Δειγματοληψίας

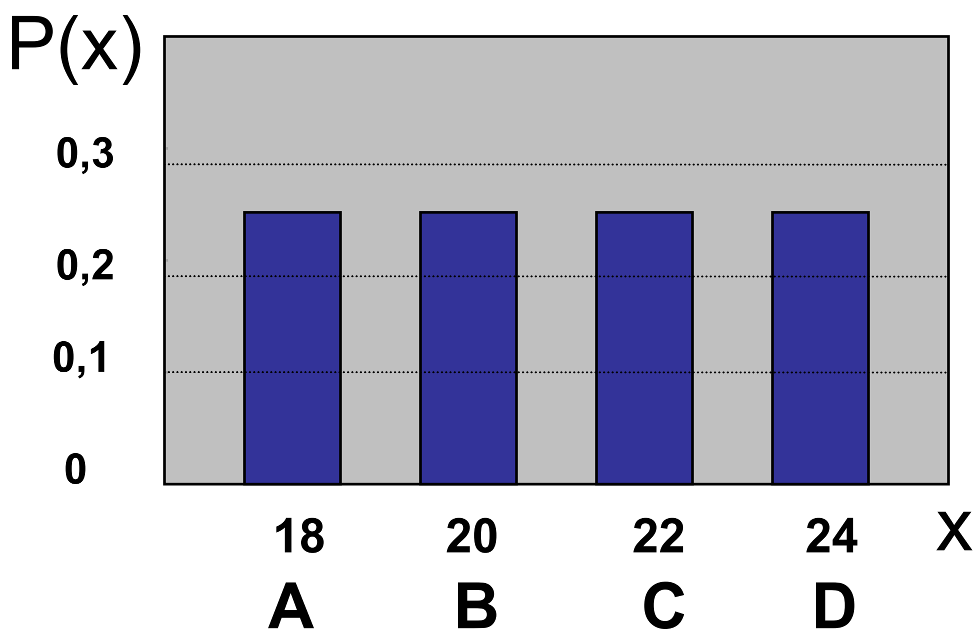
(συνέχεια)

Στατιστικά Μέτρα για την Κατανομή Πληθυσμού:

$$\mu = \frac{\sum X_i}{N}$$

$$= \frac{18+20+22+24}{4} = 21$$

$$\sigma = \sqrt{\frac{\sum (X_i - \mu)^2}{N}} = 2.236$$



Ομοιόμορφη Κατανομή

Ανάπτυξη Κατανομής Δειγματοληψίας

(συνέχεια)

Τώρα υπολογίστε όλα τα πιθανά δείγματα μεγέθους n=2

1st Παρ.	2nd Παρατήρηση			
	18	20	22	24
18	18,18	18,20	18,22	18,24
20	20,18	20,20	20,22	20,24
22	22,18	22,20	22,22	22,24
24	24,18	24,20	24,22	24,24

16 πιθανά δείγματα (δειγματοληψία με επανατοποθέτηση)

16 Δειγματικοί Μέσοι

1st Παρ	2nd Παρατήρηση			
	18	20	22	24
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

Ανάπτυξη Κατανομής Δειγματοληψίας

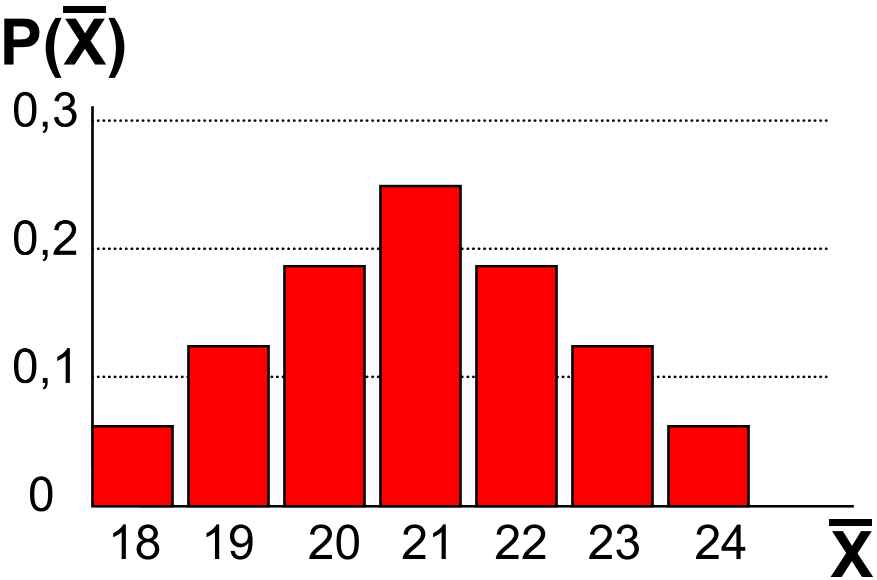
(συνέχεια)

Κατανομή Δειγματοληψίας Όλων των Δειγματικών Μέσων

16 Δειγματικοί Μέσοι

1st Παρ	2nd Παρατήρηση 18	20	22	24
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

Κατανομή Δειγματικών Μέσων



(δεν είναι πλέον ομοιόμορφη)

Ανάπτυξη Κατανομής Δειγματοληψίας

(συνέχεια)

Στατιστικά Μέτρα αυτής της Κατανομής Δειγματοληψίας:

$$\mu_{\bar{x}} = \frac{18+19+19+\cdots+24}{16} = 21$$

$$\sigma_{\bar{x}} = \sqrt{\frac{(18-21)^2+(19-21)^2+\cdots+(24-21)^2}{16}} = 1.58$$

Σημείωση: Εδώ διαιρούμε με το 16 επειδή υπάρχουν 16 διαφορετικά δείγματα μεγέθους 2.

