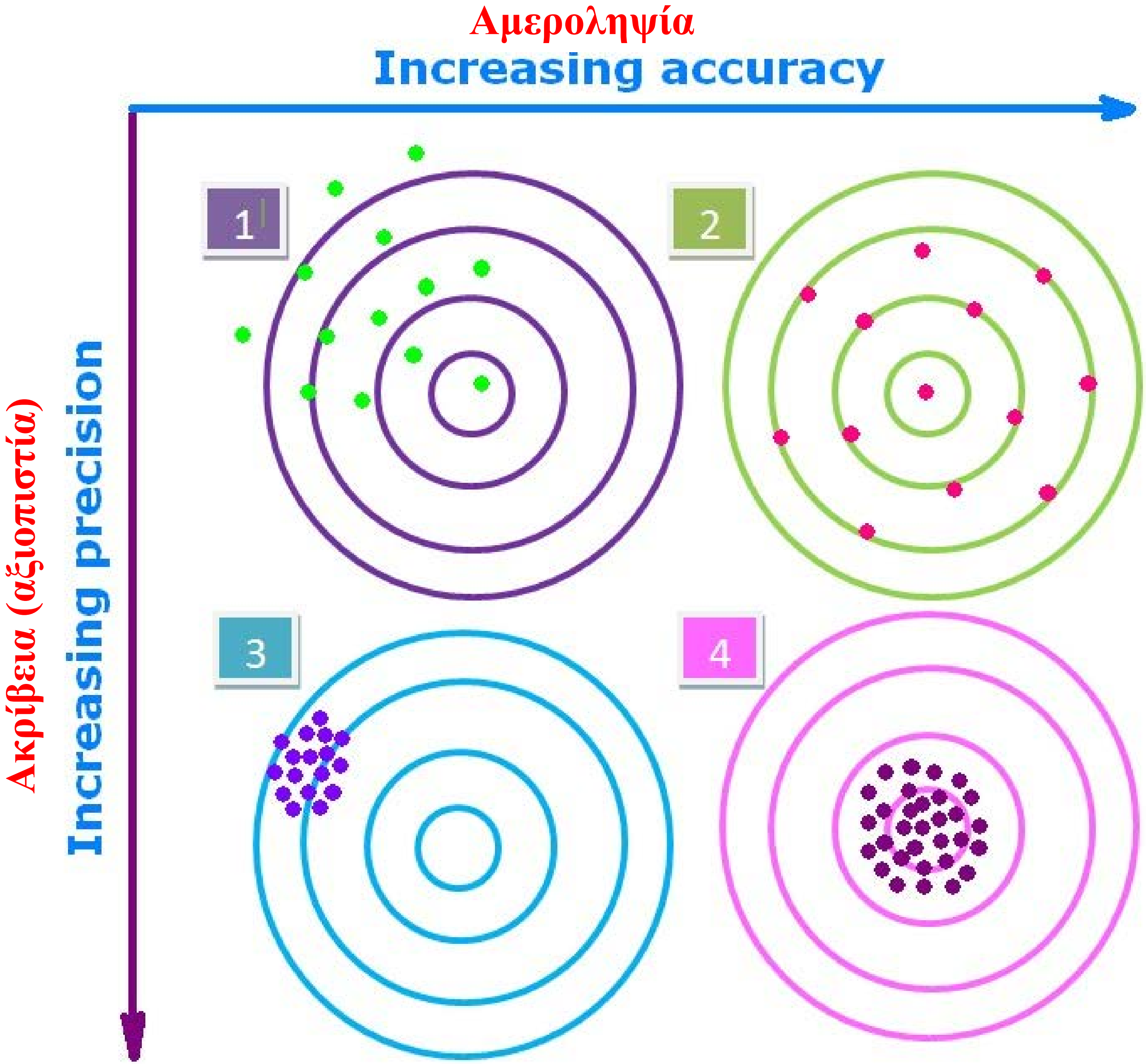
ΒΕΛΤΙΣΤΟ ΜΕΓΕΘΟΣ ΔΕΙΓΜΑΤΟΣ

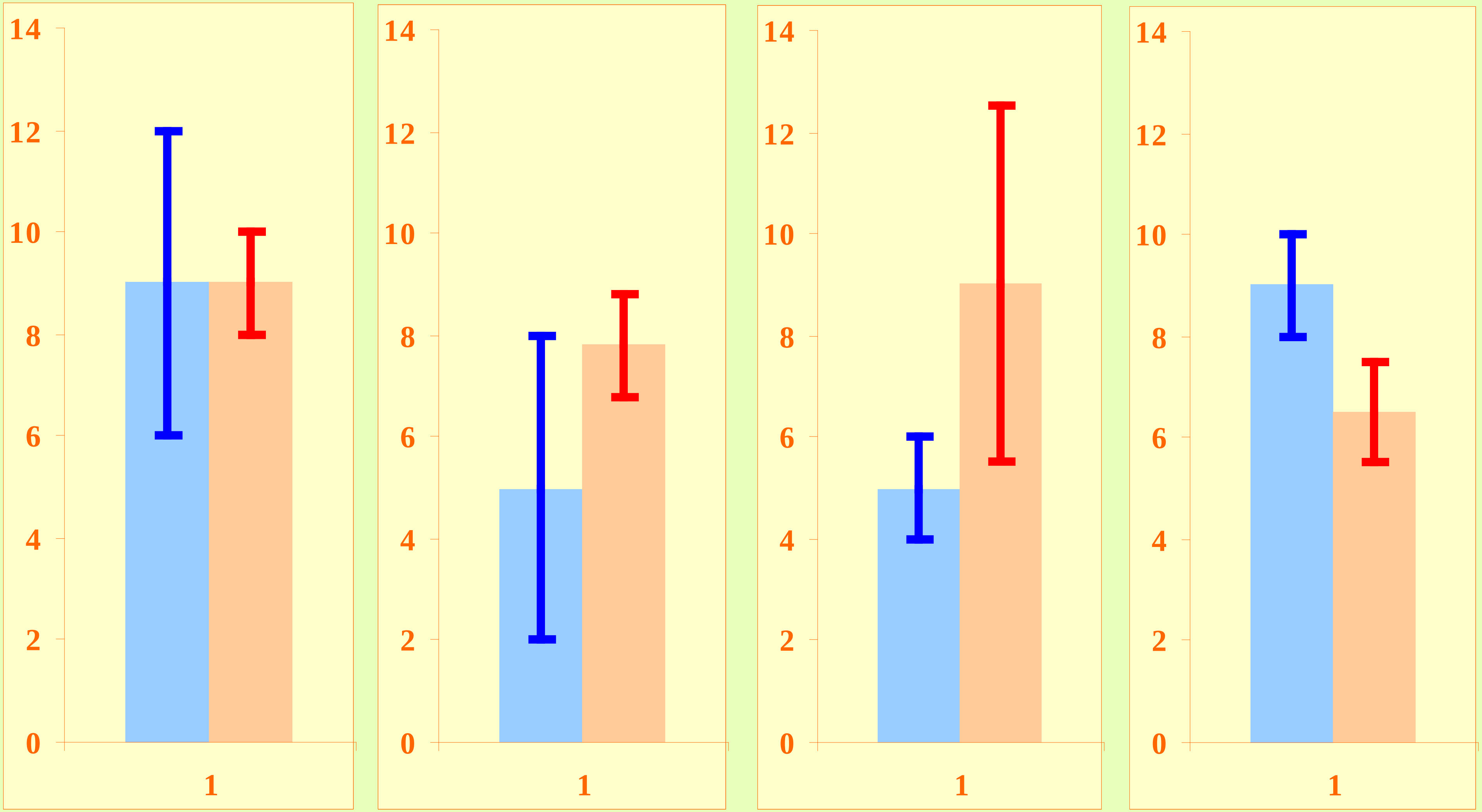
Κόστος πειράματος

Dafnia magna

Περιορισμοί σε χρόνο – χώρο, κ.λπ.
Προστασία σπάνιων ειδών….
Μπορεί να κρίνουμε ότι τελικά δεν αξίζει τον κόπο..!!!!

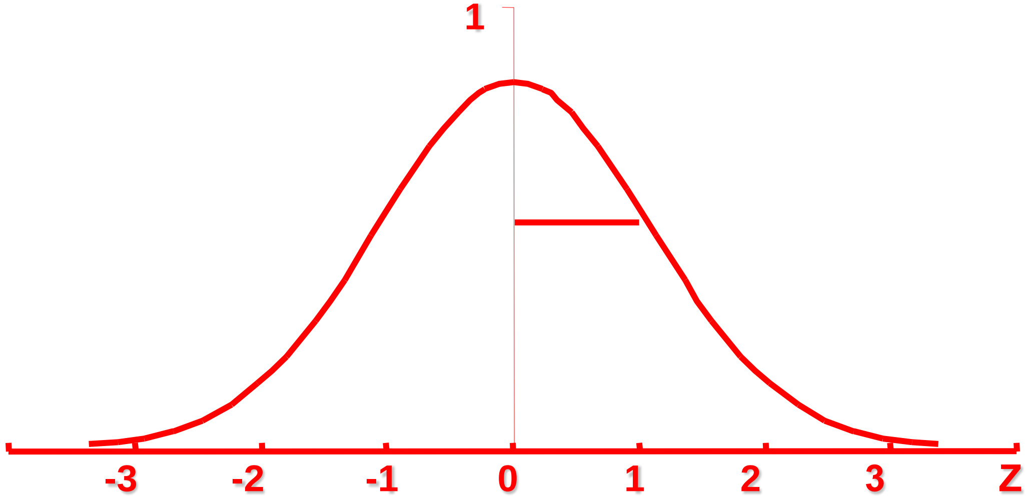

Αμεροληψία
Increasing accuracy
Ακρίβεια (αξιοπιστία)
Increasing precision
1
2
3
4

