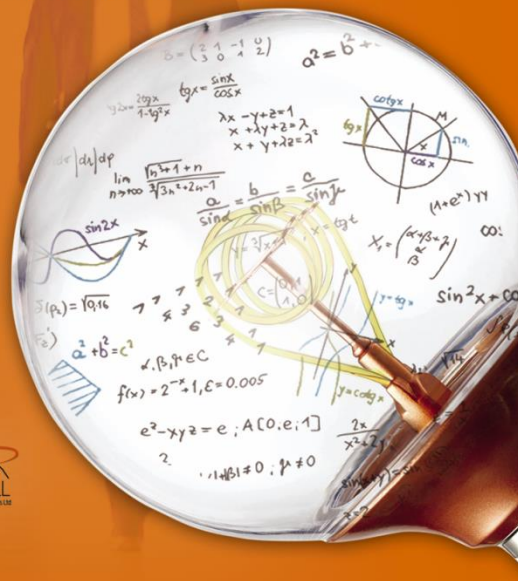


ΣΤΑΤΙΣΤΙΚΗ

Βασικές Αρχές με Έμφαση στην
ΟΙΚΟΝΟΜΙΑ και τις ΕΠΙΧΕΙΡΗΣΕΙΣ

Levine Szabat Stephan

Γενική Επιμέλεια - Πρόλογος
Στέφανος Γ. Γιακουμάτος
Στυλιανός Ι. Κουκούμιλος
Σταύρος Η. Αρβανίτης



Κεφάλαιο 8

Εκτίμηση Διαστήματος Εμπιστοσύνης

Στόχοι

- **Σε αυτό το κεφάλαιο, μαθαίνετε:**
- Την κατασκευή και ερμηνεία των εκτιμήσεων διαστημάτων εμπιστοσύνης για τον μέσο και το ποσοστό
- Τον καθορισμό του μεγέθους του δείγματος που είναι απαραίτητο στην κατασκευή εκτιμήσεων διαστημάτων εμπιστοσύνης για τον μέσο ή το ποσοστό

Συνοπτική περιγραφή κεφαλαίου

Περιεχόμενα κεφαλαίου

- Διαστήματα Εμπιστοσύνης για τον Μέσο όρο του πληθυσμού, μ
 - Όταν στην Τυπική Απόκλιση του πληθυσμού το σ είναι Γνωστό
 - Όταν στην Τυπική Απόκλιση του πληθυσμού το σ είναι Άγνωστο
- Διαστήματα Εμπιστοσύνης για το Ποσοστό του Πληθυσμού, π
- Προσδιορισμός του Απαιτούμενου Μεγέθους Δείγματος

Σημειακή Εκτίμηση και Εκτίμηση Εμπιστοσύνης

- Μία **σημειακή εκτίμηση** είναι ένας αριθμός,
- Ένα **διάστημα εμπιστοσύνης** παρέχει πρόσθετες πληροφορίες σχετικά με τη μεταβλητότητα της εκτίμησης



Σημειακές Εκτιμήσεις

Μπορούμε να εκτιμήσουμε μία παράμετρο πληθυσμού...		με ένα Στατιστικό Δείγμα (Μία Σημειακή Εκτίμηση)
Μέσος	μ	\bar{X}
Ποσοστό	π	p

Διαστήματα εμπιστοσύνης

- Πόση αβεβαιότητα συνδέεται με μια εκτίμηση σημείου μιας παραμέτρου πληθυσμού;
- Μία **εκτίμηση διαστήματος** παρέχει περισσότερες πληροφορίες σχετικά με ένα χαρακτηριστικό πληθυσμού από μια **σημειακή εκτίμηση**
- Αυτές οι εκτιμήσεις διαστήματος αποκαλούνται **διαστήματα εμπιστοσύνης**

Εκτίμηση Διαστήματος Εμπιστοσύνης

- Ένα διάστημα δίνει ένα εύρος τιμών:
 - Λαμβάνει υπόψη τη μεταβλητότητα των στατιστικών δειγμάτων από δείγμα σε δείγμα
 - Με βάση τις παρατηρήσεις από 1 δείγμα
 - Παρέχει πληροφορίες σχετικά με την εγγύτητα σε άγνωστες παραμέτρους του πληθυσμού
 - Καθορισμένη όσον αφορά το επίπεδο εμπιστοσύνης
 - π.χ. 95% εμπιστοσύνη, 99% εμπιστοσύνη
 - Ποτέ δεν μπορεί να είναι η εμπιστοσύνη 100%

Παράδειγμα Εκτίμησης Διαστήματος

Παράδειγμα συσκευασίας δημητριακών

Ο πληθυσμός έχει $\mu = 368$ and $\sigma = 15$.

- Αν πάρετε δείγμα μεγέθους $n = 25$ γνωρίζετε
 - $368 \pm 1,96 * 15 / \sqrt{25} = (362,12, 373,88)$. Το 95% των διαστημάτων που σχηματίζονται με αυτόν τον τρόπο θα περιέχουν μ .
 - Όταν δεν γνωρίζετε το μ , χρησιμοποιείτε \bar{X} για να εκτιμήσετε το μ
 - Αν $\bar{X} = 362,3$ το διάστημα είναι $362,3 \pm 1,96 * 15 / \sqrt{25} = (356,42, 368,18)$
 - Όταν $356,42 \leq \mu \leq 368,18$ το διάστημα που βασίζεται σε αυτό το δείγμα αποτελεί μια σωστή δήλωση για το μ .

Αλλά τι γίνεται με τα διαστήματα από άλλα πιθανά δείγματα μεγέθους 25;

Παράδειγμα Εκτίμησης Διαστήματος

(συνέχεια)

Δείγμα #	\bar{X}	Χαμηλότερο Όριο	Υψηλότερο Όριο	Περιέχει μ;
1	362,30	356,42	368,18	Ναι
2	369,50	363,62	375,38	Ναι
3	360,00	354,12	365,88	Όχι
4	362,12	356,24	368,00	Ναι
5	373,88	368,00	379,76	Ναι