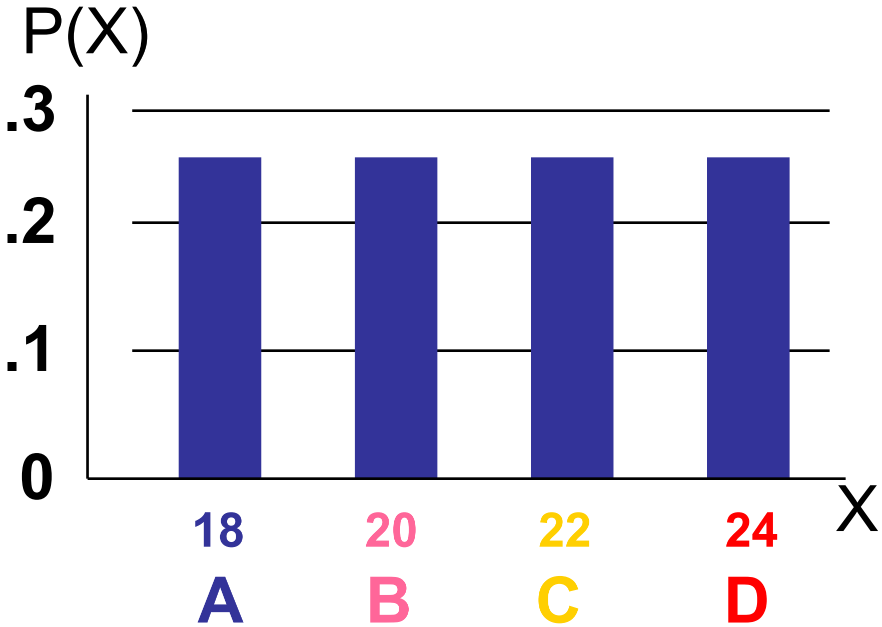
Σύγκριση Της Κατανομής Του Πληθυσμού Με Την Κατανομή Των Δειγματικών Μέσων

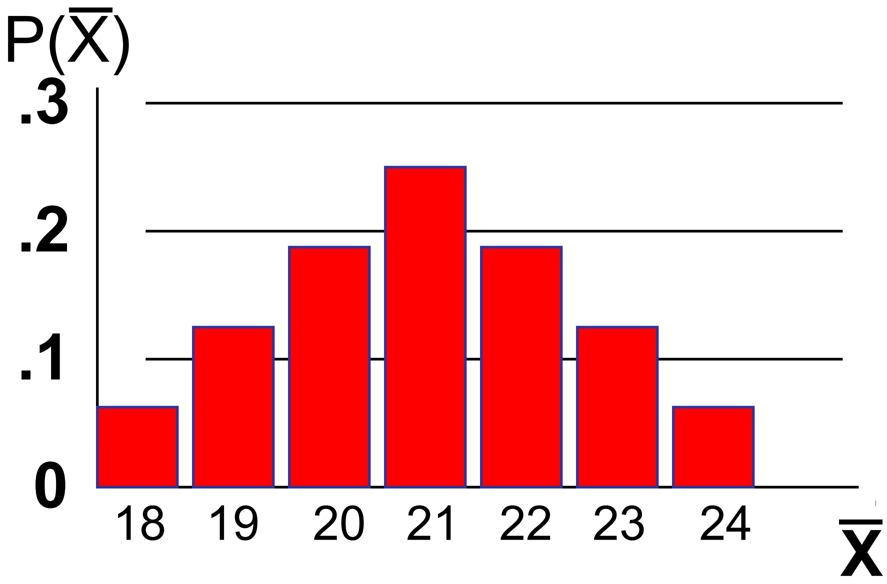
Πληθυσμός
N = 4

$\mu = 21 \qquad \sigma = 2.236$

Κατανομή Δειγματικών Μέσων
n = 2

$\mu_{\bar{X}} = 21 \qquad \sigma_{\bar{X}} = 1.58$





Κατανομή Δειγματοληψίας του Δειγματικού Μέσου: Τυπικό Σφάλμα Του Μέσου

- Διαφορετικά δείγματα του ιδίου μεγέθους από τον ίδιο πληθυσμό θα δώσουν διαφορετικούς δειγματικούς μέσους
- Ένα μέτρο της μεταβλητότητας του μέσου όρου από δείγμα σε δείγμα δίνεται από το Τυπικό Σφάλμα του Μέσου:

  (Αυτό προυποθέτει οτι η δειγματοληψία είναι με επανατοποθέτηση ή οτι η δειγματοληψία είναι χωρίς επανατοποθέτηση από έναν άπειρο πληθυσμό)

$$\boxed{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}}$$

- Σημειώστε οτι το τυπικό σφάλμα του μέσου μειώνεται όσο αυξάνεται το μέγεθος του δείγματος

Κατανομή Δειγματοληψίας του Δειγματικού Μέσου: Αν ο Πληθυσμός είναι Κανονικός

- Αν ο πληθυσμός είναι κανονικός με μέσο όρο μ και τυπική απόκλιση σ, η κατανομή δειγματοληψίας του $\bar{X}$ είναι επίσης κανονικά κατανεμημένη με

$$\boxed{\mu_{\bar{X}} = \mu} \quad \text{και} \quad \boxed{\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}}$$

Z-τιμή για την Κατανομή Δειγματοληψίας του Μέσου

- Z-τιμή για την κατανομή δειγματοληψίας του $\bar{X}$:

$$Z = \frac{(\bar{X} - \mu_{\bar{X}})}{\sigma_{\bar{X}}} = \frac{(\bar{X} - \mu)}{\frac{\sigma}{\sqrt{n}}}$$

όπου:

$\bar{X}$ = δειγματικός μέσος

μ = μέσος όρος πληθυσμού

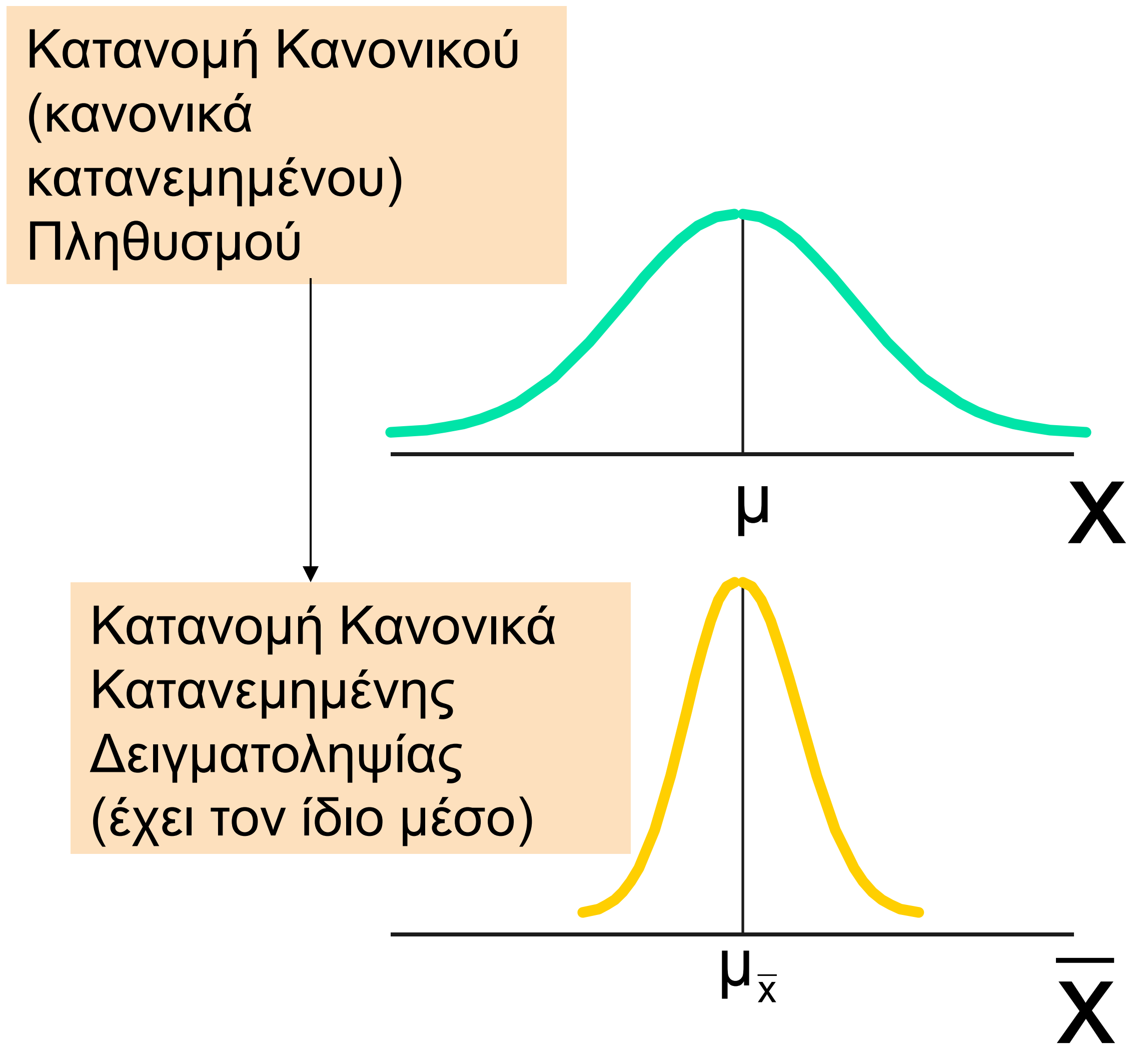
σ = τυπική απόκλιση πληθυσμού

n = μέγεθος δείγματος

Ιδιότητες Κατανομής Δειγματοληψίας

$$\mu_{\bar{x}} = \mu$$

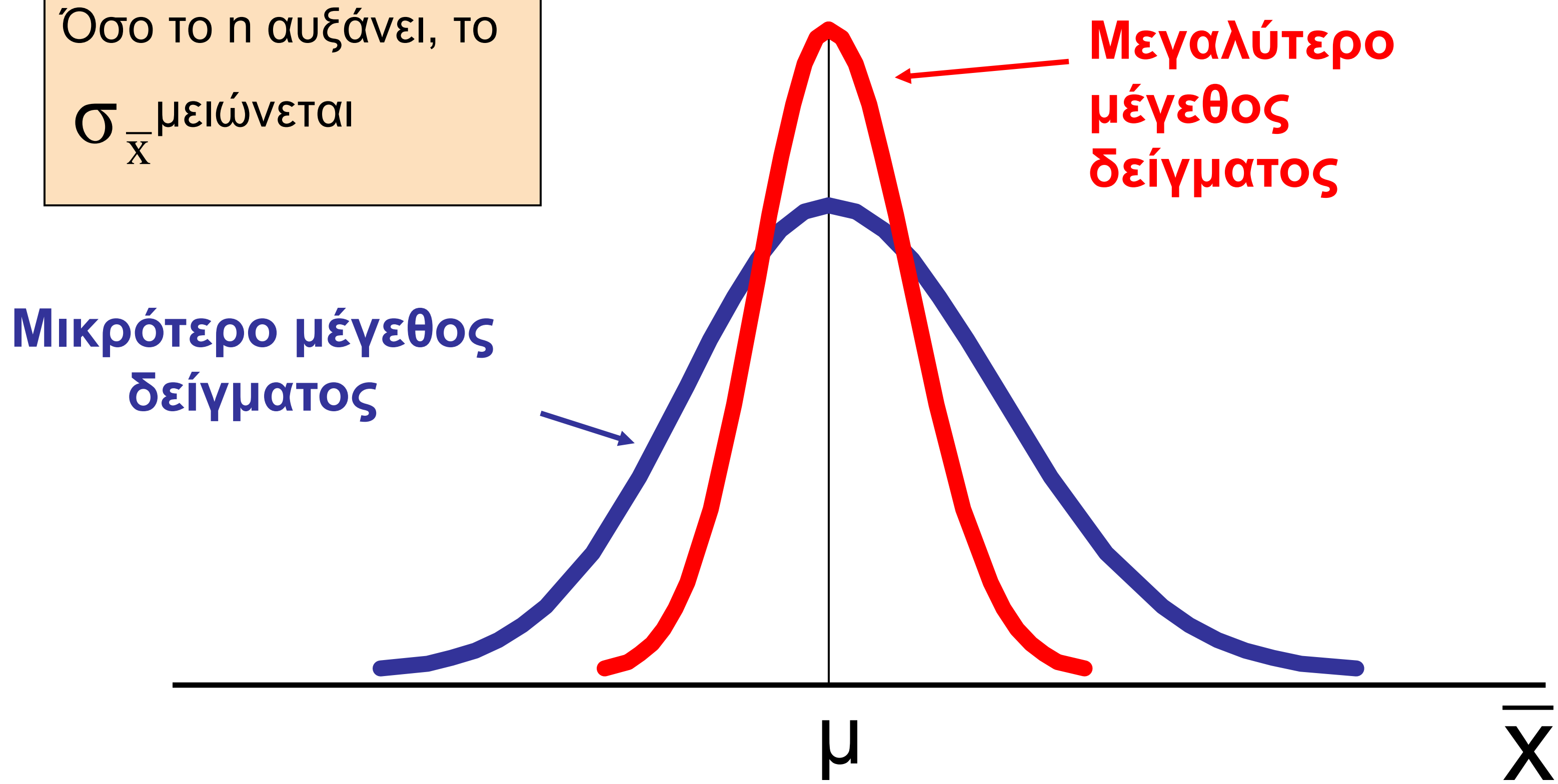
(δηλ. το $\bar{X}$ είναι αμερόληπτο)