± διασπορά, ± τυπική απόκλιση, ± τυπικό σφάλμα, ± διάστημα εμπιστοσύνης

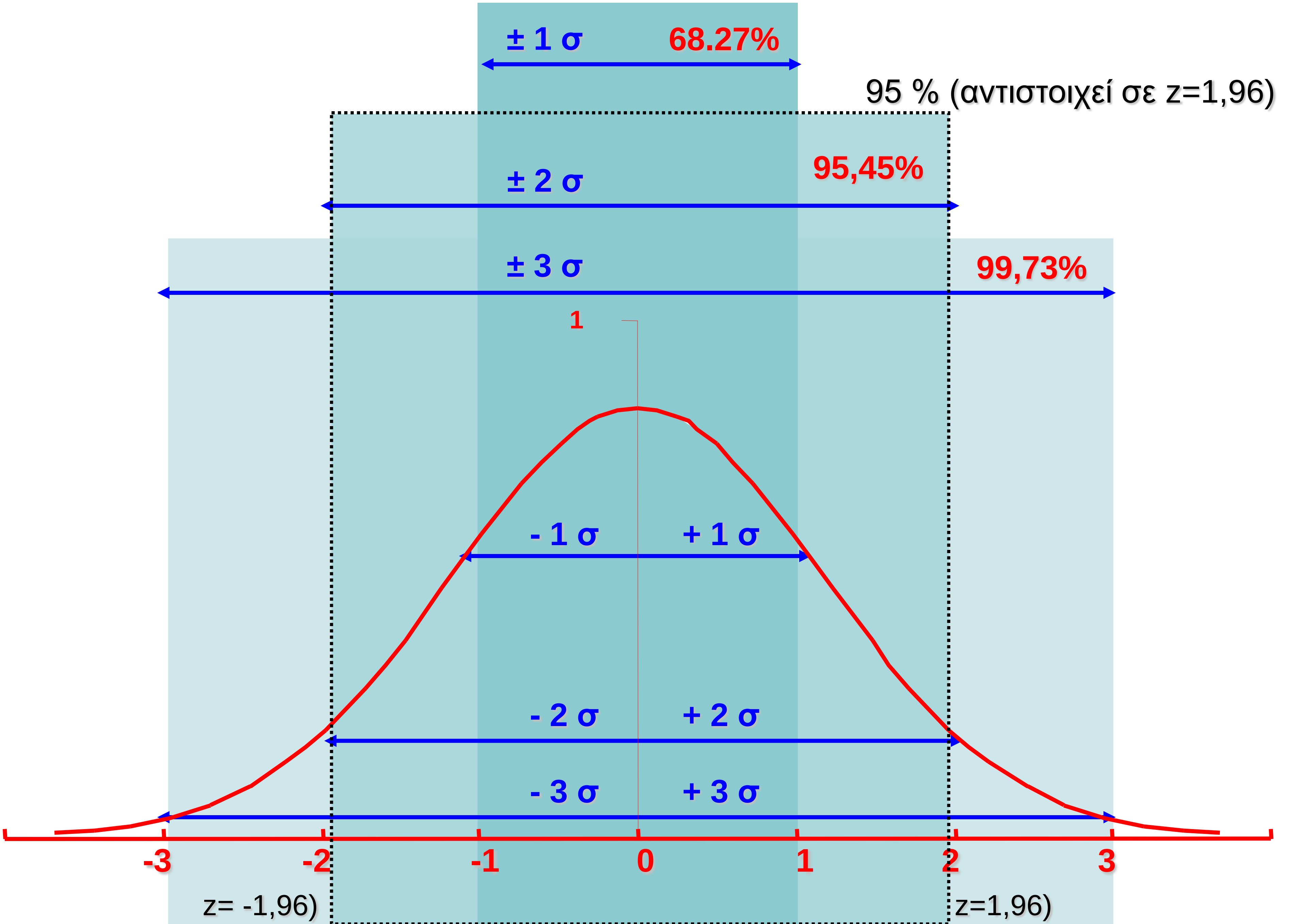


Δείγμα $$\frac{X_i - \overline{X}}{sd}$$

X_i	X transformed
78	0.536
74	0.708
53	1.609
74	0.708
93	-0.107
125	-1.480
99	-0.365
124	-1.437
76	0.622
104	-0.579
Mean 90	**Mean 0**
SD 23,3	**SD 1**

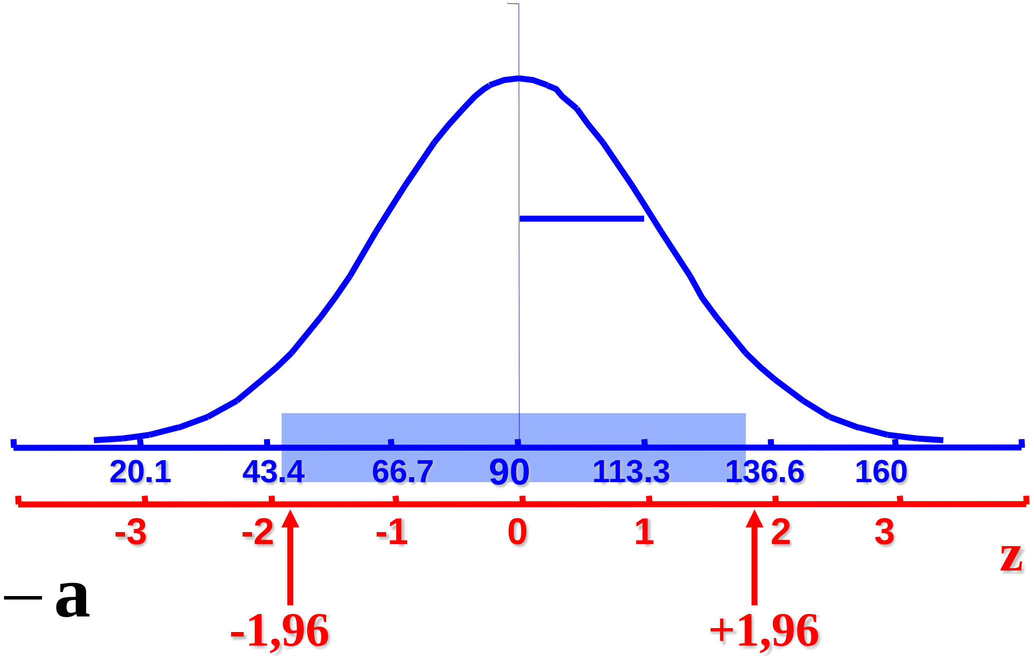


Πληθυσμός: $$\frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$



± 1 σ
68.27%
95 % (αντιστοιχεί σε z=1,96)
± 2 σ
95,45%
± 3 σ
99,73%
1
- 1 σ
+ 1 σ
- 2 σ
+ 2 σ
- 3 σ
+ 3 σ
-3
-2
-1
0
1
2
3
z= -1,96)
z=1,96)

Πληθυσμός Μέσων όρων $$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$$



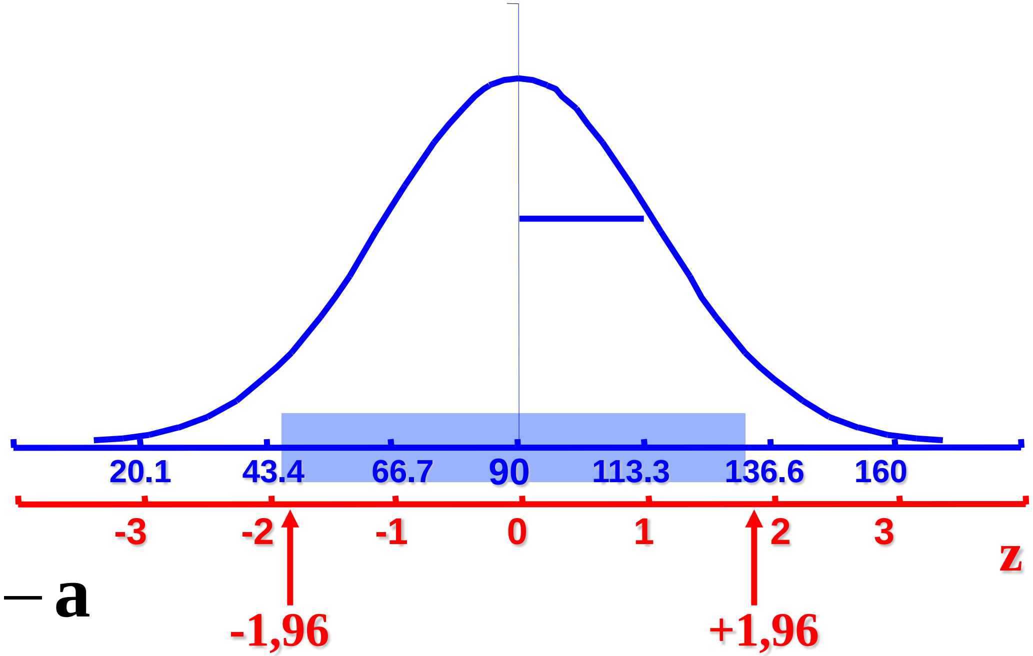
$$\Pr\left(-z_{a/2} < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < z_{a/2}\right) = 1-a$$

$$\Pr\left(\overline{X} - z_{a/2}\cdot\sigma/\sqrt{n} < \mu < \overline{X} + z_{a/2}\cdot\sigma/\sqrt{n}\right) = 1-a$$

Η πραγματική μέση τιμή μ βρίσκεται με πιθανότητα 95% γύρω από τη μέση τιμή του δείγματος $\overline{X}$ σε ένα διάστημα:

$$\left(\overline{X} \pm z_{a/2}\cdot\sigma/\sqrt{n}\right) = 1-a = 1\text{-}0.05 = 95\%$$

Πληθυσμός Μέσων όρων $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}$



$$\Pr\left(-z_{a/2} < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < z_{a/2}\right) = 1-a$$

$$\Pr\left(\overline{X}-z_{a/2}\cdot\sigma/\sqrt{n} < \mu < \overline{X}+z_{a/2}\cdot\sigma/\sqrt{n}\right) = 1-a$$

Η πραγματική μέση τιμή μ βρίσκεται με πιθανότητα 95% γύρω από τη μέση τιμή του δείγματος $\overline{X}$ σε ένα διάστημα:

$$\left(\overline{X} \pm z_{a/2}\cdot\sigma/\sqrt{n}\right) = 1-a = 1\text{-}0.05 = 95\%$$

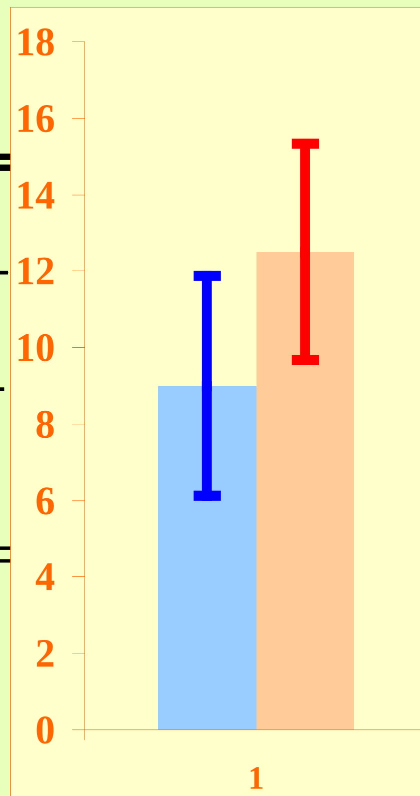
n	Δείγμα I	Δείγμα II	Δείγμα I	Δείγμα II	Δείγμα I	Δείγμα II	Δείγμα I	Δείγμα II
1	12	15	12	15	12	15	12	15
2	6	10	6	10	6	10	6	10
3			11	15	11	15	11	15
4			7	10	7	10	7	10
5					8	12	8	12
6					10	13	10	13
7							9	13
8							9	12
			1.5	1.4	1.0	0.9	0.8	0.8
			9.0	12.5	9.0	12.5	9.0	12.5
			2.9	2.8	1.9	1.8	1.4	1.3

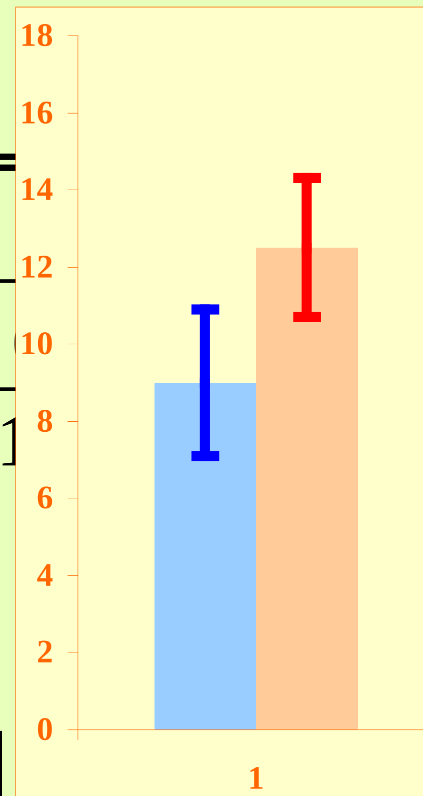
$$\bar{\mathbf{x}} = \frac{\sum_{i}^{n} \mathbf{x_i}}{\mathbf{n}} =$$