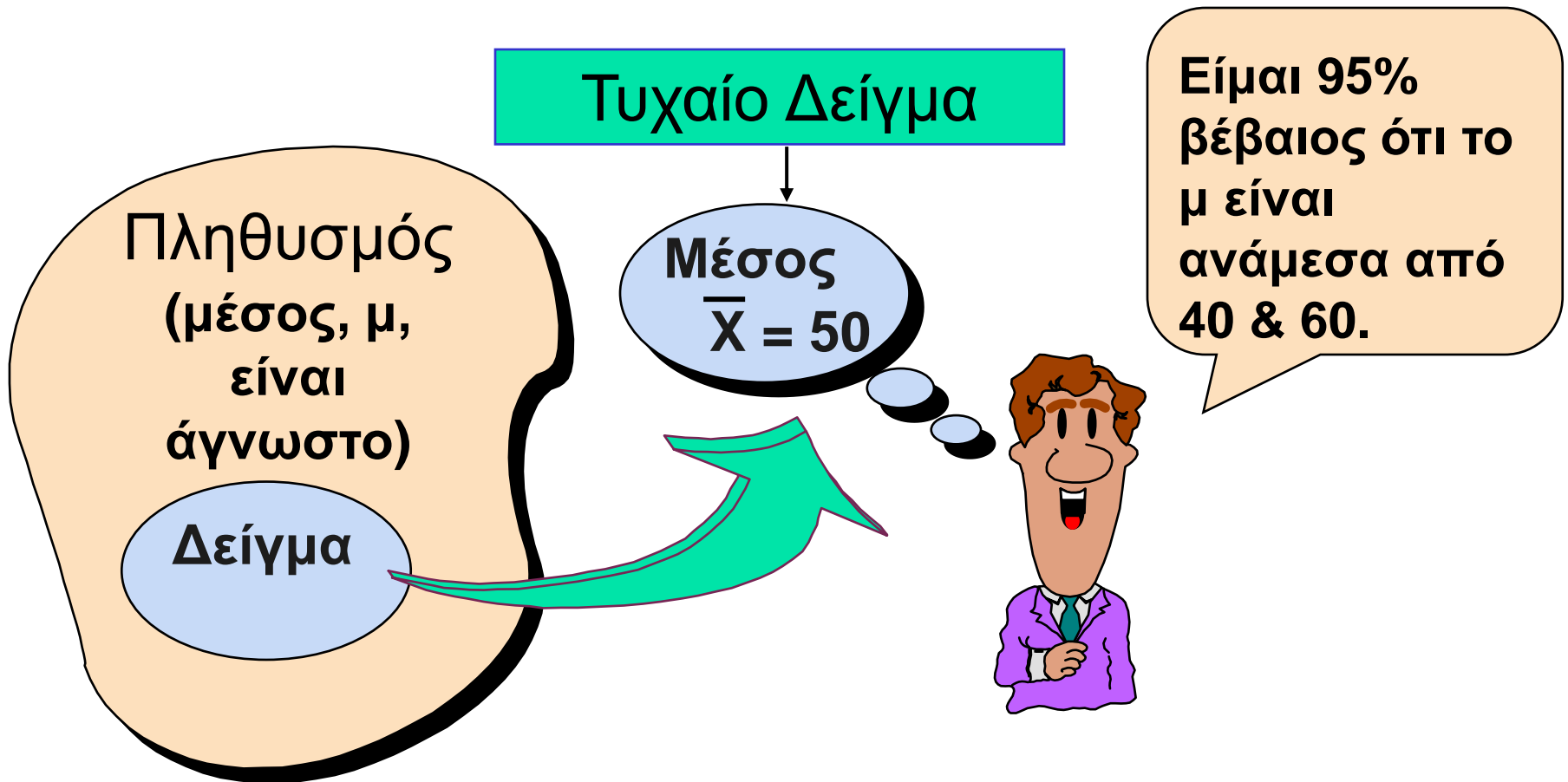
Παράδειγμα Εκτίμησης Διαστήματος

(συνέχεια)

- Στην πράξη λαμβάνετε μόνο ένα δείγμα μεγέθους n
- Στην πράξη δεν γνωρίζετε το μ έτσι δεν ξέρετε αν το διάστημα πραγματικά περιέχει μ
- Ωστόσο, γνωρίζετε ότι το 95% των διαστημάτων που σχηματίζονται με αυτόν τον τρόπο θα περιέχουν μ
- Έτσι, με βάση το ένα δείγμα, επιλέξατε πραγματικά ότι μπορείτε να είστε 95% σίγουροι ότι το διάστημα σας θα περιέχει μ (αυτό είναι ένα 95% **διάστημα εμπιστοσύνης**)

Σημείωση: Το 95% εμπιστοσύνη βασίζεται στο γεγονός ότι χρησιμοποιήσαμε $Z = 1,96$.

Διαδικασία Εκτίμησης



Γενικός Τύπος

- Ο γενικός τύπος για όλα τα διαστήματα εμπιστοσύνης είναι:

Σημειακή Εκτίμηση \pm (Κρίσιμη Τιμή)(Τυπικό Σφάλμα)

Όπου:

- **Σημειακή Εκτίμηση** είναι το στατιστικό δείγμα που υπολογίζει την πληθυσμιακή παράμετρο ενδιαφέροντος
- **Κρίσιμη Τιμή** είναι μια τιμή πίνακα βάσει της κατανομής δειγματοληψίας της εκτίμησης σημείων και του επιθυμητού επιπέδου εμπιστοσύνης
- **Τυπικό Σφάλμα** είναι η τυπική απόκλιση της σημειακής εκτίμησης

Επίπεδο εμπιστοσύνης

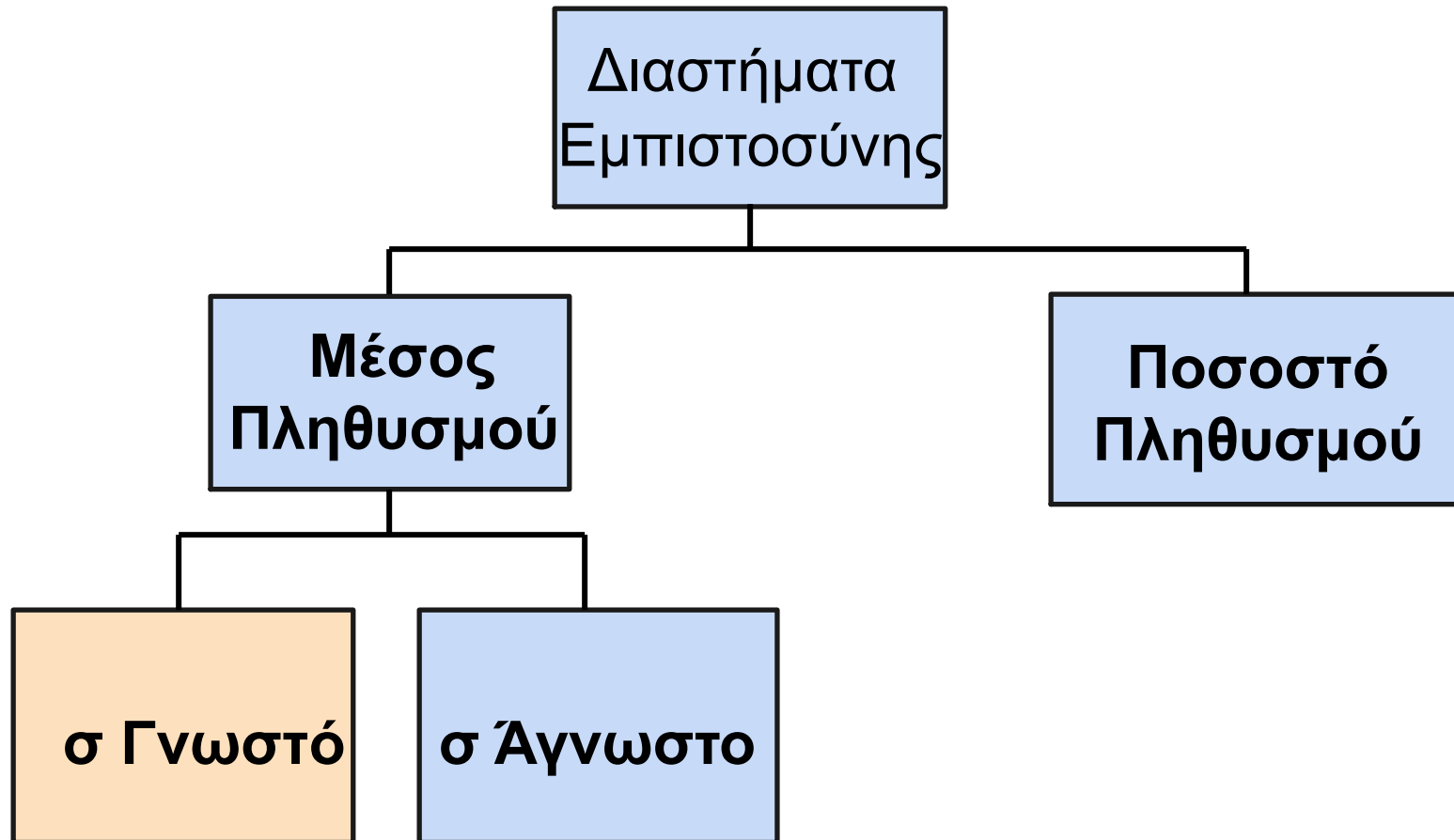
- Την εμπιστοσύνη που θα περιέχει το διάστημα την άγνωστη παράμετρο του πληθυσμού
- Ένα ποσοστό (μικρότερο από 100%)

Επίπεδο Εμπιστοσύνης, $(1-\alpha)$

(συνέχεια)

- Υποθετικό επίπεδο εμπιστοσύνης = 95%
- Επίσης γράφεται $(1 - \alpha) = 0,95$, (έτσι $\alpha = 0,05$)
- Μια ερμηνεία σχετικής συχνότητας :
 - Το 95% όλων των διαστημάτων εμπιστοσύνης που μπορούν να κατασκευαστούν θα περιέχουν την άγνωστη πραγματική παράμετρο
 - Ένα συγκεκριμένο διάστημα είτε θα περιέχει ή δεν θα περιέχει την πραγματική παράμετρο
 - Δεν υπάρχει πιθανότητα σε συγκεκριμένο διάστημα

Διαστήματα Εμπιστοσύνης



Διαστήματα Εμπιστοσύνης για μ (σ Γνωστό)