Ιδιότητες Κατανομής Δειγματοληψίας

(συνέχεια)

Όσο το n αυξάνει, το $\sigma_{\bar{x}}$ μειώνεται



Προσδιορισμός του Διαστήματος που Περιλαμβάνει Ένα Σταθερό Ποσοστό Των Δειγματικών Μέσων

Βρείτε ένα συμμετρικά κατανεμημένο διάστημα γύρω από το μ που περιλαμβάνει το 95% των δειγματικών μέσων όταν $\mu = 368$, $\sigma = 15$, και $n = 25$.

- Αφού το διάστημα περιλαμβάνει το 95% των δειγματικών μέσων, το 5% των δειγματικών μέσων θα βρίσκεται εκτός διαστήματος
- Αφού το διάστημα είναι συμμετρικό το 2,5% θα είναι πάνω από το άνω άκρο και το 2,5% θα είναι κάτω από το κάτω άκρο.
- Από τον τυποποιημένο κανονικό πίνακα, η Z-score (τιμή) με 2,5% (0,0250) κάτω είναι -1,96 και η Z-score (τιμή) με 2,5% (0,0250) πάνω είναι 1,96.

Προσδιορισμός του Διαστήματος που Περιλαμβάνει Ένα Σταθερό Ποσοστό Των Δειγματικών Μέσων

(συνέχεια)

- Υπολογισμός του κάτω άκρου του διαστήματος

$$\bar{X}_L = \mu + Z\frac{\sigma}{\sqrt{n}} = 368 + (-1.96)\frac{15}{\sqrt{25}} = 362.12$$

- Υπολογισμός του άνω άκρου του διαστήματος

$$\bar{X}_U = \mu + Z\frac{\sigma}{\sqrt{n}} = 368 + (1.96)\frac{15}{\sqrt{25}} = 373.88$$

- Το 95% όλων των δειγματικών μέσων του μεγέθους δείγματος 25 είναι μεταξύ του 362,12 και του 373,88

Κατανομή Δειγματοληψίας του Δειγματικού Μέσου: Αν ο Πληθυσμός δεν είναι Κανονικός (Κανονικά Κατανεμημένος)

- Μπορούμε να εφαρμόσουμε το Κεντρικό Οριακό Θεώρημα:
  - Ακόμα και αν ο πληθυσμός δεν είναι κανονικός,
  - …οι δειγματικοί μέσοι από τον πληθυσμό θα είναι κατά προσέγγιση κανονικά κατανεμημένοι εφόσον το μέγεθος του δείγματος είναι αρκετά μεγάλο.

Ιδιότητες της κατανομής δειγματοληψίας:

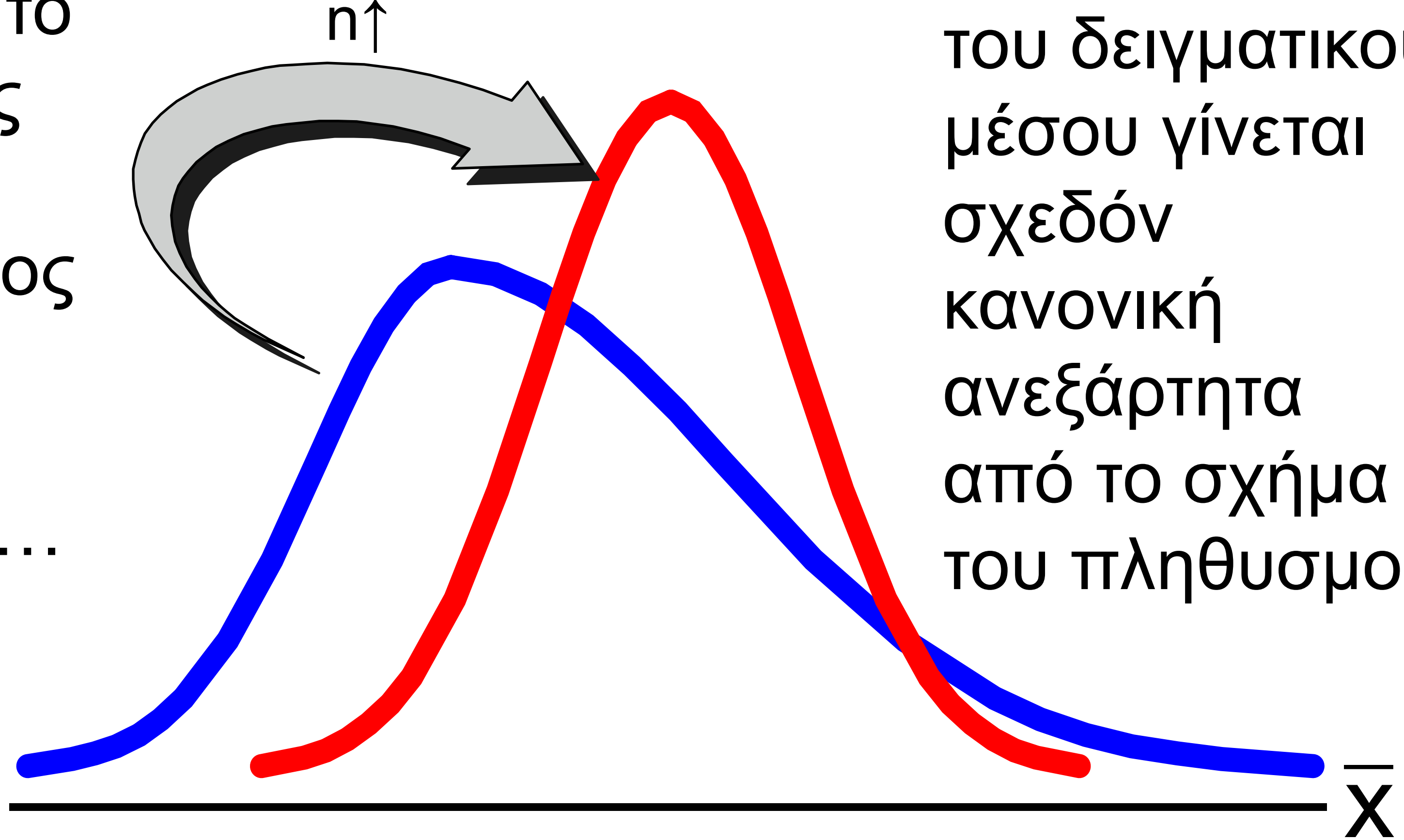
$$\boxed{\mu_{\bar{x}} = \mu}$$

και

$$\boxed{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}}$$

Κεντρικό Οριακό Θεώρημα

Καθώς το μέγεθος του δείγματος γίνεται αρκετά μεγάλο…



Η κατανομή δειγματοληψίας του δειγματικού μέσου γίνεται σχεδόν κανονική ανεξάρτητα από το σχήμα του πληθυσμού

Είδαμε …

- **ΚΕΝΤΡΙΚΟ ΟΡΙΑΚΟ ΘΕΩΡΗΜΑ:**

Η κατανομή του δειγματικού μέσου σε τυχαίο δείγμα κάθε πληθυσμού ακολουθεί κατά προσέγγιση την κανονική κατανομή. Όσο μεγαλύτερο το μέγεθος του δείγματος τόσο περισσότερο η κατανομή του προσεγγίζει την κανονική κατανομή.

Κατανομή δειγματικού μέσου (χαρακτηριστικά):

$$\mu_{\bar{x}} = \mu$$

$$\sigma_{\bar{x}}^2 = \sigma^2 / n \quad and \quad \sigma_{\bar{x}} = \sigma / \sqrt{n}$$

Η κατανομή του δειγματικού μέσου μπορεί να εκφραστεί ως: $Z = \dfrac{\overline{X} - \mu}{\sigma / \sqrt{n}}$

Αναλογίες (ποσοστά)

$$E(\hat{P}) = p$$

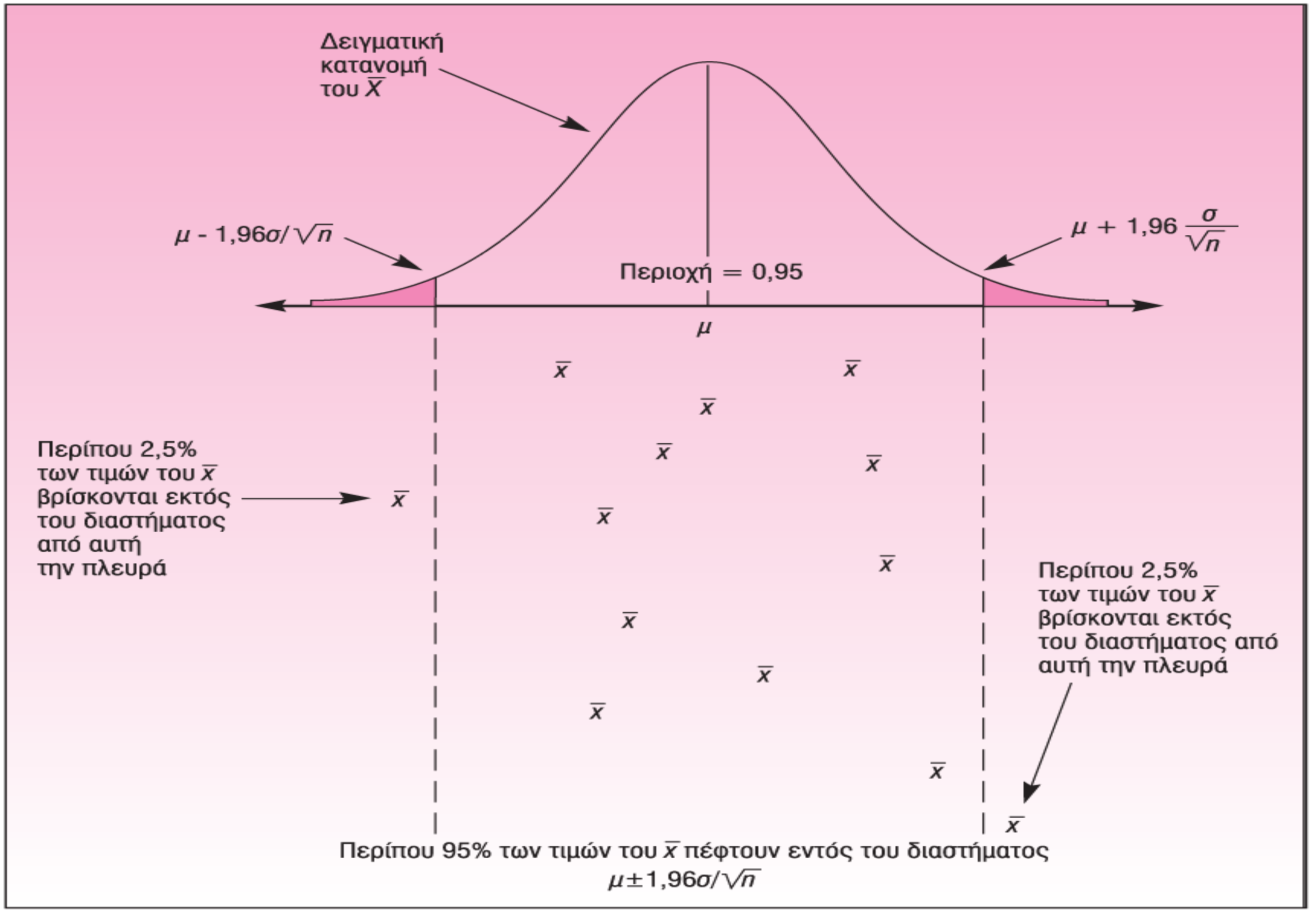
$$V(\hat{P}) = \sigma_{\hat{p}}^2 = \frac{p(1-p)}{n}$$