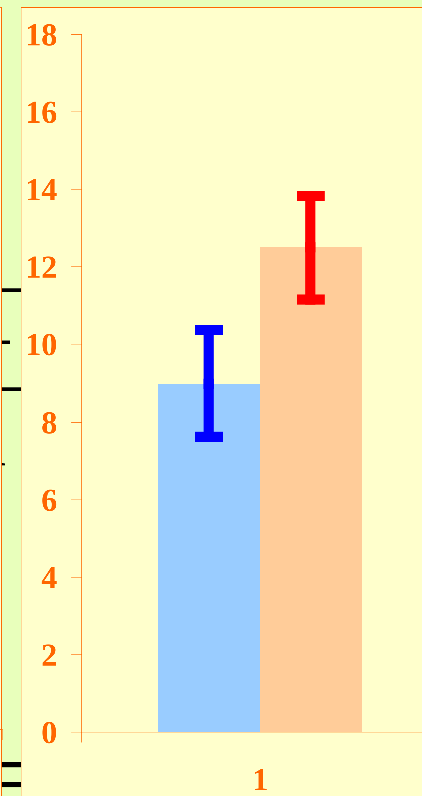
$$s = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}}$$

$$se = \frac{s}{\sqrt{n}} = \frac{4.2}{\sqrt{2}} =$$

$$C.I. = 1.96$$







ΣΥΝΕΧΕΙΣ ΜΕΤΑΒΛΗΤΕΣ
Κανονική κατανομή

1ο Κριτήριο $\sigma_{\bar{x}}$
Θέλουμε το ΤΥΠΙΚΟ ΣΦΑΛΜΑ
να έχει κάποια συγκεκριμένη τιμή

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \qquad n = \frac{\sigma^2}{\sigma^2_{\bar{x}}} \qquad \hat{n} = \frac{s^2}{s^2_{\bar{x}}}$$

Δεν είναι πολύ καλό μέτρο.
Π.χ. Μια εκτίμηση πληθυσμιακού μέσου με μ=10 και με $\sigma_{\bar{x}} = 1,2$ *έχει πολύ μικρότερη αξιοπιστία από ότι μια εκτίμηση πληθυσμιακού μέσου μ=100 και με* $\sigma_{\bar{x}} = 1,2$ *!!*

Έστω ότι θέλουμε να εκτιμήσουμε τα μέσα ύψη των δένδρων σε τέσσερις συστάδες Α, Β, Γ, και Δ. Από τις προκαταρκτικές δειγματοληψίες έχουμε εκτιμήσεις για τα μέσα ύψη και τις τυπικές αποκλίσεις. Έστω ότι θέλουμε να εκφράσουμε την αξιοπιστία με τυπικό σφάλμα $s_{\bar{x}} = 0{,}5$

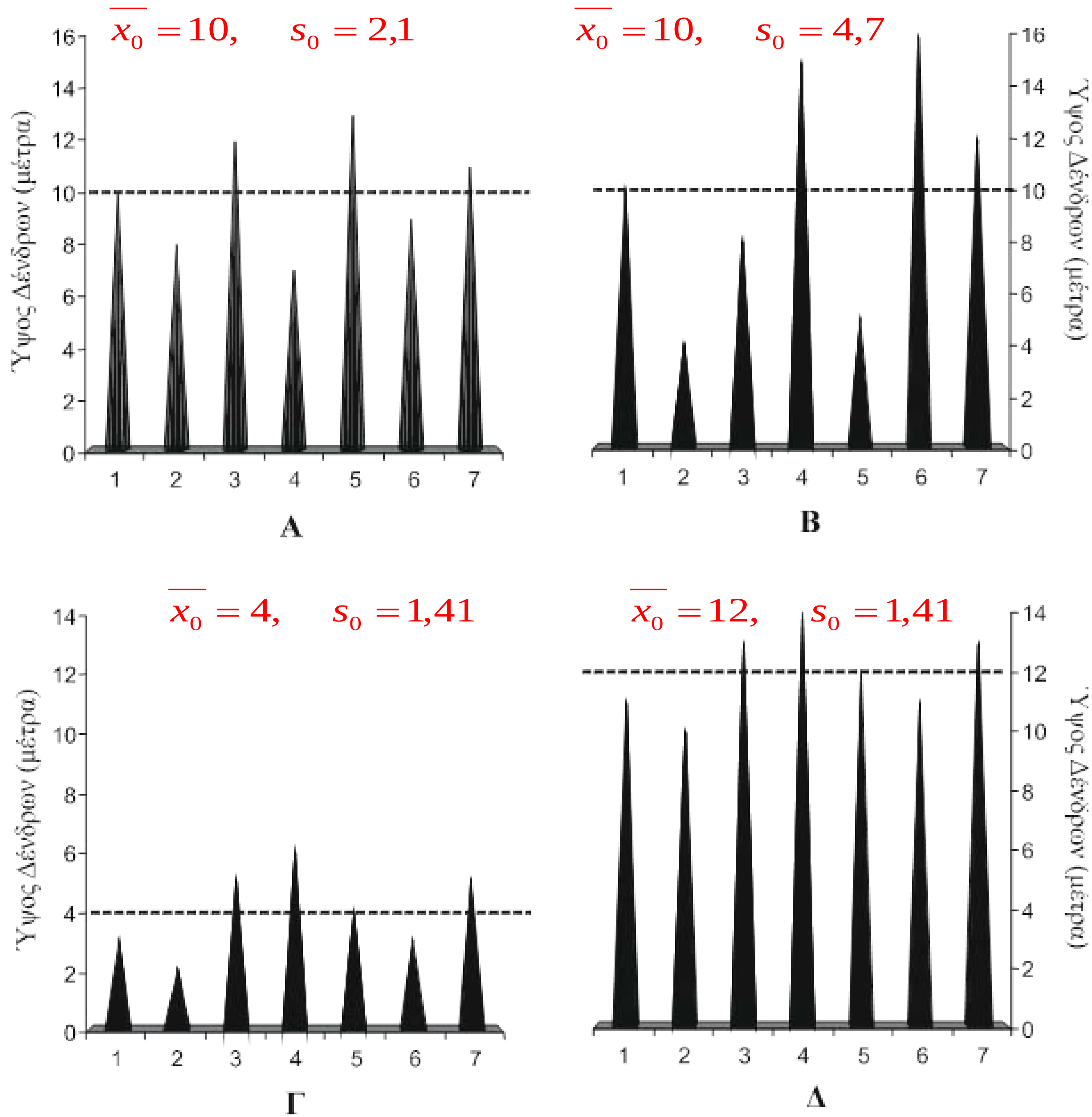
$$\hat{n} = \frac{s_0^2}{s_{\bar{x}}^2}$$

$${}_{A}\hat{n} = \frac{(2{,}1)^2}{(0{,}5)^2} = 17{,}64 \quad \delta\acute{\varepsilon}\nu\delta\rho\alpha$$

$${}_{B}\hat{n} = \frac{(4{,}7)^2}{(0{,}5)^2} = 88{,}36 \quad \delta\acute{\varepsilon}\nu\delta\rho\alpha$$

$${}_{\Gamma}\hat{n} = \frac{(1{,}41)^2}{(0{,}5)^2} = 7{,}97 \quad \delta\acute{\varepsilon}\nu\delta\rho\alpha$$

$${}_{\Delta}\hat{n} = \frac{(1{,}41)^2}{(0{,}5)^2} = 7{,}97 \quad \delta\acute{\varepsilon}\nu\delta\rho\alpha$$



2^o Κριτήριο

Θέλουμε ο συντελεστής παραλλακτικότητας (CV ή C) να έχει κάποια συγκεκριμένη τιμή

$$CV(\overline{X}) = \frac{\sigma_{\bar{x}}}{\mu_{\bar{x}}} = \frac{\sigma / \sqrt{n}}{\mu} = \frac{\sigma}{\mu\sqrt{n}}$$

$$n = \frac{\sigma^2}{\mu^2 CV^2} = \left(\frac{\sigma}{\mu CV}\right)^2$$

$$\hat{n} = \left(\frac{S_0}{\overline{X}_0 CV}\right)^2$$

Έστω ότι θέλουμε να εκτιμήσουμε τα μέσα ύψη των δένδρων σε τέσσερις συστάδες Α, Β, Γ, και Δ. Από τις προκαταρκτικές δειγματοληψίες έχουμε εκτιμήσεις για τα μέσα ύψη και τις τυπικές αποκλίσεις. Έστω ότι θέλουμε να εκφράσουμε την αξιοπιστία με το συντελεστή παραλλακτικότητας να είναι $CV = 0{,}06 \quad \delta\eta\lambda. \quad 6\%$

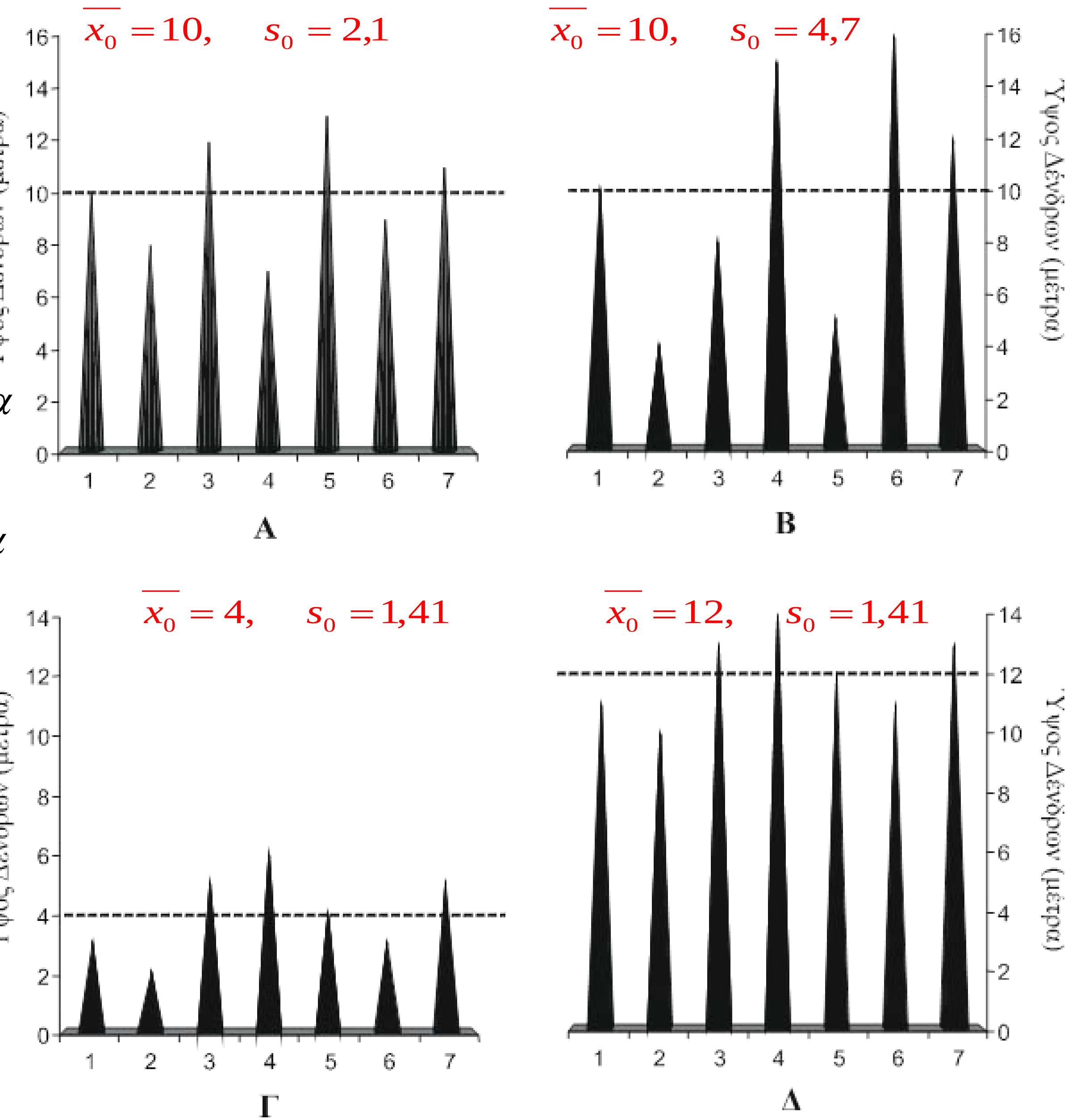
$$\hat{n} = \frac{s_0^2}{\overline{X}_0^2 \cdot CV^2}$$