- Υποθέσεις
 - Η τυπική απόκλιση πληθυσμού σ είναι γνωστή
 - Ο πληθυσμός είναι κανονικά κατανομημένος
 - Αν ο πληθυσμός δεν είναι κανονικός, χρησιμοποιείτε μεγάλο δείγμα ($n > 30$)
- Εκτίμηση διαστήματος εμπιστοσύνης :

$$\bar{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

όπου \bar{X} είναι η σημειακή εκτίμηση

$Z_{\alpha/2}$ είναι η κανονική κρίσιμη τιμή διανομής για μια πιθανότητα $\alpha / 2$ σε κάθε άκρο

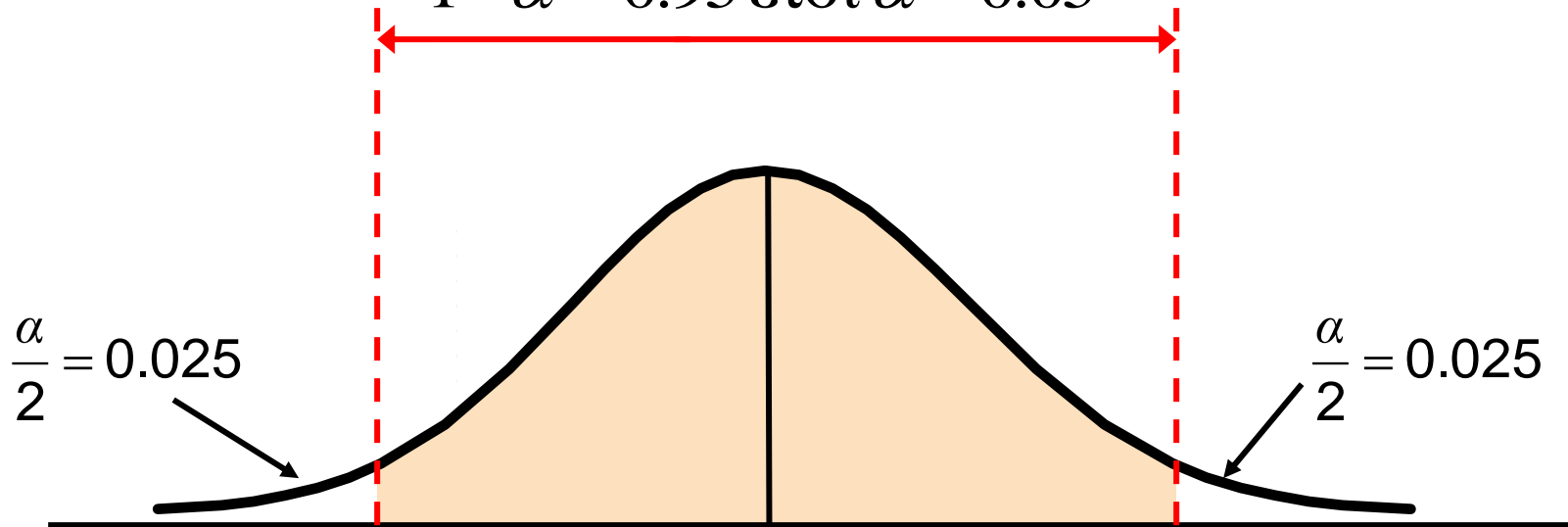
σ/\sqrt{n} είναι το τυπικό σφάλμα

Εύρεση της Κρίσιμης Τιμής, $Z_{\alpha/2}$

$$Z_{\alpha/2} = \pm 1.96$$

- Εξετάστε ένα διάστημα εμπιστοσύνης 95% :

$$1 - \alpha = 0.95 \text{ \acute{e}\tau\sigma\iota } \alpha = 0.05$$



Z μονάδες: $Z_{\alpha/2} = -1,96$

0

$Z_{\alpha/2} = 1,96$

X μονάδες: Χαμηλότερο
Όριο
Σημειακή Εκτίμηση
Εμπιστοσύνης

Υψηλότερο
Όριο
Εμπιστοσύνης

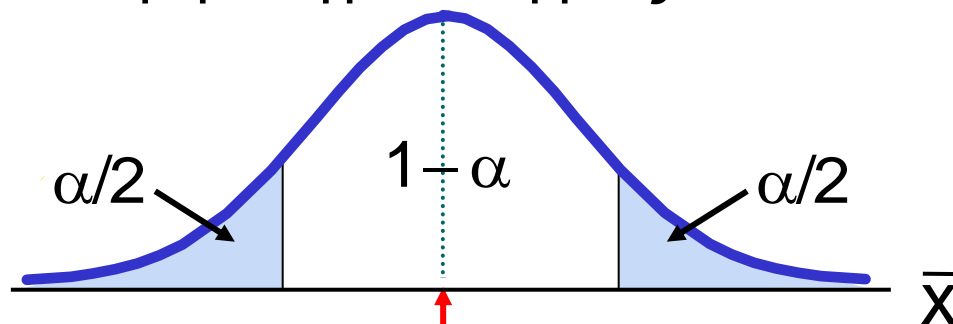
Συνήθη Επίπεδα Εμπιστοσύνης

- Τα συνήθως χρησιμοποιούμενα επίπεδα εμπιστοσύνης είναι 90%, 95% και 99%

<i>Επίπεδο Εμπιστοσύνης</i>	<i>Συντελεστής εμπιστοσύνης, $1 - \alpha$</i>	<i>$Z_{\alpha/2}$ τιμή</i>
80%	0,80	1,28
90%	0,90	1,645
95%	0,95	1,96
98%	0,98	2,33
99%	0,99	2,58
99,8%	0,998	3,08
99,9%	0,999	3,27

Διαστήματα και Επίπεδο Εμπιστοσύνης

Κατανομή Δειγματοληψίας του Μέσου

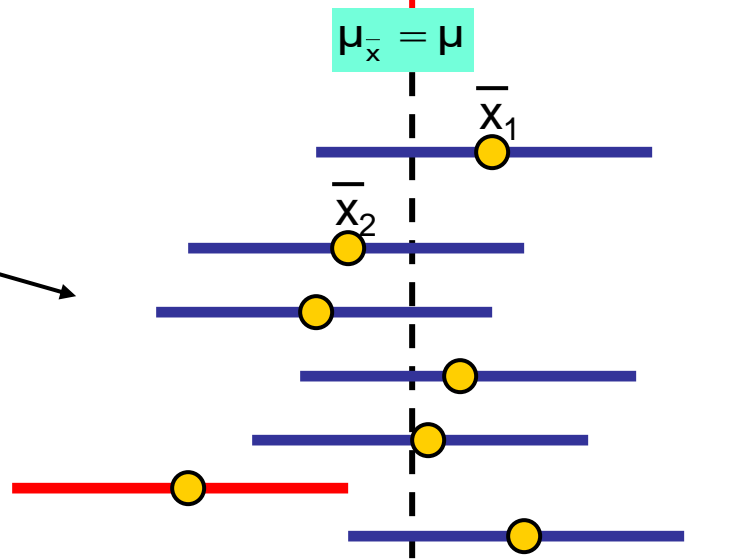
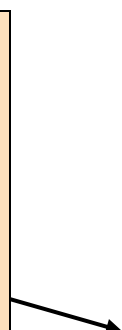


Τα διαστήματα
κυμαίνονται
από

$$\bar{X} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

έως

$$\bar{X} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$



Διαστήματα Εμπιστοσύνης

$(1-\alpha)100\%$
των
κατασκευασμένων
διαστημάτων
περιέχουν μ
 $(\alpha)100\%$ δεν
περιέχουν.

Παράδειγμα

- Ένα δείγμα 11 μαθητών από ένα μεγάλο κανονικό πληθυσμό ξοδεύει στο κυλικείο κατά μέσο όρο (\bar{x}) 2,20 ευρώ. Γνωρίζουμε από προηγούμενες μετρήσεις ότι η τυπική απόκλιση (σ) του πληθυσμού είναι 0,35 ευρώ.
- Προσδιορίστε ένα διάστημα εμπιστοσύνης 95% για το μέσο όρο (μ) του πληθυσμού.