$$\sigma_{\hat{p}} = \sqrt{p(1-p)/n}$$

$$Z = \frac{\hat{P} - p}{\sqrt{p(1-p)/n}}$$

Που θα φτάσουμε…

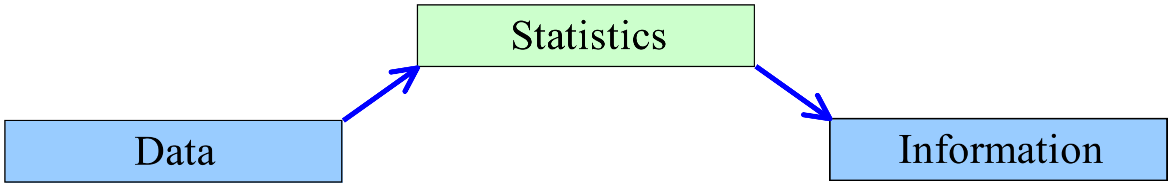
Ωστόσο, σε όλες σχεδόν τις ρεαλιστικές καταστάσεις πληθυσμιακοί παράμετροι είναι άγνωστοι. Θα χρησιμοποιήσουμε τη κατανομή δειγματοληψίας για να εξαγάγουμε συμπεράσματα σχετικά με τις άγνωστες πληθυσμιακές παραμέτρους.

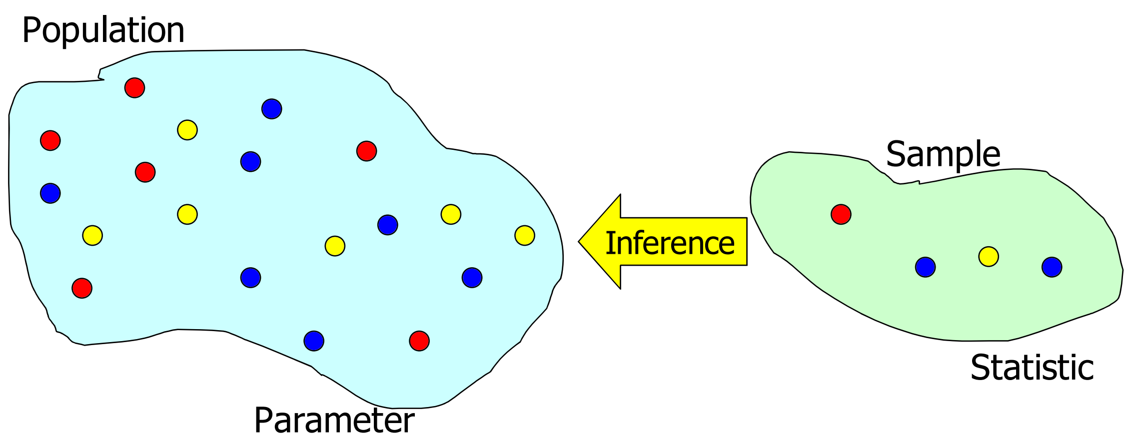


Διάγραμμα 6-1 Κατανομή πιθανοτήτων του $\bar{X}$ και κάποιες τιμές του στατιστικού σε επαναλαμβανόμενες δειγματοληψίες.

Επαγωγική Στατιστική (*Statistical Inference*)…

Στατιστική Συμπερασματολογία (inference) *είναι η διαδικασία με την οποία συλλέγουμε πληροφορίες και εξάγουμε συμπεράσματα σχετικά με πληθυσμούς μέσα από κατάλληλα δείγματα.*





Προκειμένου να γίνει συμπερασματολογία, χρειαζόμαστε τις δεξιότητες & γνώση της Περιγραφικής στατιστικής, κατανομές πιθανότητας, και κατανομές δειγματοληψίας.

Κατανομή Δειγματοληψίας Δειγματικού Μέσου: Αν ο Πληθυσμός δεν είναι Κανονικός (κανονικά κατανεμημένος)

(συνέχεια)

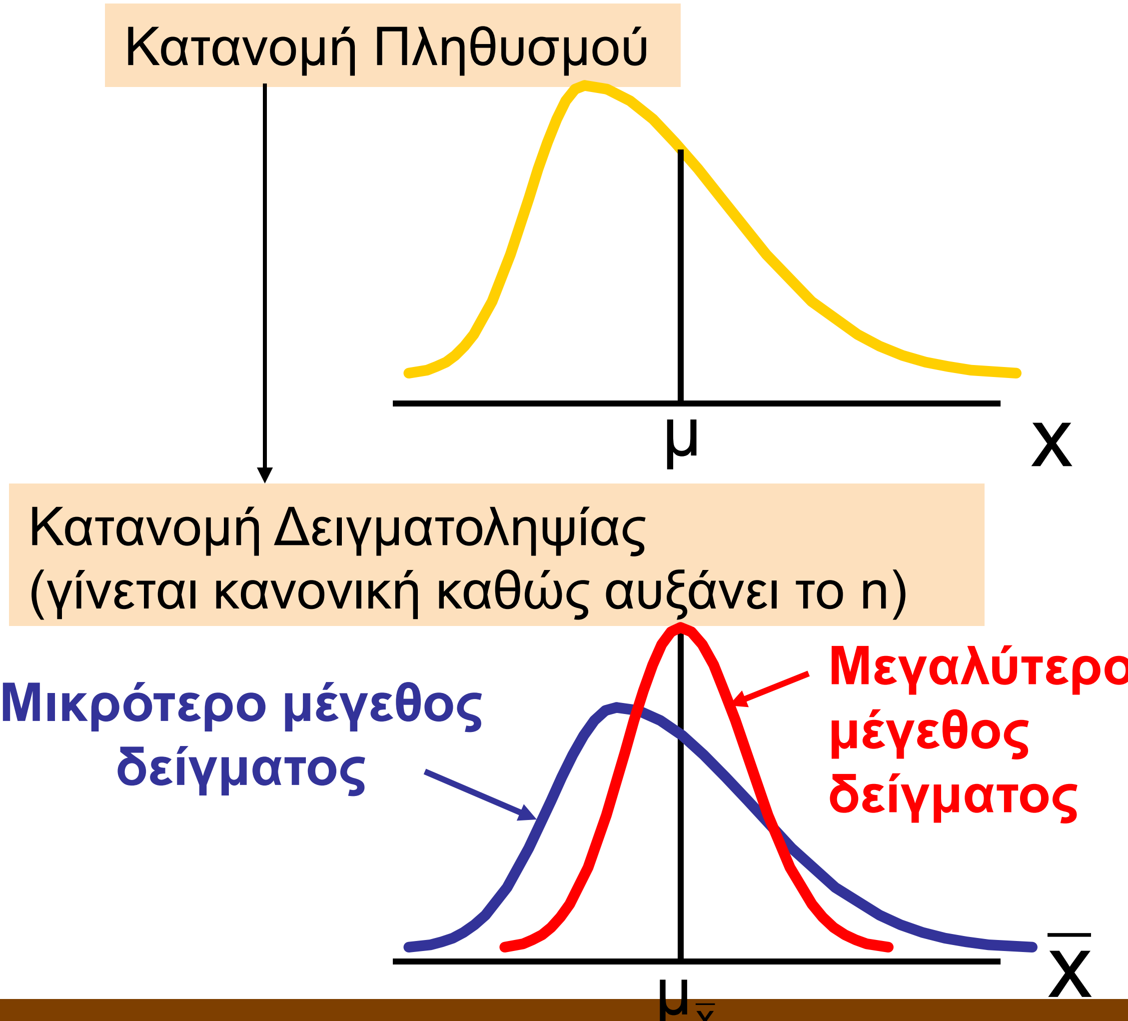
Ιδιότητες κατανομής δειγματοληψίας:

Κεντρική Τάση

$$\mu_{\bar{x}} = \mu$$

Μεταβλητότητα

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$



Πόσο Μεγάλο είναι το Αρκετά Μεγάλο;

- Για τις περισσότερες κατανομές, το $n > 30$ θα δώσει μια κατανομή δειγματοληψίας που είναι σχεδόν κανονική
- Για αρκετά συμμετρικές κατανομές, $n > 15$
- Για μια κατανομή κανονικού πληθυσμού, η κατανομή δειγματοληψίας του μέσου είναι πάντα κανονικά κατανεμημένη

Παράδειγμα

- Θεωρείστε έναν πληθυσμό με μέσο όρο $\mu = 8$ και τυπική απόκλιση $\sigma = 3$. Υποθέστε οτι επιλέγεται ένα τυχαίο δείγμα μεγέθους $n = 36$.
- Ποια είναι η πιθανότητα ο δειγματικός μέσος να είναι μεταξύ 7,8 και 8,2;

Παράδειγμα

(συνέχεια)

Λύση:

- Αν και ο πληθυσμός δεν είναι κανονικά κατανεμημένος, το κεντρικό οριακό θεώρημα μπορεί να χρησιμοποιηθεί ($n > 30$)
- … έτσι η κατανομή δειγματοληψίας του $\bar{X}$ είναι κατά προσέγγιση κανονική
- … με μέσο όρο $\mu_{\bar{X}} = 8$
- …και τυπική απόκλιση $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{36}} = 0.5$

ή τυπικό σφάλμα του μέσου (standard error)

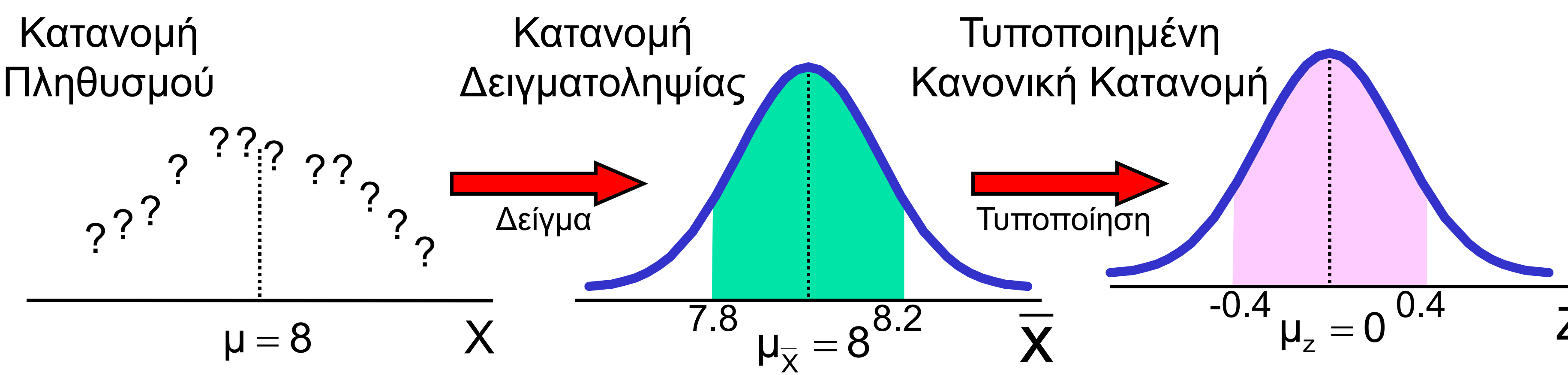
Παράδειγμα

(συνέχεια)

Λύση (συνέχεια):

$$P(7.8 < \bar{X} < 8.2) = P\left(\frac{7.8-8}{3/\sqrt{36}} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < \frac{8.2-8}{3/\sqrt{36}}\right)$$

$$= P(-0.4 < Z < 0.4) = 0.6554 - 0.3446 = \boxed{0.3108}$$



Ποσοστά Πληθυσμού

π = το ποσοστό του πληθυσμού που έχει κάποιο χαρακτηριστικό

- Ποσοστό δείγματος (p) παρέχει μια εκτίμηση του π:

$$p = \frac{X}{n} = \frac{\text{αριθμός στοιχείων στο δείγμα με τη υπό εξέταση χαρακτηριστικό}}{\text{μέγεθος δείγματος}}$$