$${}_A\hat{n} = \frac{(2{,}1)^2}{(10)^2(0.06)^2} = 12{,}25 \quad \delta\acute{\varepsilon}\nu\delta\rho\alpha$$

$${}_B\hat{n} = \frac{(4{,}7)^2}{(10)^2(0.06)^2} = 61{,}36 \quad \delta\acute{\varepsilon}\nu\delta\rho\alpha$$

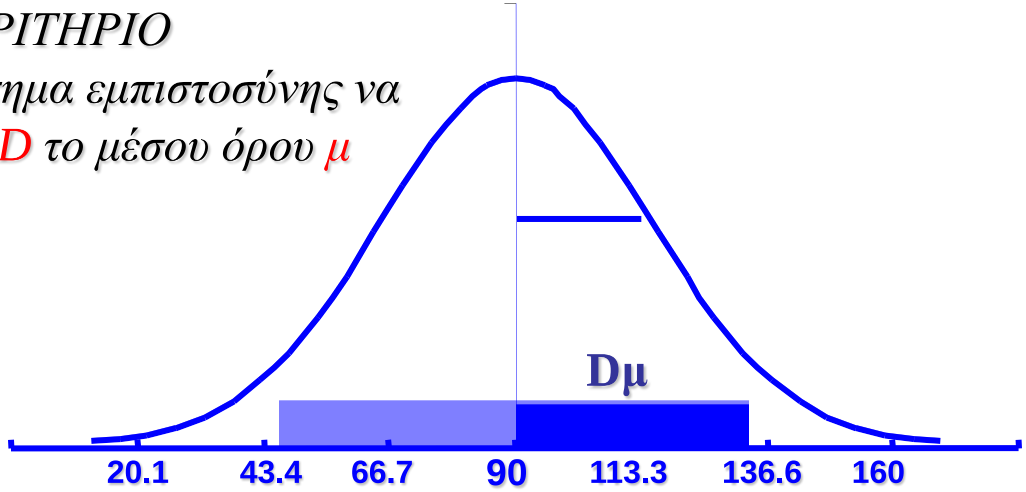
$${}_\Gamma\hat{n} = \frac{(1{,}41)^2}{(4)^2(0.06)^2} = 34{,}51 \quad \delta\acute{\varepsilon}\nu\delta\rho\alpha$$

$${}_\Delta\hat{n} = \frac{(1{,}41)^2}{(12)^2(0.06)^2} = 3{,}86 \quad \delta\acute{\varepsilon}\nu\delta\rho\alpha$$



3^o ΚΡΙΤΗΡΙΟ

Θέλουμε το διάστημα εμπιστοσύνης να είναι ποσοστό D το μέσου όρου μ



$$\left(z_{a/2}\cdot \sigma/\sqrt{n}\right)=D\cdot \mu$$

$$n=\left(\frac{z_{a/2}\cdot \sigma}{D\cdot \mu}\right)^2$$

$$n=\left(\frac{z_{a/2}\cdot S_0}{D\cdot \bar{x}_0}\right)^2$$

Αν τα προκαταρκτικά δείγματα είναι μικρά (<30) [που συνήθως είναι] αντί για κανονική κατανομή έχω t κατανομή οπότε αντί για $Z_{a/2}$ βάζω $t_{a/2,n-1}$

και επιλύω με την μέθοδο των επανειλημμένων δοκιμών

$$\hat{n}=\left(\frac{t_{a/2,n-1}\cdot S_0}{D\cdot \bar{X}_0}\right)^2$$

$$\hat{n} = \left(\frac{t_{a/2,n-1} \cdot S_0}{D \cdot \bar{X}_0} \right)^2$$

X
78
74
53
74
93
127
99
127
76
104

Άθροισμα 905

$\bar{x}=$ 90.5

$S_0 =$ 22.9

$D =$ 0,10

$a=$ 0.05

$n_{δοκιμής}$	$t_{a/2,n-1}$	$\hat{n}_{εκτιμούμενο}$
5	2,776	49,34
10	2,262	32,76
20	2,093	28,04
24	2,069	27,39
26	2,060	27,15
27	2,056	27,05

Έστω ότι θέλουμε να εκτιμήσουμε τα μέσα ύψη των δένδρων σε τέσσερις συστάδες Α, Β, Γ, και Δ. Από τις προκαταρκτικές δειγματοληψίες, με μεγέθη δειγμάτων >30, έχουμε εκτιμήσεις για τα μέσα ύψη και τις τυπικές αποκλίσεις. Έστω ότι θέλουμε να εκφράσουμε την αξιοπιστία με το διάστημα εμπιστοσύνης να είναι D=0,05 της μέσης τιμής

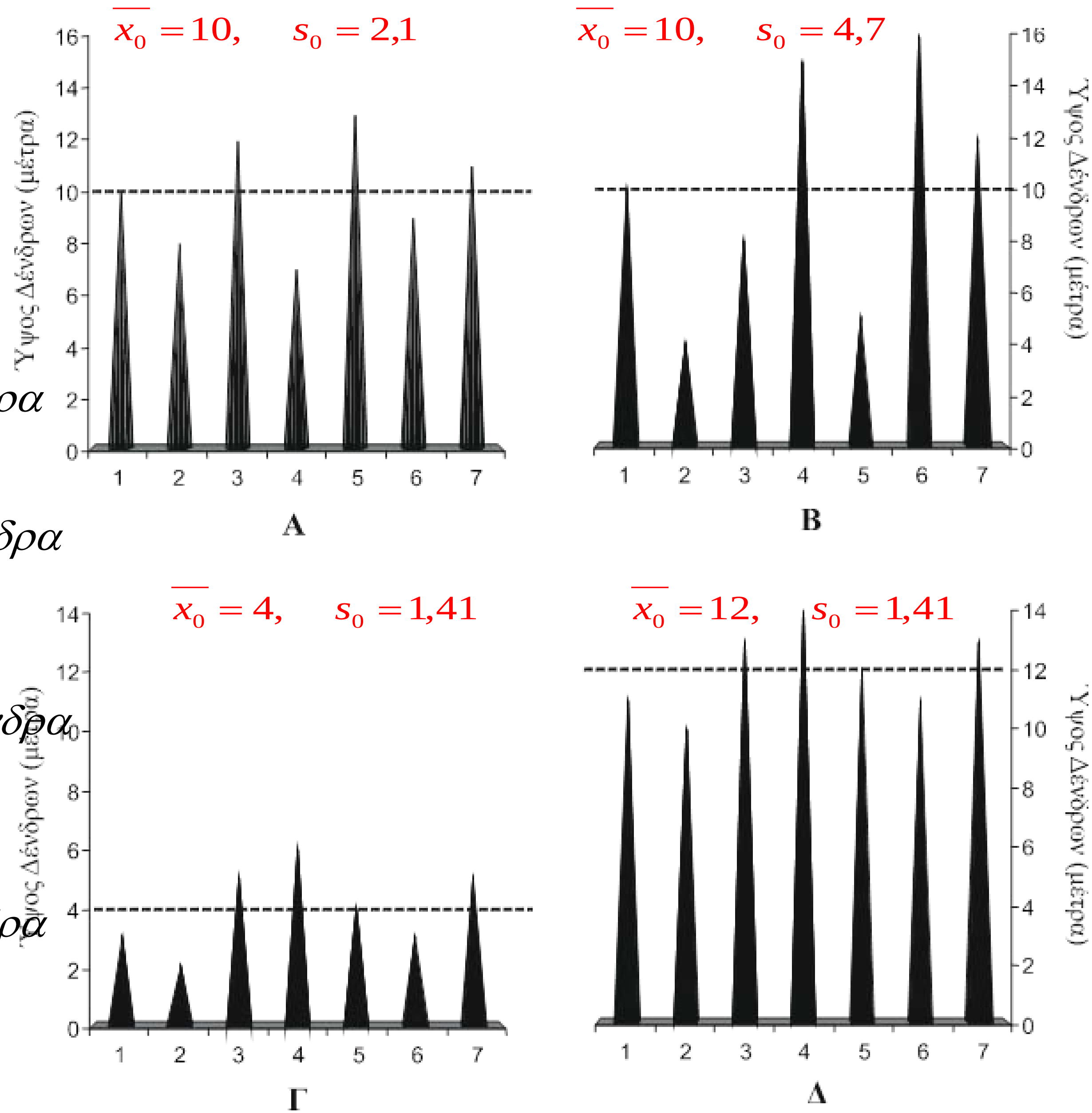
$$\widehat{\mathbf{n}} = \left(\frac{\mathbf{z_{a/2}} \cdot S_0}{\mathbf{D} \cdot \overline{\mathbf{x}}_0} \right)^2$$

$${}_{A}\hat{n} = \frac{(1{,}96)^2 (2{,}1)^2}{(10)^2 (0{,}05)^2} = 67{,}76 \quad \delta\acute{\varepsilon}\nu\delta\rho\alpha$$

$${}_{B}\hat{n} = \frac{(1{,}96)^2 (4{,}7)^2}{(10)^2 (0{,}05)^2} = 339{,}44 \quad \delta\acute{\varepsilon}\nu\delta\rho\alpha$$