Παράδειγμα

(συνέχεια)

■ Λύση:

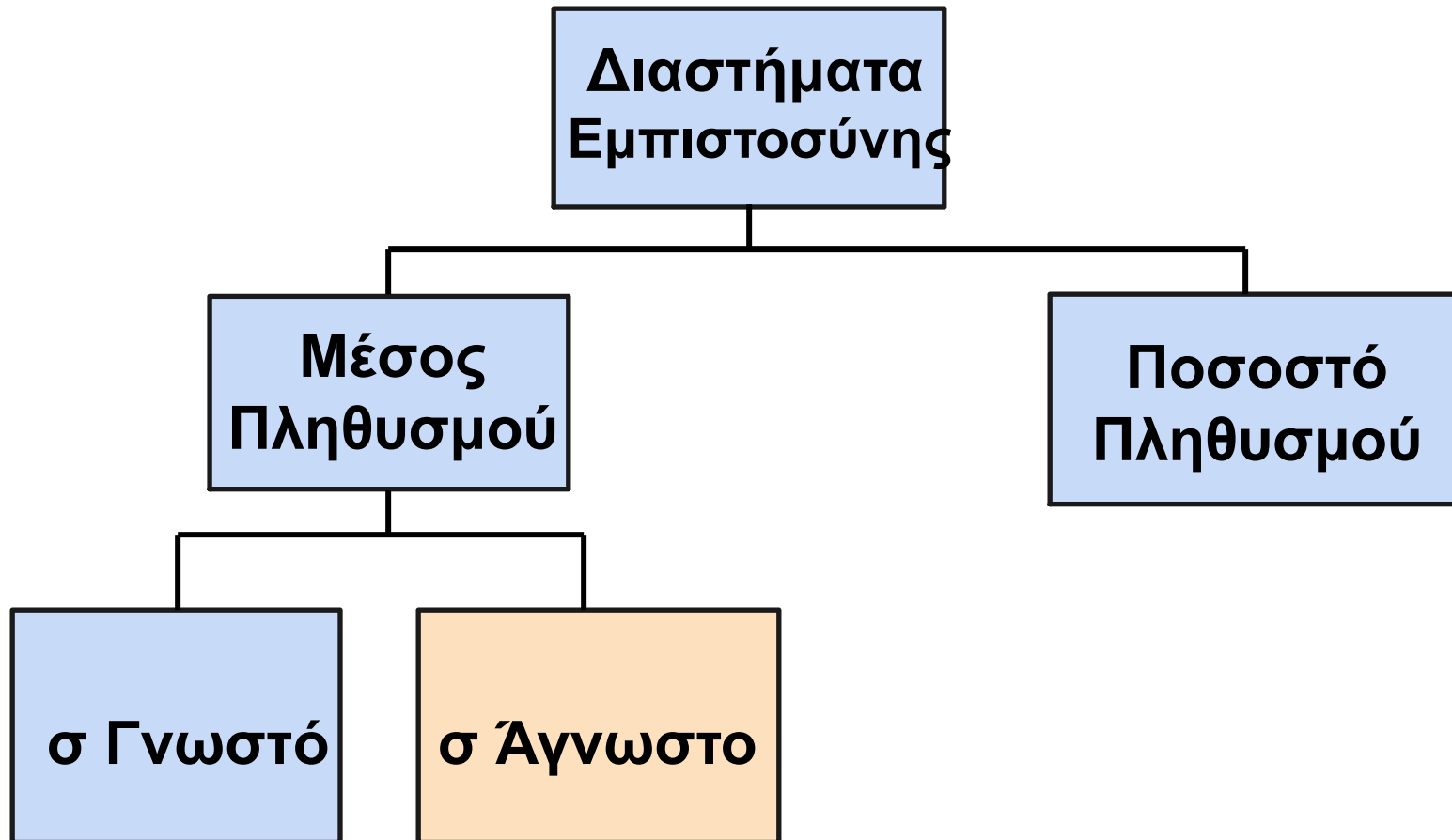
$$\begin{aligned}\bar{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \\ = 2,20 \pm 1,96 (0,35/\sqrt{11}) \\ = 2,20 \pm 0,2068\end{aligned}$$

$$1,9932 \leq \mu \leq 2,4068$$

Ερμηνεία

- Είμαστε 95% σίγουροι ότι η μέσος όρος του πληθυσμού (μ) είναι από 1,9932 έως 2,4068 ευρώ
- Παρόλο που ο πραγματικός μέσος μπορεί ή δεν μπορεί να είναι σε αυτό το διάστημα, 95% των διαστημάτων που σχηματίζονται με αυτόν τον τρόπο θα περιέχουν τον πραγματικό μέσο

Διαστήματα Εμπιστοσύνης



Γνωρίζετε πραγματικά το σ ;

- Πιθανόν όχι!
- Σε όλες σχεδόν τις πραγματικές επιχειρηματικές καταστάσεις, το σ δεν είναι γνωστό.
- Αν υπάρχει μια κατάσταση όπου το σ είναι γνωστό τότε το μ είναι επίσης γνωστό (δεδομένου ότι για να υπολογίσετε σ πρέπει να γνωρίζετε το μ).
- Εάν γνωρίζετε πραγματικά το μ , δεν θα χρειαζόταν να συγκεντρωθεί ένα δείγμα για να το εκτιμήσετε.

Διάστημα Εμπιστοσύνης για μ (σ Άγνωστο)

- Εάν η τυπική απόκλιση του πληθυσμού σ είναι άγνωστη, μπορούμε να αντικαταστήσουμε την τυπική απόκλιση του δείγματος, S
- Αυτό εισάγει επιπλέον αβεβαιότητα, καθώς το S είναι μεταβλητή από δείγμα σε δείγμα
- Επομένως, χρησιμοποιούμε την κατανομή t αντί για την κανονική κατανομή

Διάστημα Εμπιστοσύνης για μ (σ Άγνωστο)

(συνέχεια)

- Υποθέσεις
 - Η τυπική απόκλιση του πληθυσμού είναι άγνωστη
 - Ο πληθυσμός κατανέμεται κανονικά
 - Εάν ο πληθυσμός δεν είναι φυσιολογικός, χρησιμοποιήστε μεγάλο δείγμα ($n > 30$)
- Χρησιμοποιήστε την κατανομή Student t
- Εκτίμηση Διαστήματος Εμπιστοσύνης:

$$\bar{X} \pm t_{\alpha/2} \frac{S}{\sqrt{n}}$$

(όπου $t_{\alpha/2}$ είναι η κρίσιμη τιμή της κατανομής t με $n - 1$ βαθμούς ελευθερίας και μια περιοχή $\alpha / 2$ σε κάθε άκρο)

Κατανομή Student t

- Το t είναι μία οικογένεια κατανομών
- Η τιμή $t_{\alpha/2}$ εξαρτάται από τους βαθμούς ελευθερίας (d.f.)
 - Ο αριθμός των παρατηρήσεων που είναι ελεύθερες να μεταβληθούν μετά τον μέσο δείγματος έχουν υπολογιστεί

$$d.f. = n - 1$$

Βαθμοί Ελευθερίας (df)

Ιδέα: Ο αριθμός των παρατηρήσεων που είναι ελεύθερες να μεταβληθούν μετά τον μέσο δείγματος έχουν υπολογιστεί

Παράδειγμα: Υποθέστε ότι ο μέσος 3 αριθμών είναι 8,0

Θεωρήστε $X_1 = 7$
Θεωρήστε $X_2 = 8$
Πόσο είναι το X_3 ;