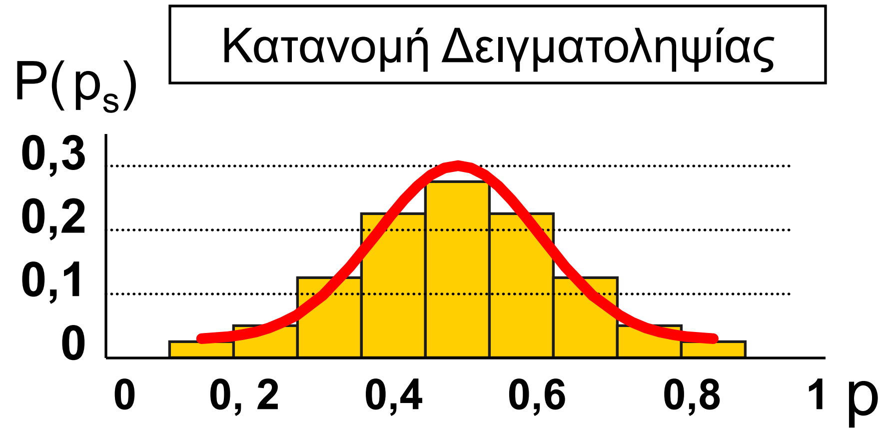
- $0 \leq p \leq 1$
- Το p είναι κατά προσέγγιση κατανεμημένο ως κανονική κατανομή όταν το n είναι μεγάλο

(υποθέτοντας δειγματοληψία με επανατοποθέτηση από έναν πεπερασμένο πληθυσμό ή χωρίς επανατοποθέτηση από έναν άπειρο πληθυσμό)

Κατανομή Δειγματοληψίας του p

- Προσέγγιση από κανονική κατανομή αν:
  -

$$n\pi \geq 5$$
και
$$n(1-\pi) \geq 5$$



όπου

$$\mu_p = \pi$$ και $$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$$

(όπου π = ποσοστό πληθυσμού)

Z-τιμή για Ποσοστά

Τυποποίηση του p σε μια Z τιμή με την βοήθεια του τύπου:

$$Z = \frac{p - \pi}{\sigma_p} = \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$$

Παράδειγμα

- Αν το πραγματικό ποσοστό των ψηφοφόρων που στηρίζει την Πρόταση Α είναι π = 0,4, ποια είναι η πιθανότητα ένα δείγμα μεγέθους 200 να αποφέρει ένα ποσοστό δείγματος μεταξύ 0,40 και 0,45;

- δηλ.: **αν π = 0,4 και n = 200, ποια είναι η**

  $P(0{,}40 \le p \le 0{,}45)$;

Παράδειγμα

(συνέχεια)

- **αν π = 0,4 και n = 200, ποια είναι η P(0,40 ≤ p ≤ 0,45) ;**

Βρίσκω σ_p : $\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0.4(1-0.4)}{200}} = 0.03464$

Μετατροπή σε τυποποιημένη κανονική: $P(0.40 \le p \le 0.45) = P\left(\frac{0.40-0.40}{0.03464} \le Z \le \frac{0.45-0.40}{0.03464}\right)$

$$= P(0 \le Z \le 1.44)$$

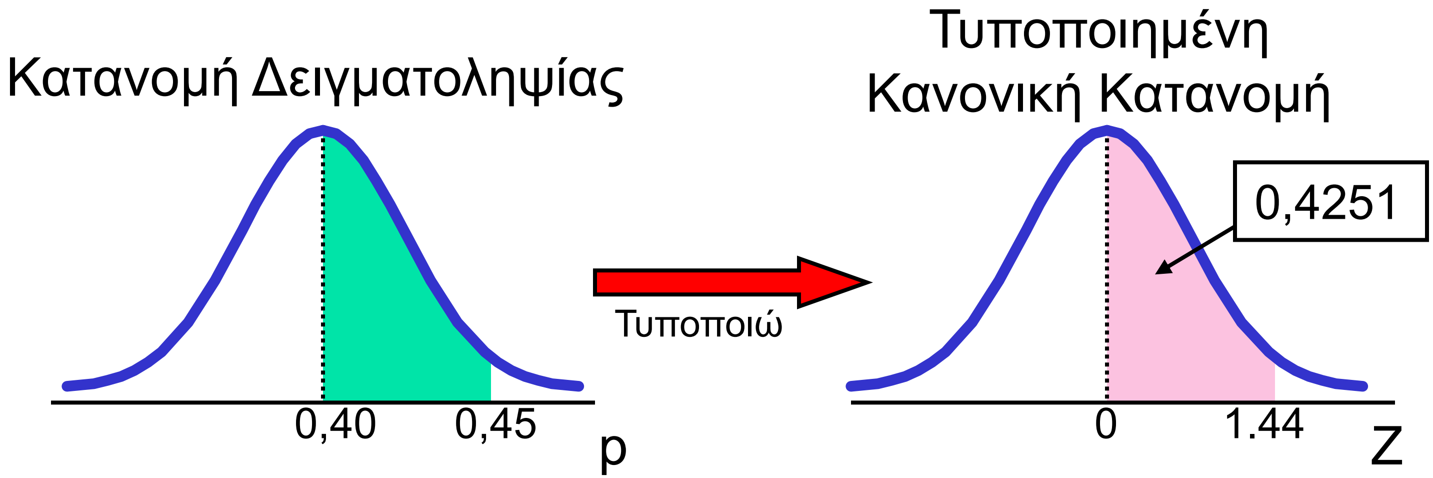
Παράδειγμα

(συνέχεια)

- **αν π = 0,4 και n = 200, ποια είναι η P(0,40 ≤ p ≤ 0,45) ;**

Χρησιμοποιώ τον αθροιστικό κανονικό πίνακα:

P(0 ≤ Z ≤ 1,44) = 0,9251 – 0,5000 = $\boxed{0{,}4251}$



Περίληψη Κεφαλαίου

Σε αυτό το κεφάλαιο αναλύσαμε:

- Την έννοια της κατανομής δειγματοληψίας
- Τον υπολογισμό πιθανοτήτων που σχετίζονται με τον δειγματικό μέσο και το δειγματικό ποσοστό
- Την σπουδαιότητα του Κεντρικού Οριακού Θεωρήματος