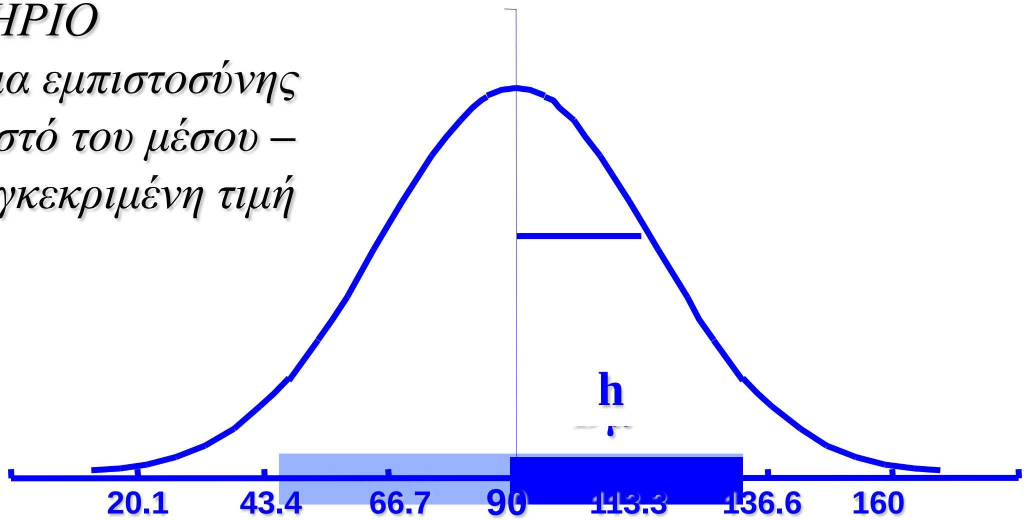
$${}_{\Gamma}\hat{n} = \frac{(1{,}96)^2 (1{,}41)^2}{(4)^2 (0{,}05)^2} = 190{,}92 \quad \delta\acute{\varepsilon}\nu\delta\rho\alpha$$

$${}_{\Delta}\hat{n} = \frac{(1{,}96)^2 (1{,}41)^2}{(12)^2 (0{,}05)^2} = 21{,}21 \quad \delta\acute{\varepsilon}\nu\delta\rho\alpha$$



4ο ΚΡΙΤΗΡΙΟ

Θέλουμε το διάστημα εμπιστοσύνης – όχι να είναι ποσοστό του μέσου – αλλά να έχει μια συγκεκριμένη τιμή



$$\left(z_{a/2} \cdot \sigma / \sqrt{n} \right) = h$$

$$n = \left(\frac{z_{a/2} \cdot \sigma}{h} \right)^2$$

Αν τα προκαταρκτικά δείγματα είναι μικρά (<30) [που συνήθως είναι] αντί για κανονική κατανομή έχω t κατανομή οπότε αντί για $Z_{a/2}$ βάζω $t_{a/2,n-1}$

και επιλύω με την μέθοδο των επανειλημμένων δοκιμών

$$\hat{n} = \left(\frac{t_{a/2,n-1} \cdot S_0}{h} \right)^2$$

$$n = \left(\frac{t_{a/2,n-1} \cdot S_0}{h} \right)$$

$$\hat{n} = \left(\frac{2,074 \cdot 22,9}{10} \right)^2 = \left(\frac{47,49}{10} \right)^2 = 4,749^2 = 22,55$$

X
78
74
53
74
93
127
99
127
76
104

Άθροισμα 905

$\bar{x} =$ 90.5

$S_0 =$ 22.9

h = 0,10

a= 0.05

$n_{δοκιμής}$	$t_{a/2,n-1}$	$\hat{n}_{εκτιμούμενο}$
5	……	……
8	……	……
12	……	……
15	……	……
20	……	……
23	2,074	22,55

Έστω ότι θέλουμε να εκτιμήσουμε τα μέσα ύψη των δένδρων σε τέσσερις συστάδες Α, Β, Γ, και Δ. Από τις προκαταρκτικές δειγματοληψίες, με μεγέθη δειγμάτων >30, έχουμε εκτιμήσεις για τα μέσα ύψη και τις τυπικές αποκλίσεις. Έστω ότι θέλουμε να εκφράσουμε την αξιοπιστία με το διάστημα εμπιστοσύνης να είναι h=0,8

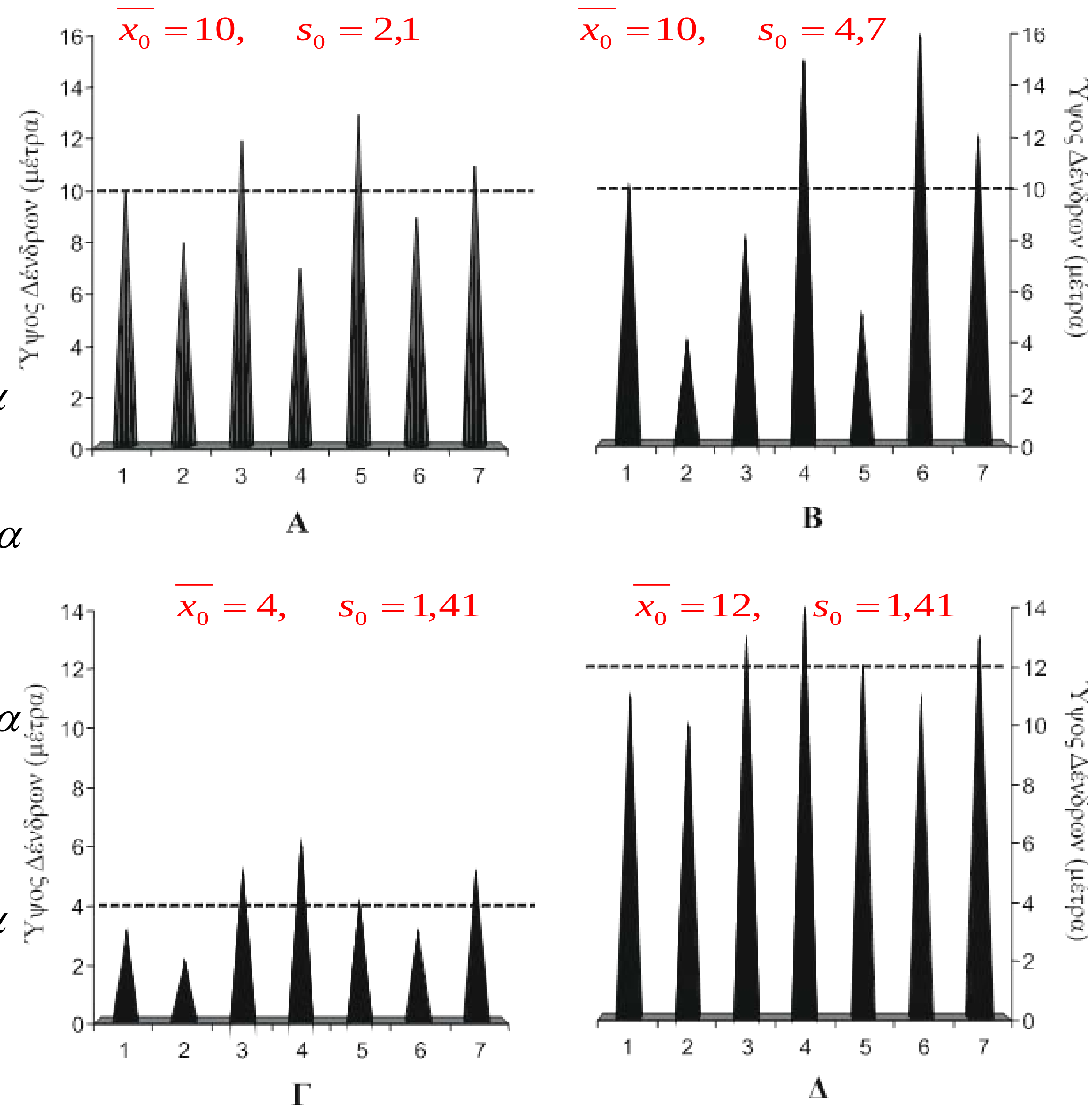
$$\hat{\mathbf{n}} = \left(\frac{\mathbf{z_{a/2}} \cdot s_0}{h} \right)^2$$

$${}_A\hat{n} = \frac{(1{,}96)^2 (2{,}1)^2}{(0{,}8)^2} = 26{,}47 \quad \delta \acute{\varepsilon} \nu \delta \rho \alpha$$

$${}_B\hat{n} = \frac{(1{,}96)^2 (4{,}7)^2}{(0{,}8)^2} = 132{,}59 \quad \delta \acute{\varepsilon} \nu \delta \rho \alpha$$

$${}_\Gamma\hat{n} = \frac{(1{,}96)^2 (1{,}41)^2}{(0{,}8)^2} = 11{,}93 \quad \delta \acute{\varepsilon} \nu \delta \rho \alpha$$

$${}_\Delta\hat{n} = \frac{(1{,}96)^2 (1{,}41)^2}{(0{,}8)^2} = 11{,}93 \quad \delta \acute{\varepsilon} \nu \delta \rho \alpha$$



Πίνακας 1.1. Αναγωγή σε Διάστημα Εμπιστοσύνης της αξιοπιστίας της εκτίμησης με βάση τα υπολογισθέντα μεγέθη του δείγματος.

Παράμετροι Συστάδων	(α) Τυπικό Σφάλμα $S_{\bar{x}} = 0,5$ μέτρα	(β) Συντελ. Παραλ-λακτικότητας C=0,06	(γ) Διάστημα Εμπιστοσύνης (i) $D.\mu = \pm 0,05.\mu$	(γ) Διάστημα Εμπιστοσύνης (ii) $h = \pm 0,8$
Α $_A S_0 = 2,1$ $_A \bar{x}_0 = 10$	$\hat{n} = 17,64$ 9,02 – 10,98	$\hat{n} = 12,25$ 8,82 – 11,17	$\hat{n} = 67,76$ 9,5 – 10,5	$\hat{n} = 26,47$ 9,2 – 10,8
Β $_B S_0 = 4,7$ $_B \bar{x}_0 = 10$	$\hat{n} = 88,36$ 9,02 – 10,98	$\hat{n} = 61,36$ 8,82 – 11,17	$\hat{n} = 339,44$ 9,5 – 10,5	$\hat{n} = 132,59$ 9,2 – 10,8
Γ $_\Gamma S_0 = 1,41$ $_\Gamma \bar{x}_0 = 4$	$\hat{n} = 7,95$ 3,02 – 4,98	$\hat{n} = 34,51$ 3,53 – 4,47	$\hat{n} = 190,92$ 3,8 – 4,2	$\hat{n} = 11,93$ 3,2 – 4,8
Δ $_\Lambda S_0 = 1,41$ $_\Delta \bar{x}_0 = 12$	$\hat{n} = 7,95$ 11,02 – 12,98	$\hat{n} = 3,83$ 10,59 – 13,41	$\hat{n} = 21,21$ 11,4 – 12,6	$\hat{n} = 11,93$ 11,2 – 12,8

Συγκρίνουμε τα μικρότερα εκτιμώμενα δείγματα κάθε περίπτωσης. Από δείγμα μεγέθους 12,2 για την συστάδα Α – που δίδει Συντελ. Παραλλακτ. 0,06 – προκύπτει μάλλον μεγάλο διάστημα (8,82-11,17) εμπιστοσύνης. Επίσης, το ίδιο διάστημα προκύπτει για την συστάδα Β, αν και το δείγμα είναι αρκετά μεγαλύτερο, προφανώς λόγω της μεγάλης διακύμανσης στον πληθυσμό. Για την συστάδα Γ, το μικρό (7,95) δείγμα το οποίο απαιτείται για τυπικό σφάλμα 0,5 που διερευνήσαμε, σαφώς δίδει μεγάλο σχετικά διάστημα εμπιστοσύνης. Παρόμοιο σχόλιο μπορεί να γίνει και για τον Συντελ. Παραλλακτικότητας, της συστάδας Δ, για την οποία το πολύ μικρό (3,83) δείγμα, δίδει μεγάλο εύρος του διαστήματος εμπιστοσύνης.