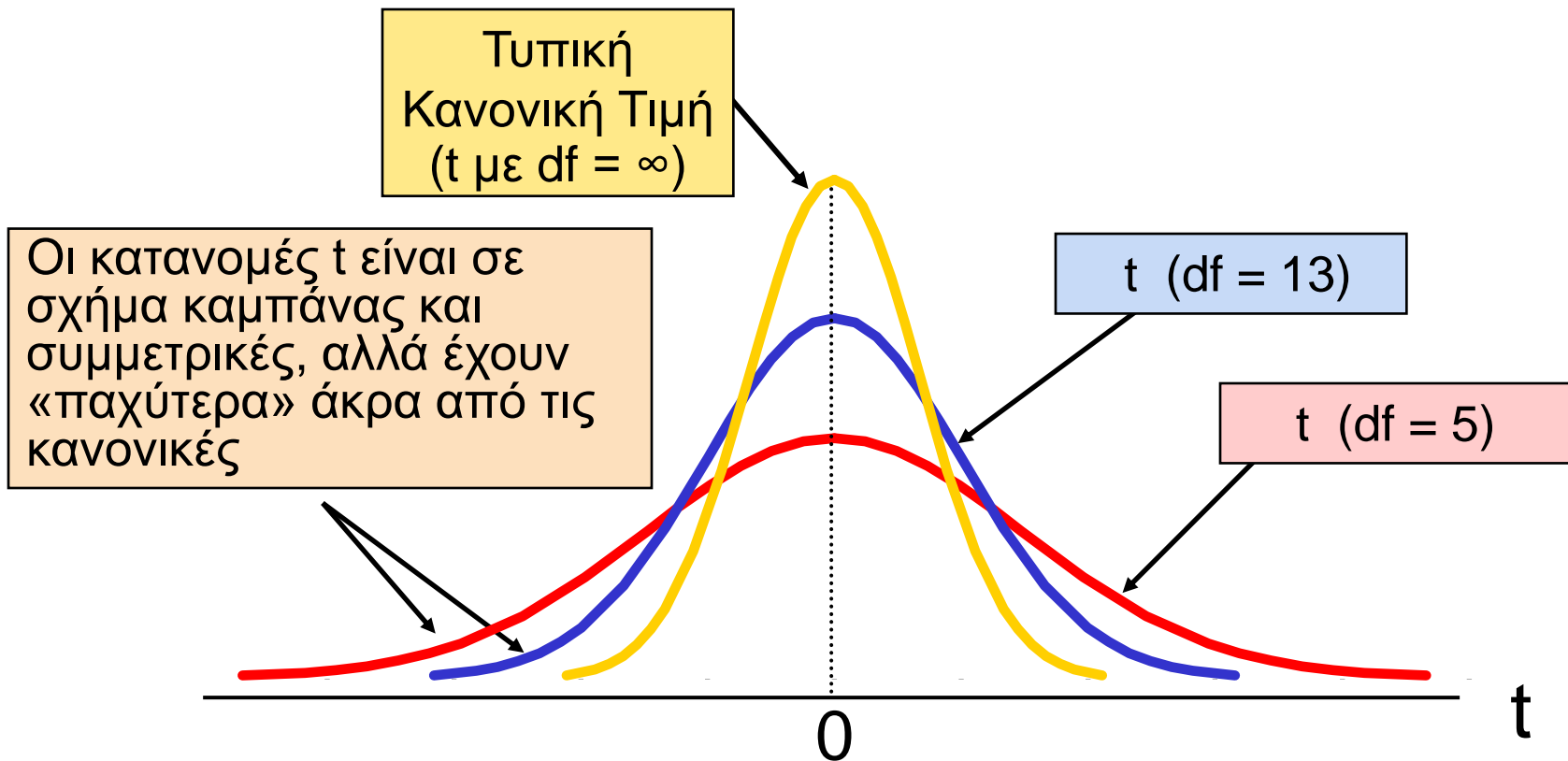
Αν ο μέσος αυτών των τριών τιμών είναι 8,0, τότε το X_3 πρέπει να είναι 9 (π.χ., το X_3 δεν είναι ελεύθερο να ποικίλλει)

Εδώ, $n = 3$, έτσι βαθμοί ελευθερίας = $n - 1 = 3 - 1 = 2$

(2 τιμές μπορεί να είναι οποιοσδήποτε αριθμός, αλλά ο τρίτος δεν είναι ελεύθερος να μεταβάλλεται για έναν δεδομένο μέσο)

Κατανομή Student t

Σημείωση: $t \rightarrow Z$ όσο το n αυξάνεται

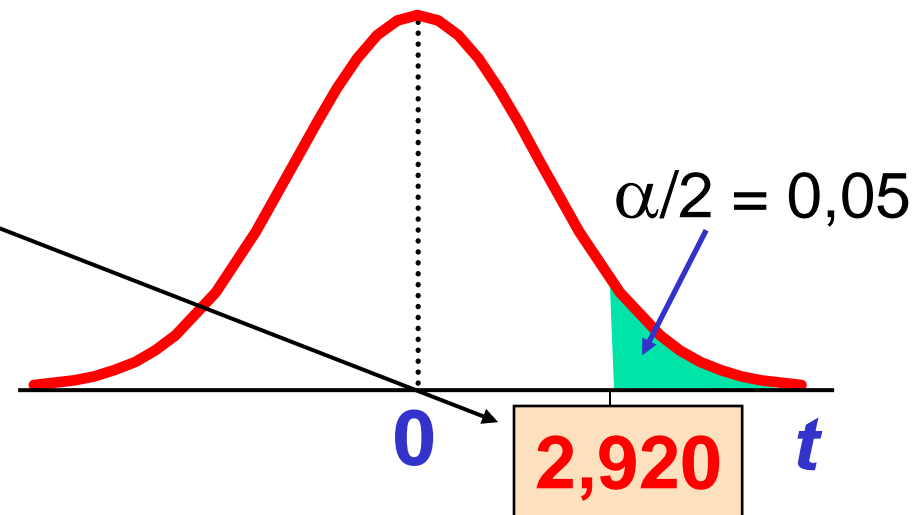


Πίνακας Student t

		Περιοχή Άνω Άκρου		
df	,10	,05	,025	
1	3,078	6,314	12,706	
2	1,886	2,920	4,303	
3	1,638	2,353	3,182	

Η συλλογή του πίνακα περιέχει τιμές t , όχι πιθανότητες

Θεωρήστε: $n = 3$
 $df = n - 1 = 2$
 $\alpha = 0,10$
 $\alpha/2 = 0,05$



Επιλεγμένες τιμές κατανομής t

Με σύγκριση με την τιμή Z

Επίπεδο	t	t	t	Z
<u>Εμπιστοσύνης (10 d.f.)</u>		<u>(20 d.f.)</u>	<u>(30 d.f.)</u>	<u>(∞ d.f.)</u>
0,80	1,372	1,325	1,310	1,28
0,90	1,812	1,725	1,697	1,645
0,95	2,228	2,086	2,042	1,96
0,99	3,169	2,845	2,750	2,58

Σημείωση: t \rightarrow Z καθώς το n αυξάνεται

Παράδειγμα διαστήματος εμπιστοσύνης κατανομής t

Ένα τυχαίο δείγμα $n = 25$ έχει $\bar{X} = 50$ και $S = 8$. Δημιουργήστε ένα διάστημα εμπιστοσύνης 95% για μ

- d.f. = $n - 1 = 24$, έτσι $t_{\alpha/2} = t_{0.025} = 2.0639$

Το διάστημα εμπιστοσύνης είναι

$$\bar{X} \pm t_{\alpha/2} \frac{S}{\sqrt{n}} = 50 \pm (2.0639) \frac{8}{\sqrt{25}}$$

$$46,698 \leq \mu \leq 53,302$$

Επίπεδο εμπιστοσύνης	0,800	0,900	0,950	0,980	0,990	0,995	0,998	0,999	
Μονόπλευρος	0,1000	0,0500	0,0250	0,0100	0,0050	0,0025	0,0010	0,0005	
Δίπλευρος	0,2000	0,1000	0,0500	0,0200	0,0100	0,005	0,0020	0,0010	
1	3,078	6,314	12,706	31,820	63,657	127,321	318,309	636,619	
2	1,886	2,920	4,303	6,965	9,925	14,089	22,327	31,599	
3	1,638	2,353	3,182	4,541	5,841	7,453	10,215	12,924	
4	1,533	2,132	2,776	3,747	4,604	5,598	7,173	8,610	
5	1,476	2,015	2,571	3,365	4,032	4,773	5,893	6,869	
6	1,440	1,943	2,447	3,143	3,707	4,317	5,208	5,959	
7	1,415	1,895	2,365	2,998	3,499	4,029	4,785	5,408	
8	1,397	1,860	2,306	2,897	3,355	3,833	4,501	5,041	
B	9	1,383	1,833	2,262	2,821	3,250	3,690	4,297	4,781
α	10	1,372	1,812	2,228	2,764	3,169	3,581	4,144	4,587
θ	11	1,363	1,796	2,201	2,718	3,106	3,497	4,025	4,437
μ	12	1,356	1,782	2,179	2,681	3,055	3,428	3,930	4,318
ο	13	1,350	1,771	2,160	2,650	3,012	3,372	3,852	4,221
ί	14	1,345	1,761	2,145	2,625	2,977	3,326	3,787	4,140
	15	1,341	1,753	2,131	2,602	2,947	3,286	3,733	4,073
E	16	1,337	1,746	2,120	2,584	2,921	3,252	3,686	4,015
λ	17	1,333	1,740	2,110	2,567	2,898	3,222	3,646	3,965
ε	18	1,330	1,734	2,101	2,552	2,878	3,197	3,610	3,922
υ	19	1,328	1,729	2,093	2,539	2,861	3,174	3,579	3,883
θ	20	1,325	1,725	2,086	2,528	2,845	3,153	3,552	3,850
ε	21	1,323	1,721	2,080	2,518	2,831	3,135	3,527	3,819
ρ	22	1,321	1,717	2,074	2,508	2,819	3,119	3,505	3,792
ί	23	1,319	1,714	2,069	2,500	2,807	3,104	3,485	3,768
α	24	1,318	1,711	2,064	2,492	2,797	3,090	3,467	3,745
ς	25	1,316	1,708	2,060	2,485	2,787	3,078	3,450	3,725
	26	1,315	1,706	2,056	2,479	2,779	3,067	3,435	3,707
	27	1,314	1,703	2,052	2,473	2,771	3,057	3,421	3,690
	28	1,313	1,701	2,048	2,467	2,763	3,047	3,408	3,674
	29	1,311	1,699	2,045	2,462	2,756	3,038	3,396	3,659
	30	1,310	1,697	2,042	2,457	2,750	3,030	3,385	3,646

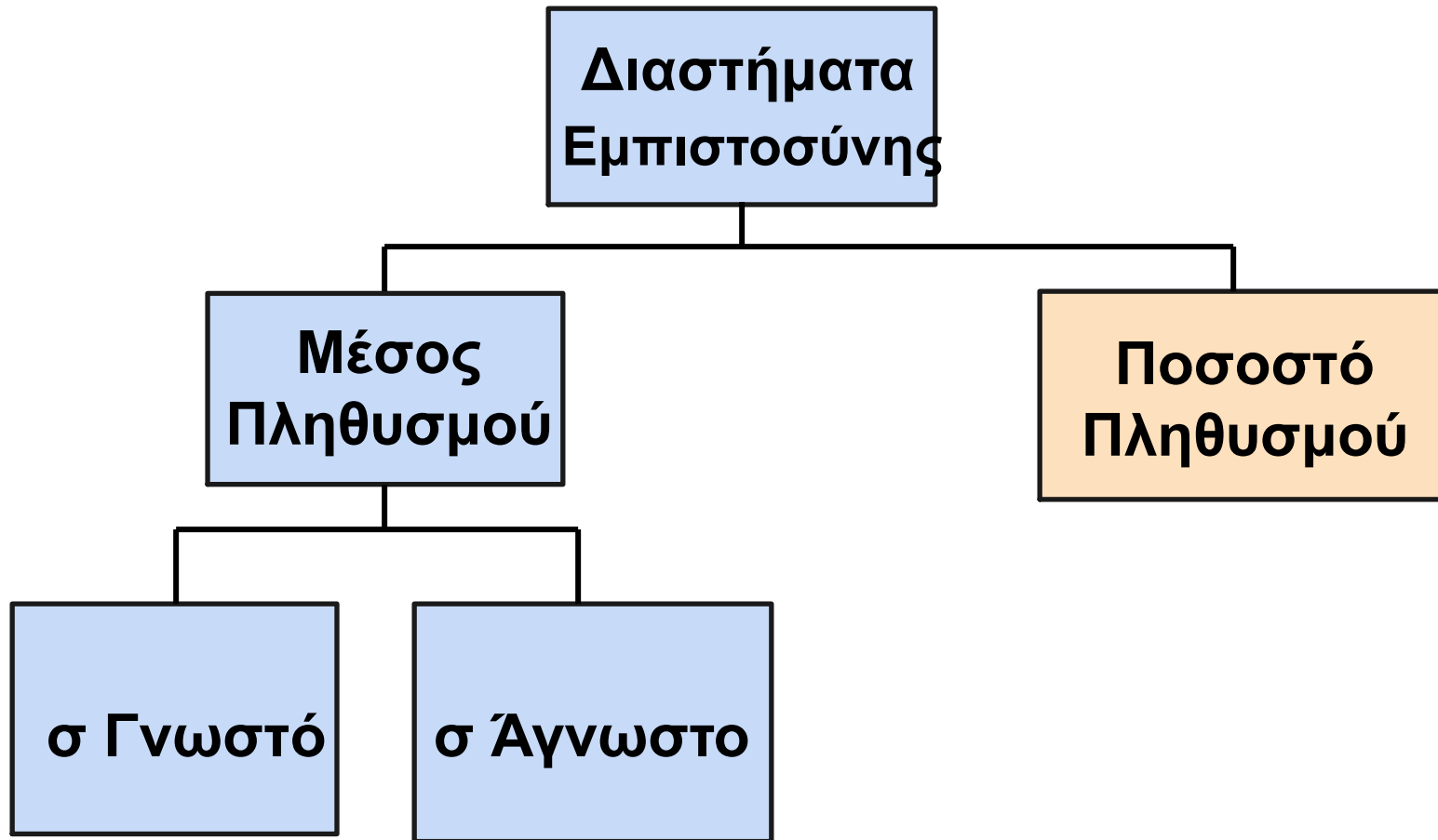
Πίνακας Student t

Παράδειγμα διαστήματος εμπιστοσύνης κατανομής t

(συνέχεια)

- Η ερμηνεία αυτού του διαστήματος απαιτεί την υπόθεση ότι ο πληθυσμός από τον οποίο λαμβάνετε δειγματοληψία είναι περίπου μια κανονική κατανομή (ειδικά επειδή το n είναι μόνο 25).
- Αυτή η κατάσταση μπορεί να ελεγχθεί δημιουργώντας ένα:
 - Διάγραμμα κανονικής πιθανότητας ή
 - Θηκόγραμμα

Διαστήματα Εμπιστοσύνης



Διαστήματα Εμπιστοσύνης για Το Ποσοστό Πληθυσμού, π

- Μια εκτίμηση διαστήματος για το ποσοστό του πληθυσμού (π) μπορεί να υπολογιστεί με την προσθήκη μιας απαίτησης για αβεβαιότητα στο ποσοστό του δείγματος (p)

Διαστήματα Εμπιστοσύνης για Το Ποσοστό Πληθυσμού, π

(συνέχεια)

- Θυμηθείτε ότι η κατανομή του ποσοστού του δείγματος είναι περίπου κανονική εάν το μέγεθος του δείγματος είναι μεγάλο, με τυπική απόκλιση.

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

- Θα το υπολογίσουμε με δείγματα δεδομένων:

$$\sqrt{\frac{p(1-p)}{n}}$$

Άκρα Διαστήματος Εμπιστοσύνης

- Τα ανώτερα και κατώτερα όρια εμπιστοσύνης για το ποσοστό του πληθυσμού υπολογίζονται με τον τύπο

$$p \pm Z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}$$

- **Όπου**
 - $Z_{\alpha/2}$ είναι η τυπική κανονική τιμή για το επιθυμητό επίπεδο εμπιστοσύνης
 - p είναι το ποσοστό του δείγματος
 - n είναι το μέγεθος του δείγματος
- Σημείωση: πρέπει να έχει $np > 5$ και $n(1-p) > 5$

Παράδειγμα

- Ένα τυχαίο δείγμα 100 ατόμων δείχνει ότι τα 25 είναι αριστερόχειρες
- Δημιουργήστε ένα διάστημα εμπιστοσύνης 95% για το πραγματικό ποσοστό των αριστερόχειρων



Παράδειγμα

(συνέχεια)

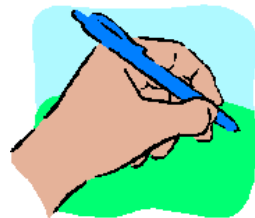
- Ένα τυχαίο δείγμα 100 ατόμων δείχνει ότι τα 25 είναι αριστερόχειρες. Δημιουργήστε ένα διάστημα εμπιστοσύνης 95% για το πραγματικό ποσοστό των αριστερόχειρων.