ΔΙΩΝΥΜΙΚΗ ΚΑΤΑΝΟΜΗ
Παρουσία-απουσία χαρακτηριστικού

$$f(x) = P(X = \chi) = \left(\frac{n!}{\chi!(n-\chi)!}\right) p^{\chi}(1-p)^{n-x}$$

Όπου η X παίρνει τιμές 0, 1, 2, 3.....n

Δίνει την πιθανότητα ένα γεγονός να πραγματοποιηθεί
χ φορές σε n ανεξάρτητες δοκιμές Bernoulli
Δηλαδή να έχουμε χ επιτυχίες και n-χ αποτυχίες

Η μέση τιμή $\mu = nP$
και η διασπορά $\sigma^2 = np(1-p)$

$\mu \Rightarrow p$

$\bar{x} \Rightarrow$

$V(\bar{x}) \Rightarrow V(\hat{p})$

$\bar{x}_0 \Rightarrow \hat{p}_0$

Τυπική απόκλιση

$$\sigma = \sqrt{np(1-p)}$$

Αν πάρουμε και εξετάσουμε τυχαία n φυτά από ένα πληθυσμό στον οποίο το ποσοστό προσβολής είναι p τότε η πιθανότητα X από τα n φυτά να είναι προσβεβλημένα ακολουθεί την διωνυμική κατανομή

Αν σημειώνουμε την παρουσία ή απουσία κάπου χαρακτηριστικού στην κάθε δειγματοληπτική μονάδα (π.χ. παρουσία – απουσία δάκου σε καρπούς ελιάς) τότε:
Ο αριθμός X των προσβεβλημένων δ.μ. στις n εξετασθείσες δ.μ. ακολουθεί την διωνυμική κατανομή.

ΜΑΣ ΕΝΔΙΑΦΕΡΕΙ ΤΟ <u>ΠΟΣΟΣΤΟ</u> ΤΟΝ “ΠΡΟΣΒΕΒΛΗΜΕΝΩΝ” ΔΕΙΓΜΑΤΟΛΗΠΤΙΚΩΝ ΜΟΝΑΔΩΝ

Η αναλογία φύλου,
Το ποσοστό ανηλίκων ατόμων σε ένα πληθυσμό,
Το ποσοστό μολυσμένων ζώων, φρούτων κ.λπ.
Ο αριθμός εμφάνισης μιας κατηγορίας (αλόφυτα, υποείδη, κ.λπ) σε ένα μεγάλο πληθυσμό.

$$V(\hat{p}) = \frac{p(1-p)}{n}$$

Θέλουμε ο συντελεστής παραλλακτικότητας CV να έχει κάποια συγκεκριμένη τιμή

$$CV(\hat{p}) = \frac{\sigma_{\hat{x}}}{\mu} = \frac{\sigma_{\hat{p}}}{p} = \frac{\sqrt{V(\hat{p})}}{p}$$

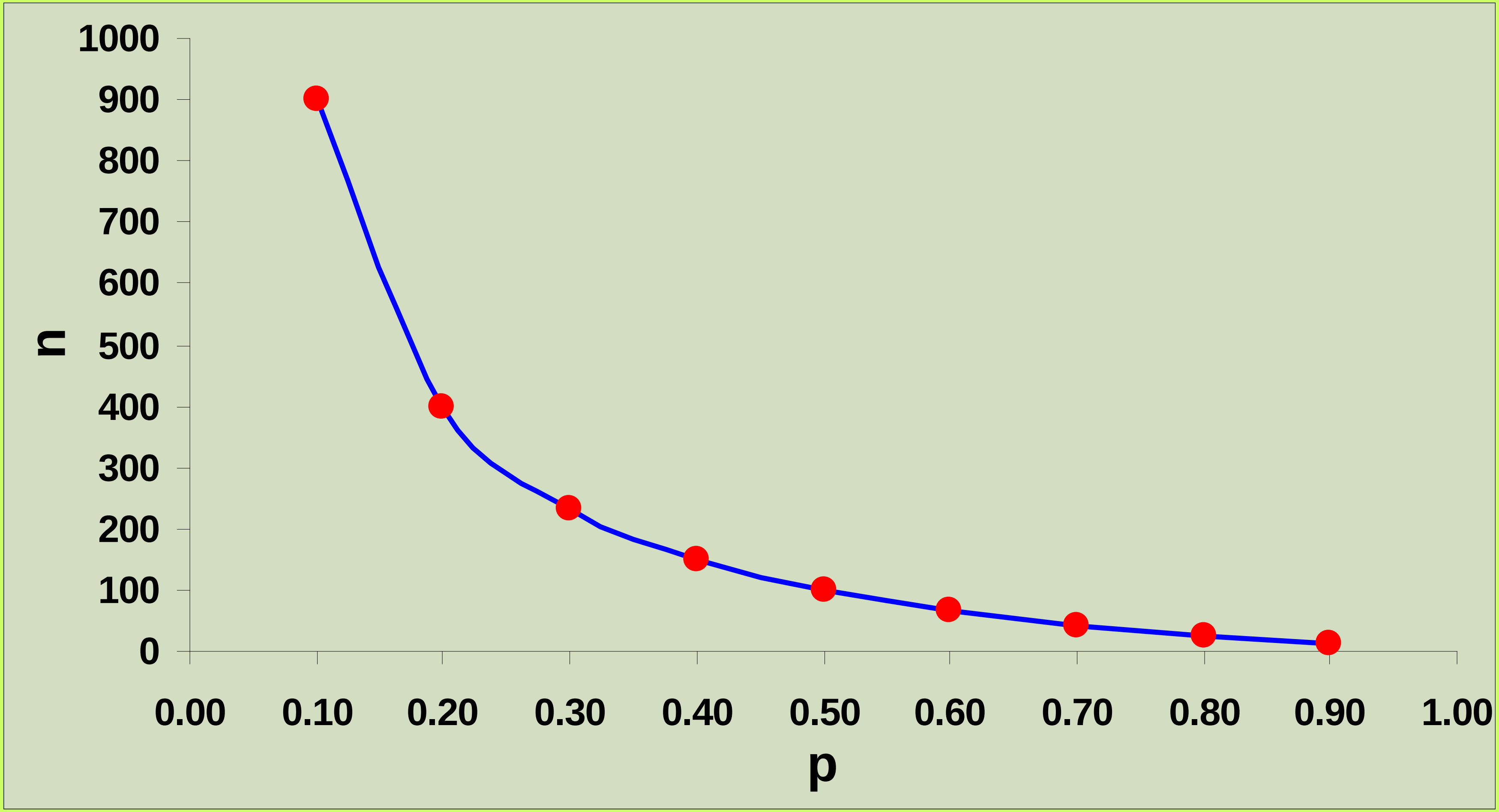
$$\hat{n} = \frac{(1-\hat{p}_0)}{\hat{p}_0 \cdot CV^2}$$

Ας υποθέσουμε ότι θέλουμε να εκτιμήσουμε την αναλογία φύλου σε ένα πληθυσμό ελαφιών.
από προκαταρκτική δειγματοληψία (ή προηγούμενη πληροφορία ή εκτίμηση) αναμένουμε τα άρρενα άτομα να είναι το 40% (δηλ. 0,4)
Αν έχουμε αμφιβολία παίρνουμε το 0,5 της τιμής που υποθέτουμε – αυτό μας κάνει τους υπολογισμούς πιο συντηρητικούς.

Θέλουμε η εκτίμησή μας να έχει συντελεστή παραλλακτικότητας CV = 0,20

$$n = \frac{1-0.40}{0.40 * 0.2^2} = 37$$

$$\hat{n} = \frac{(1-\hat{p}_0)}{\hat{p}_0 \cdot CV^2}$$



*Θέλουμε το διάστημα εμπιστοσύνης
να είναι ποσοστό το μέσου όρου*

$$Z_{a/2} * \sigma_{\hat{p}} = Dp \Rightarrow Z_{a/2} * \sqrt{\frac{p(1-p)}{n}} = Dp$$

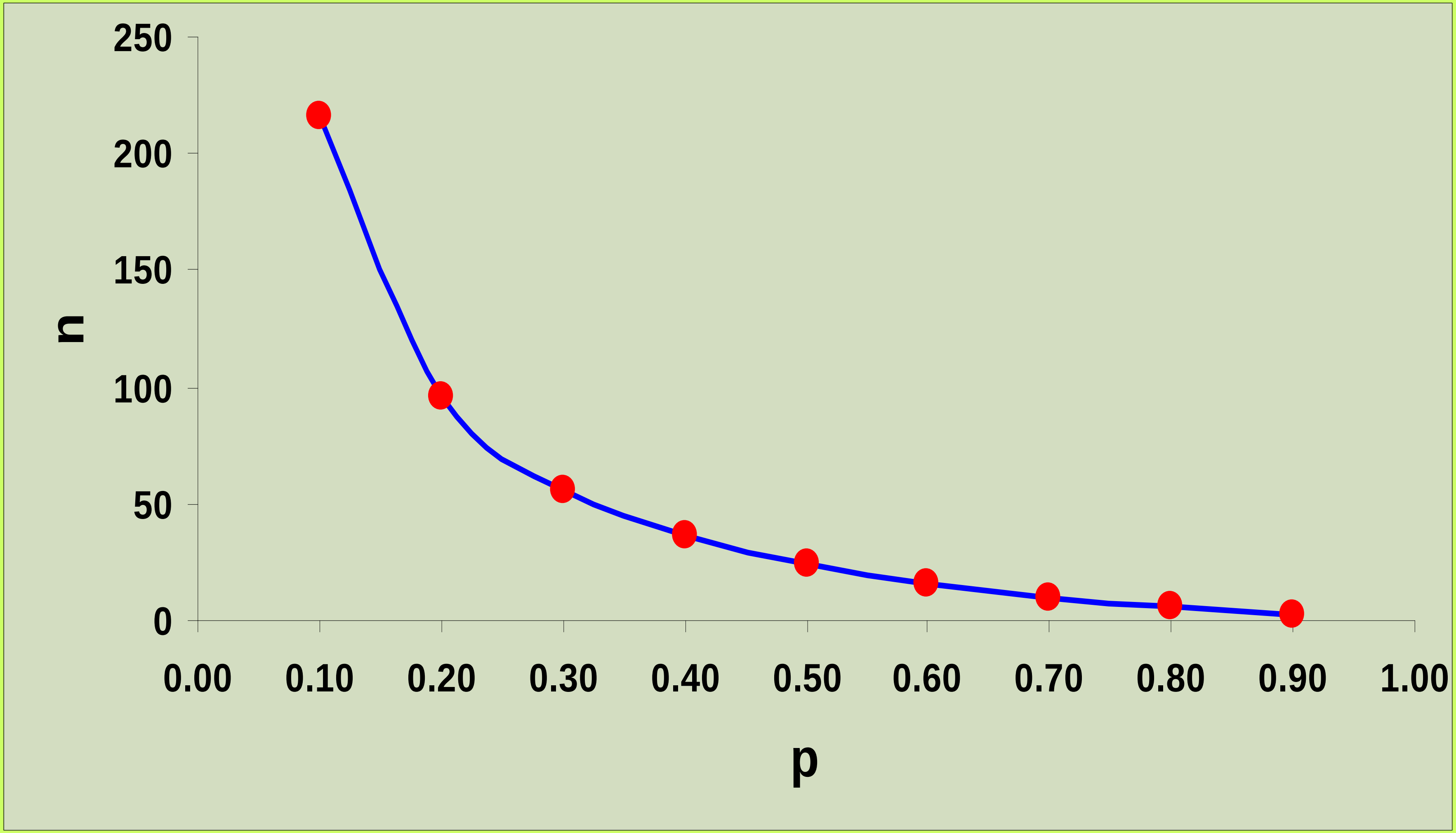
$$\hat{n} = \left(\frac{z_{a/2}}{D}\right)^2 * \left(\frac{1-\hat{p}_0}{\hat{p}_0}\right)$$

Ας υποθέσουμε ότι θέλουμε να εκτιμήσουμε την αναλογία φύλου σε ένα πληθυσμό ελαφιών.
από προκαταρκτική δειγματοληψία (ή προηγούμενη πληροφορία ή εκτίμηση) αναμένουμε τα άρρενα άτομα να είναι το 40% (δηλ. 0,4).
Αν έχουμε αμφιβολία παίρνουμε το 0,5 της τιμής που υποθέτουμε – αυτό μας κάνει τους υπολογισμούς πιο συντηρητικούς.