$$\begin{aligned} p \pm Z_{\alpha/2} \sqrt{p(1-p)/n} \\ = 25/100 \pm 1,96 \sqrt{0,25(0,75)/100} \\ = 0,25 \pm 1,96(0,0433) \\ = 0,1651 \leq \pi \leq 0,3349 \end{aligned}$$

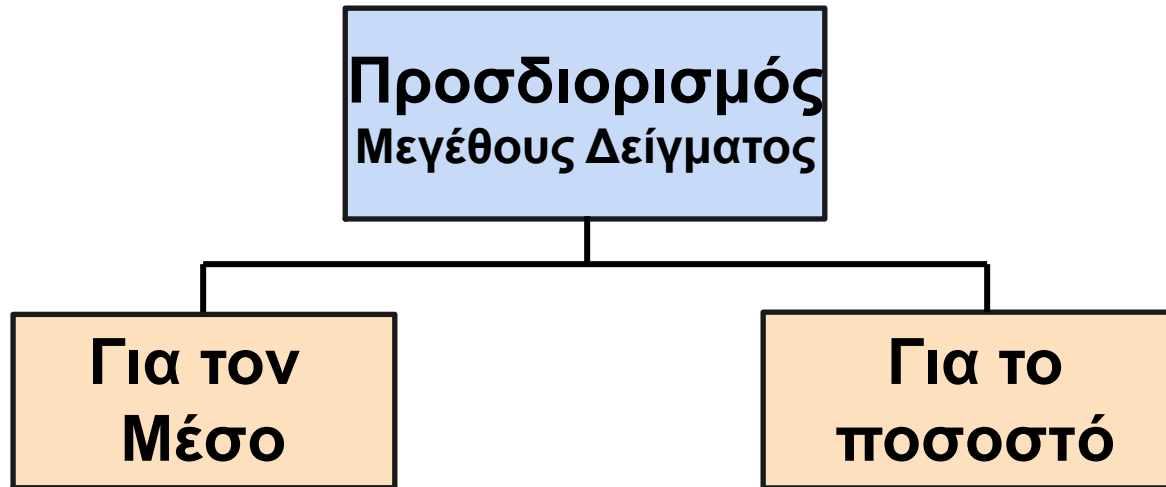


Ερμηνεία

- Είμαστε 95% βέβαιοι ότι το πραγματικό ποσοστό των αριστερόχειρων στον πληθυσμό είναι μεταξύ 16,51% και 33,49%.
- Αν και το διάστημα από 0,1651 έως 0,3349 μπορεί ή όχι να περιέχει το πραγματικό ποσοστό, το 95% των διαστημάτων που σχηματίζονται από δείγματα μεγέθους 100 με τον τρόπο αυτό θα περιέχουν το πραγματικό ποσοστό.



Προσδιορισμός Μεγέθους Δείγματος



Σφάλμα δειγματοληψίας

- Το απαιτούμενο μέγεθος δείγματος μπορεί να βρεθεί για να φτάσει σε ένα επιθυμητό **περιθώριο σφάλματος (e)** με συγκεκριμένο επίπεδο εμπιστοσύνης ($1 - \alpha$)
- Το περιθώριο λάθους καλείται επίσης **σφάλμα δειγματοληψίας**
 - το ποσό της ανακρίβειας στην εκτίμηση της παραμέτρου του πληθυσμού
 - το ποσό που προστίθεται και αφαιρείται στην σημειακή εκτίμηση για να σχηματιστεί το διάστημα εμπιστοσύνης

Προσδιορισμός μεγέθους δείγματος

Προσδιορισμός
μεγέθους δείγματος

Για τον
Μέσο

Σφάλμα δειγματοληψίας
(περιθώριο σφάλματος)

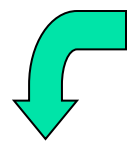
$$\bar{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$e = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

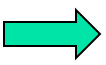
Προσδιορισμός μεγέθους δείγματος (συνέχεια)

Προσδιορισμός
Μεγέθους δείγματος

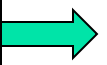
Για τον
Μέσο



$$e = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$



Λύνετε για n
για να πάρετε



$$n = \frac{Z_{\alpha/2}^2 \sigma^2}{e^2}$$

Προσδιορισμός μεγέθους δείγματος

(συνέχεια)

- Για να προσδιορίσετε το απαιτούμενο μέγεθος δείγματος για τον μέσο, πρέπει να γνωρίζετε:
 - Το επιθυμητό επίπεδο εμπιστοσύνης ($1 - \alpha$), το οποίο καθορίζει την κρίσιμη τιμή, $Z_{\alpha/2}$
 - Το αποδεκτό σφάλμα δειγματοληψίας, e
 - Την τυπική απόκλιση, σ

Παράδειγμα Απαιτούμενου Μεγέθους Δείγματος

Αν $\sigma = 45$, ποιο μέγεθος δείγματος χρειάζεται για να εκτιμηθεί ο μέσος εντός ± 5 με 90% εμπιστοσύνη;

$$n = \frac{Z^2 \sigma^2}{e^2} = \frac{(1,645)^2 (45)^2}{5^2} = 219,19$$

Άρα το απαιτούμενο μέγεθος δείγματος είναι **$n = 220$**

(Πάντα στρογγυλοποιημένο)

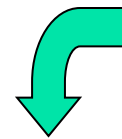
Αν το σ είναι άγνωστο

- Αν είναι άγνωστο, το σ μπορεί να εκτιμηθεί όταν χρησιμοποιείται ο απαιτούμενος τύπος μεγέθους δείγματος
 - Χρησιμοποιήστε μια τιμή για το σ που αναμένεται να είναι τουλάχιστον τόσο μεγάλη όσο η πραγματική σ
 - Επιλέξτε ένα πιλοτικό δείγμα και υπολογίστε το σ με την τυπική απόκλιση του δείγματος, S

Προσδιορισμός μεγέθους δείγματος

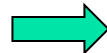
(συνέχεια)

Προσδιορισμός
Μεγέθους Δείγματος

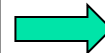


Για το
ΠΟΣΟΣΤό

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



Επιλύστε
για n για να
πάρετε



$$n = \frac{Z_{\alpha/2}^2 \pi (1 - \pi)}{e^2}$$

Προσδιορισμός μεγέθους δείγματος

(συνέχεια)

- Για να προσδιορίσετε το απαιτούμενο μέγεθος δείγματος για το ποσοστό, πρέπει να γνωρίζετε:
 - Το επιθυμητό επίπεδο εμπιστοσύνης $(1 - \alpha)$, που καθορίζει την κρίσιμη τιμή, $Z_{\alpha/2}$
 - Το αποδεκτό σφάλμα δειγματοληψίας, e
 - Το πραγματικό ποσοστό του υπό εξέταση ενδεχομένου, π
 - Το π μπορεί να εκτιμηθεί με πιλοτικό δείγμα, εάν είναι απαραίτητο (ή συντηρητικά χρησιμοποιήστε 0,5 ως εκτίμηση του π)

Παράδειγμα Απαιτούμενου Μεγέθους Δείγματος

Πόσο μεγάλο είναι ένα δείγμα που θα ήταν απαραίτητο για την εκτίμηση του πραγματικού ποσοστού των ελαττωμάτων σε ένα μεγάλο πληθυσμό εντός $\pm 3\%$, με 95% εμπιστοσύνη;
(Ας υποθέσουμε μια απόδοση πιλοτικού δείγματος $p = 0,12$)

Παράδειγμα Απαιτούμενου Μεγέθους Δείγματος

(συνέχεια)

Λύση:

Για 95% εμπιστοσύνη, χρησιμοποιείστε $Z_{\alpha/2} = 1,96$

$e = 0,03$

$p = 0,12$, χρησιμοποιείστε αυτό για να υπολογίσετε
το π

$$n = \frac{Z_{\alpha/2}^2 \pi (1 - \pi)}{e^2} = \frac{(1,96)^2 (0,12)(1 - 0,12)}{(0,03)^2} = 450,74$$