**Θέλουμε η ακρίβεια της εκτίμησή μας να είναι
D = ± 0,20 του μέσου.**

$$\mathbf{n} = \left(\frac{1.96}{0.20}\right) * \left(\frac{1 - 0.40}{0.40}\right) = 14.7$$

$$\hat{n} = \left(\frac{z_{a/2}}{D}\right)^2 * \left(\frac{1-\hat{p}_0}{\hat{p}_0}\right)$$



δηλαδή θέλουμε το διάστημα εμπιστοσύνης αλλά να έχει μια συγκεκριμένη τιμή

$$Z_{a/2} * \sigma_{\hat{p}} = h \Rightarrow Z_{a/2} * \sqrt{\frac{p(1-p)}{n}} = h$$

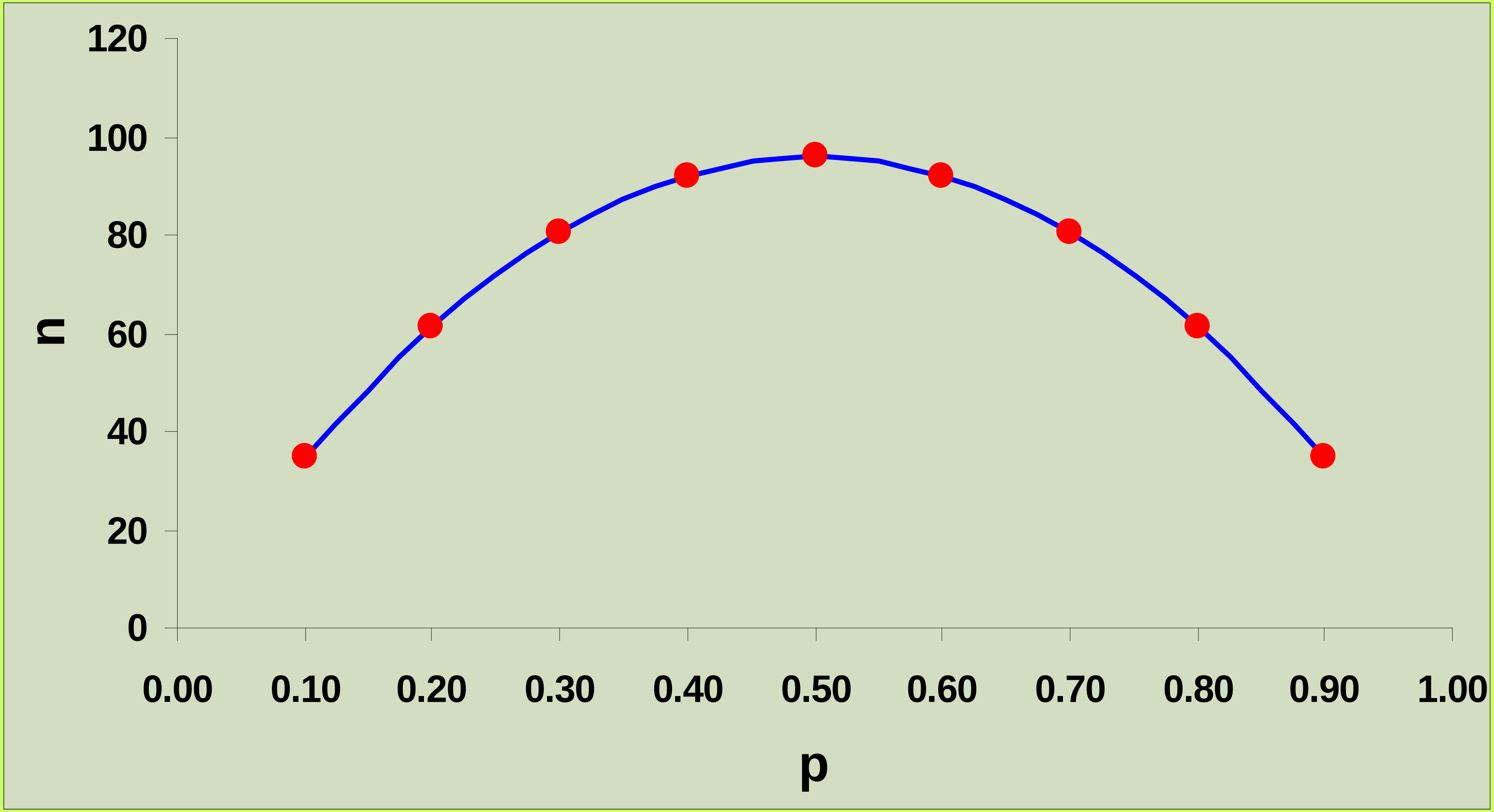
$$\hat{n} = \left(\frac{z_{a/2}}{h}\right)^2 * \hat{p}_0 (1-\hat{p}_0)$$

Ας υποθέσουμε ότι θέλουμε να εκτιμήσουμε την αναλογία φύλου σε ένα πληθυσμό ελαφιών.
από προκαταρκτική δειγματοληψία (ή προηγούμενη πληροφορία ή εκτίμηση) αναμένουμε τα άρρενα άτομα να είναι το 40% (δηλ. 0,4) .
Αν έχουμε αμφιβολία παίρνουμε το 0,5 της τιμής που υποθέτουμε – αυτό μας κάνει τους υπολογισμούς πιο συντηρητικούς.

Θέλουμε η εκτίμησή μας να έχει a = 0.05 δηλαδή $Z_{\alpha/2}$=1,96 και θέλουμε τα όρια του σφάλματος της εκτίμησής μας να είναι ± 0,02 (h)

$$\mathbf{n} = \left(\frac{1.96}{0.02}\right)^2 (0.4)(1-0.4) = 46$$

$$\hat{n} = \left(\frac{z_{a/2}}{h} \right)^2 * \hat{p}_0 (1 - \hat{p}_0)$$



ΚΑΤΑΝΟΜΗ POISSON
κατανομή των σπανίων γεγονότων

$$f(x) = P(X = \chi) = \left(\frac{e^{-\lambda}\lambda^{\chi}}{\chi!} \right)$$

Όπου η Χ παίρνει τιμές 0, 1, 2, 3.....n

Η μέση τιμή = διασπορά = $\mu = \sigma^2 = \lambda$

Τυπικό σφάλμα του μέσου $\sigma_{\hat{\lambda}} = \sqrt{\dfrac{\lambda}{n}}$

n ≧ 50 και συγχρόνως np < 5

ΠΑΡΑΔΕΙΓΜΑΤΑ poisson:
Περιγράφει την κατανομή του αριθμού των “hits”, των αφίξεων, των συμβάντων, στη μονάδα του χρόνου ή του χώρου ή της δειγματοληπτικής προσπάθειας όταν το κάθε συμβάν είναι ανεξάρτητο από τα υπόλοιπα.

Σημειώνουμε τον αριθμό νυγμάτων δάκου σε καρπούς ελιάς αλλά δεν ενδιαφερόμαστε για το ποσοστό των προσβεβλημένων ελαιοκάρπων (που θα ήταν διωνυμική) αλλά για το μέσο αριθμό νυγμάτων ανά ελιά (δ.μ.) Υποθέτουμε βέβαια ότι η χωροδιάταξη των γεγονότων είναι τυχαία, στην περίπτωσή μας δηλαδή δεν υπάρχει τάση στους δάκους να προτιμούν ελιές οι οποίες να έχουν ήδη νύγμα από άλλο δάκο. Στην αντίθετη περίπτωση η κατανομή θα ήταν η αρνητική διωνυμική.

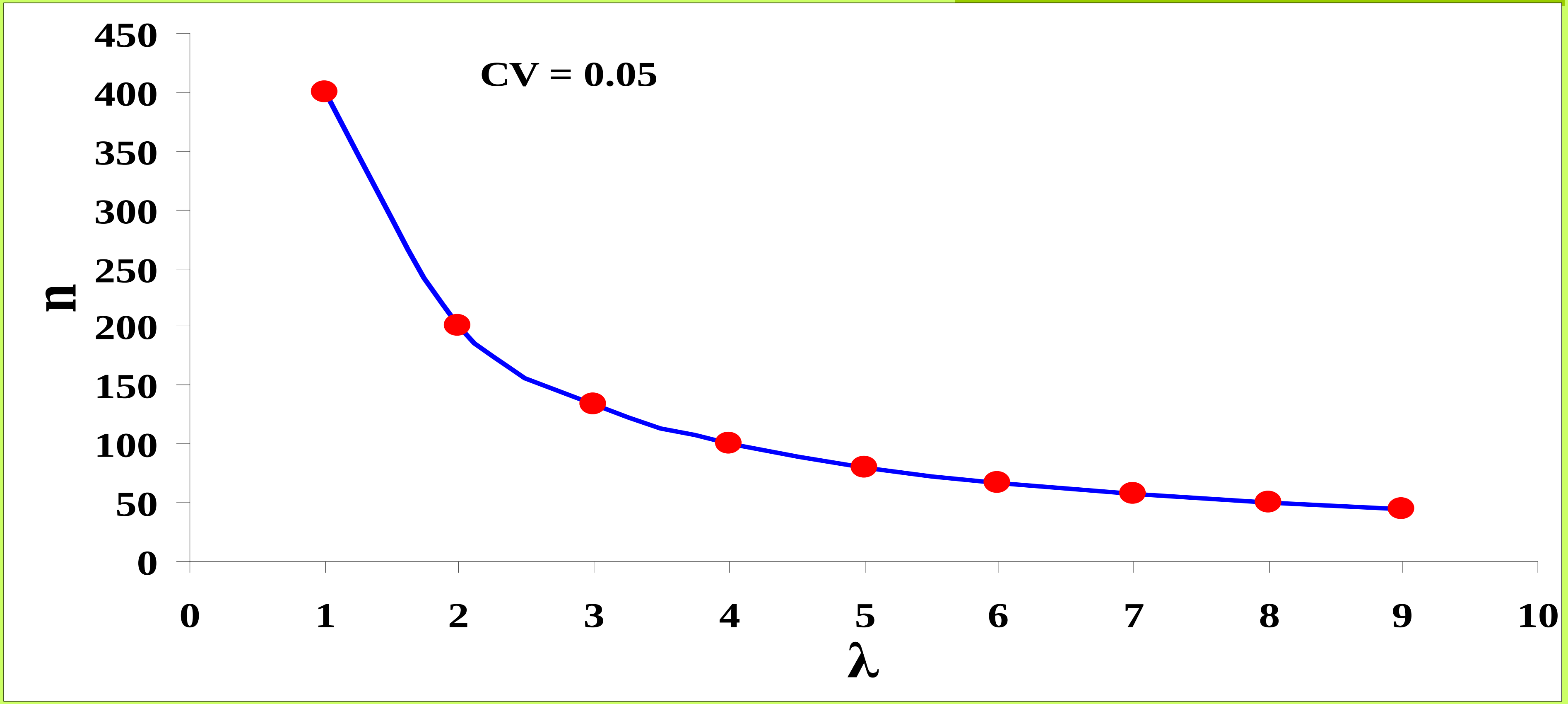
Χρησιμοποιείται για να εκφράσει την κατανομή του αριθμού των πελατών που φτάνουν στη μονάδα του χρόνου, τον αριθμό των ασφαλιστικών αξιώσεων, ο αριθμός των λαμβανόμενων τηλεφωνικών κλίσεων στη μονάδα του χρόνου, ο αριθμός εκπεμπόμενων σωματιδίων, κ.λπ.
Επίσης τον αριθμό των σπόρων, φυταρίων που υπάρχουν σε αγροτεμάχιο, ο αριθμός των απογόνων που γεννιούνται σε μια σεζόν (αυτό μπορεί να περιγραφεί και με την διωνυμική), αριθμός των θηραμάτων που θηρεύονται στη μονάδα του χρόνου, κ.λπ.
Όλα αυτά μπορούν να προσεγγιστούν με την κανονική κατανομή αν η μέση τιμή είναι αρκετά μεγάλη.