Έτσι χρησιμοποιείστε $n = 451$

Δεοντολογικά ζητήματα

- Μια εκτίμηση διαστήματος εμπιστοσύνης (που αντανακλά το σφάλμα δειγματοληψίας) θα πρέπει πάντα να περιλαμβάνεται κατά την αναφορά μιας σημειακής εκτίμησης
- Το επίπεδο εμπιστοσύνης πρέπει πάντα να αναφέρεται
- Πρέπει να αναφέρεται το μέγεθος του δείγματος
- Θα πρέπει επίσης να παρέχεται μια ερμηνεία της εκτίμησης του διαστήματος εμπιστοσύνης

Περίληψη Κεφαλαίου

Σε αυτό το κεφάλαιο αναφέραμε:

- Την κατασκευή και την ερμηνεία των εκτιμήσεων του διαστήματος εμπιστοσύνης για τον μέσο του πληθυσμού και το ποσοστό του πληθυσμού
- Τον προσδιορισμό του μεγέθους δείγματος που απαιτείται για την ανάπτυξη ενός διαστήματος εμπιστοσύνης για τον μέσο πληθυσμού ή το ποσοστό πληθυσμού

Business Statistics 7th Edition

A First Course



Levine Szabat Stephan

Κεφάλαιο 8

On Line Θέμα: Bootstrapping

Το Bootstrapping Είναι μια Μέθοδος που Χρησιμοποιείται όταν ο Πληθυσμός δεν είναι Κανονικός

DCOVA

Για να υπολογίσετε τον μέσο όρο του πληθυσμού χρησιμοποιώντας bootstrapping:

1. Επιλέξτε ένα τυχαίο δείγμα μεγέθους n χωρίς αντικατάσταση από πληθυσμό μεγέθους N .
2. Επαναλάβετε το αρχικό δείγμα επιλέγοντας τιμές n αντικαθιστώντας το αρχικό δείγμα.
3. Υπολογίστε το X από αυτό το δείγμα.
4. Επαναλάβετε τα βήματα 2 & 3 m διαφορετικές φορές.
5. Κατασκευάστε την κατανομή δειγματοληψίας εκ νέου του X .
6. Κατασκευάστε μια σειρά διάταξης όλου του συνόλου των X στα οποία πραγματοποιείται δειγματοληψία εκ νέου.
7. Σε αυτήν την σειρά διάταξης βρείτε την τιμή που κόβει το μικρότερο $\alpha/2$ (100%) και την τιμή που κόβει το μεγαλύτερο $\alpha/2$ (100%). Αυτές οι τιμές παρέχουν τα κατώτερα και τα ανώτερα όρια της εκτίμησης διαστήματος εμπιστοσύνης του bootstrap του μ .

Το Bootstrapping Απαιτεί την χρήση λογισμικού όπως Minitab ή JMP

DCOVA A

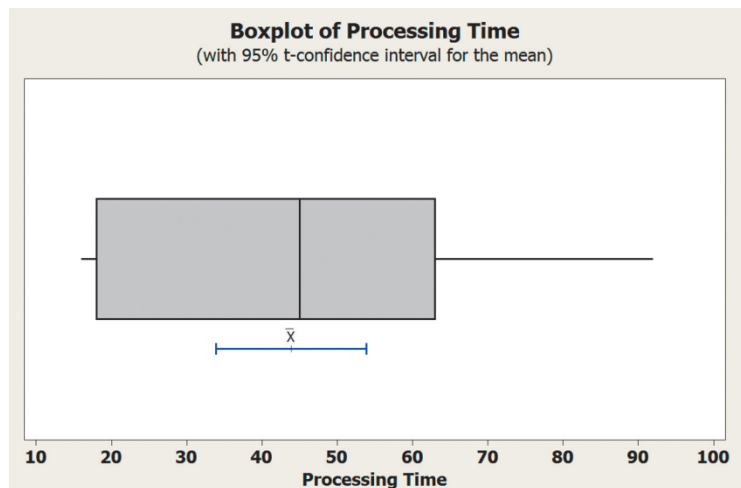
- Συνήθως χρησιμοποιείται πολύ μεγάλος αριθμός (χιλιάδες) δειγμάτων.
- Το λογισμικό είναι απαραίτητο για :
 - Την αυτοματοποίηση της διαδικασίας για την εκ νέου δειγματοληψία
 - Τον υπολογισμό της κατάλληλης στατιστικής δειγματοληψίας
 - Την δημιουργία της σειράς διάταξης
 - Την εύρεση των κατώτερων και ανώτερων ορίων εμπιστοσύνης

Παράδειγμα Bootstrapping -- Χρόνος Επεξεργασίας των Εφαρμογών Ασφάλειας Ζωής

DCOVA_A

Δείγμα που λήφθηκε 27 φορές χωρίς επανατοποθέτηση από τον πληθυσμό

73 19 16 64 28 28 31 90 60 56 31 56 22 18 45 48 17 17 17
91 92 63 50 51 69 16 17



Από το θηκόγραμμα προκύπτει ότι ο πληθυσμός δεν είναι φυσιολογικός, έτσι το διάστημα εμπιστοσύνης t δεν είναι κατάλληλο.

Χρησιμοποιήστε το bootstrapping για να διαμορφώσετε ένα διάστημα εμπιστοσύνης για μ .

Συγκρίνοντας το αρχικό δείγμα με το πρώτο δείγμα με επανατοποθέτηση

DCOVA A

Δείγμα που λήφθηκε 27 φορές χωρίς επανατοποθέτηση από τον πληθυσμό

73 19 16 64 28 28 31 90 60 56 31 56 22 18 45 48 17 17 17
91 92 63 50 51 69 16 17

Το αρχικό bootstrap ξεκινά από την παράλειψη ορισμένων τιμών (18, 45, 50, 63 και 91) που εμφανίζονται στο παραπάνω αρχικό δείγμα. Σημειώστε ότι η τιμή του 73 εμφανίζεται δύο φορές, παρόλο που στο αρχικό δείγμα εμφανίζεται μόνο μία φορά .

16 16 16 17 17 17 17 17 19 22 28 31 31 51 56 56 60 60 64
64 64 64 69 **73 73** 90 92

Πίνακας σειράς διάταξης των μέσων δειγματοληψίας για 100 δείγματα

Πέμπτο Μικρότερο

DCOVA



31,5926	33,9259	35,4074	36,5185	36,6296	36,9630	37,0370	37,0741
37,1481	37,3704	37,9259	38,1111	38,1481	38,2222	38,2963	38,7407
38,8148	38,8519	38,8889	39,0000	39,1852	39,3333	39,3704	39,6667
40,1481	40,5185	40,6296	40,9259	40,9630	41,2593	41,2963	41,7037
41,8889	42,0741	42,1111	42,1852	42,8519	43,0741	43,1852	43,3704
43,4444	43,7037	43,8148	43,8519	43,8519	43,9259	43,9630	44,1481
44,4074	44,5556	44,7778	45,0000	45,4444	45,5185	45,5556	45,6667
45,7407	45,8519	45,9630	45,9630	46,0000	46,1111	46,2963	46,2963
46,3333	46,3333	46,4815	46,6667	46,7407	46,9630	47,0741	47,2222
47,2963	47,3704	47,4815	47,4815	47,5556	47,6667	47,8519	48,5185
48,8889	49,0000	49,2222	49,4444	49,4815	49,4815	49,6296	49,6296
49,7407	50,2963	50,4074	50,5926	50,9259	51,4074	51,4815	51,5926
51,9259	52,3704	53,4074	54,3333				

Πέμπτο Μεγαλύτερο

Εύρεση ενός CI Bootstrap 90% για τον μέσο ενός πληθυσμού

DCOVA

- Για να βρούμε το 90% CI για 100 δείγματα, πρέπει να βρούμε την 0,05 (100) = 5η μικρότερη και την 5η μεγαλύτερη τιμή.
- Από τον πίνακα η 5η μικρότερη τιμή είναι 36,6296 και η 5η μεγαλύτερη τιμή είναι 51,5926.
- Το 90% bootstrap CI είναι (36,6296, 51,5926)

Περίληψη Θεμάτων

- Σε αυτό το θέμα αναλύσαμε
- Την έννοια του bootstrapping.
- Πότε το bootstrapping μπορεί να είναι κατάλληλο για χρήση.
- Πώς να χρησιμοποιήσετε το bootstrapping για να σχηματίσετε ένα διάστημα εμπιστοσύνης για τον μέσο ενός πληθυσμού.