Ας υποθέσουμε ότι θέλουμε να εκτιμήσουμε το μέγεθος του δείγματος *(αριθμό φωλεών)* που πρέπει να εξετάσουμε για να υπολογίσουμε το μέσο αριθμό αυγών σε φωλιές ψαρονιών. Από προκαταρκτική δειγματοληψία αναμένουμε ο μέσο αριθμό αυγών ανά φωλιά να είναι 6.

Θέλουμε η εκτίμησή μας να έχει συντελεστή παραλλακτικότητας CV = 0,10

$$\hat{n} = \frac{1}{\hat{\lambda}_0 \cdot CV^2}$$

$$\mathbf{n = \frac{1}{6 * 0.10^2} = 16.6}$$

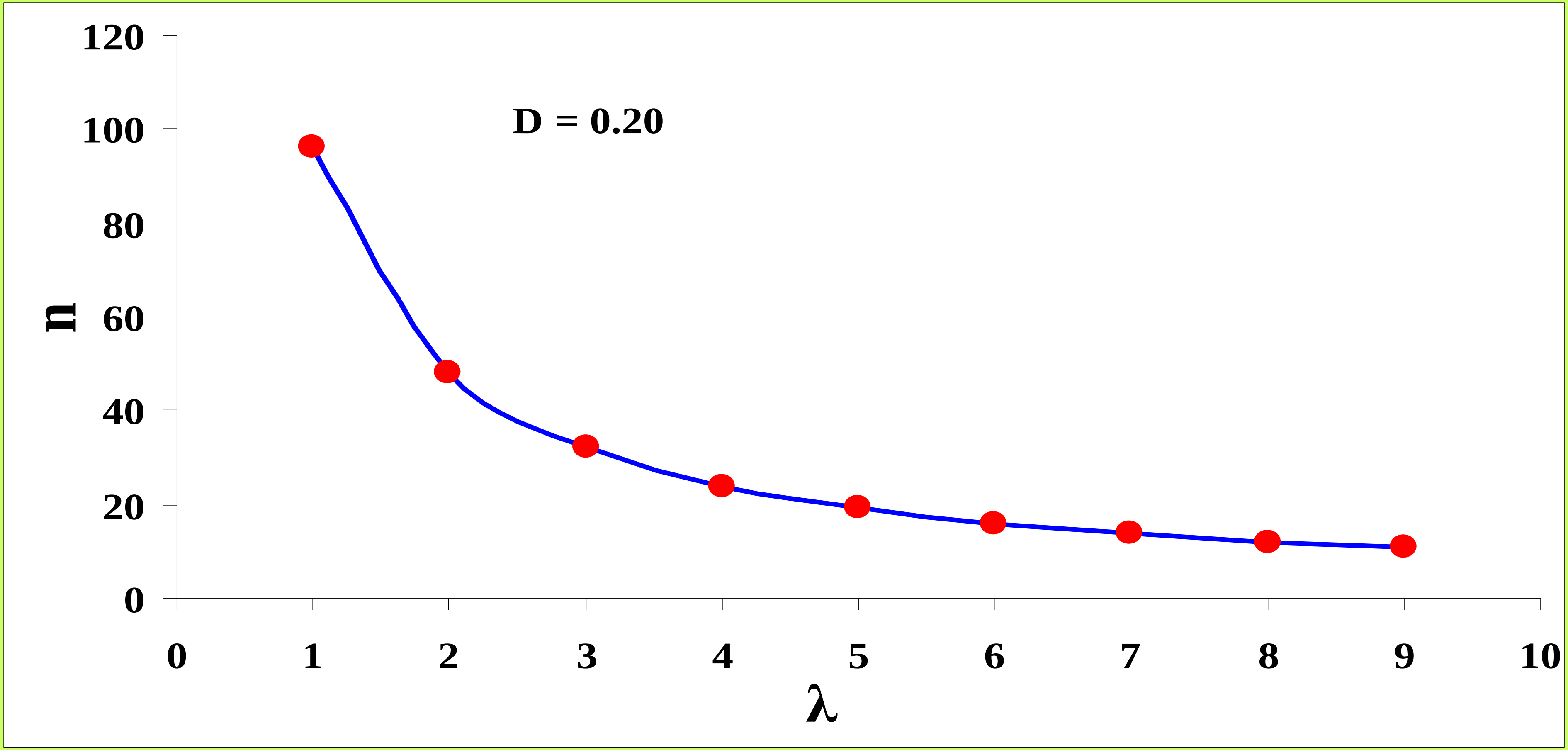


Ας υποθέσουμε ότι θέλουμε να εκτιμήσουμε το μέγεθος του δείγματος *(αριθμό φωλεών)* που πρέπει να εξετάσουμε για να υπολογίσουμε το μέσο αριθμό αυγών σε φωλιές ψαρονιών. Από προκαταρκτική δειγματοληψία αναμένουμε ο μέσο αριθμό αυγών ανά φωλιά να είναι 6.

$$\hat{n} = \left(\frac{z_{a/2}}{D} \right)^2 * \frac{1}{\hat{\lambda}}$$

Θέλουμε η ακρίβεια της εκτίμησή μας να είναι D = ±0,20 του μέσου.

$$n = \left(\frac{1.96}{0.20} \right)^2 \frac{1}{6} = 16$$

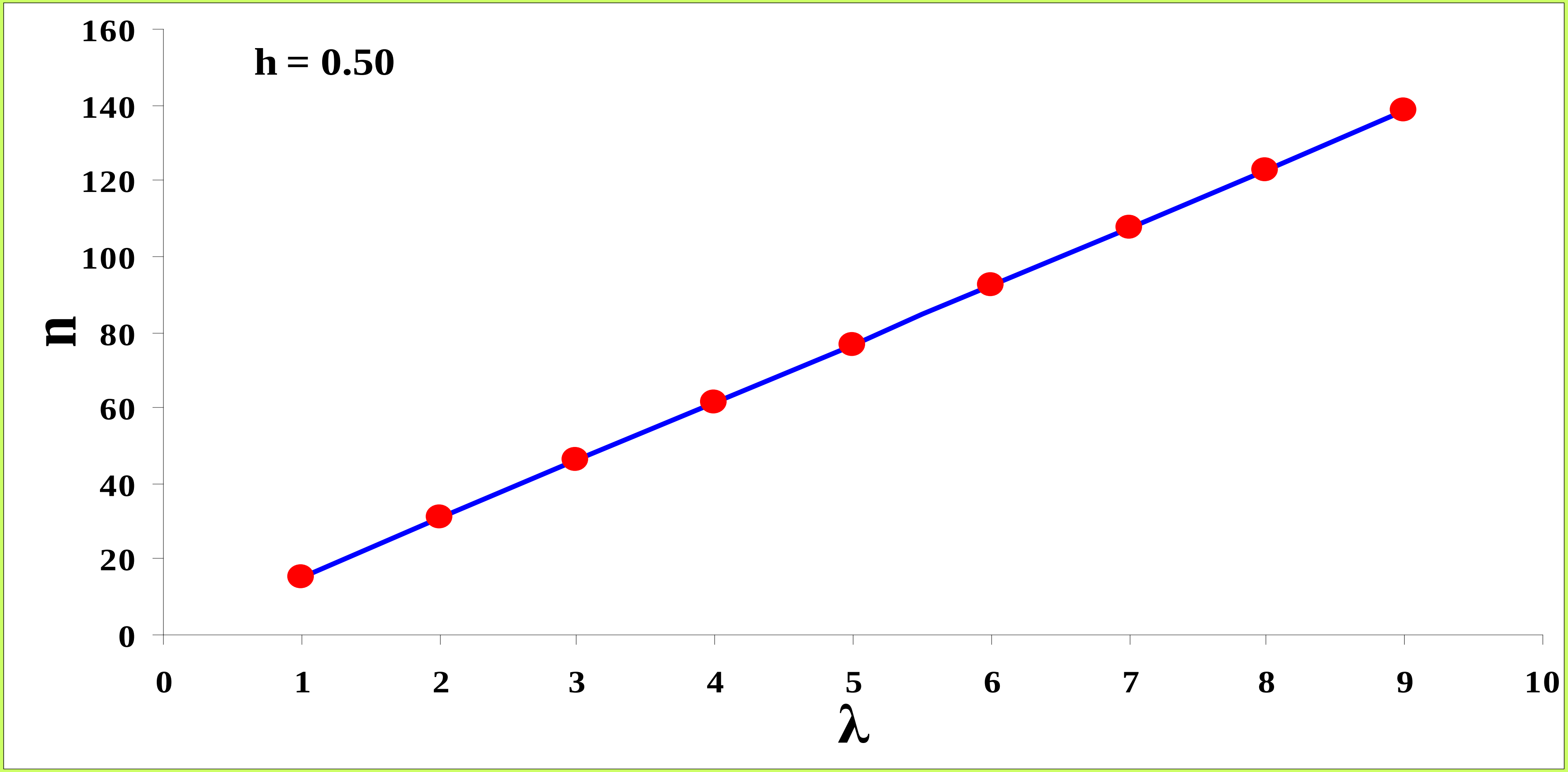


Ας υποθέσουμε ότι θέλουμε να εκτιμήσουμε το μέγεθος του δείγματος *(αριθμό φωλεών)* που πρέπει να εξετάσουμε για να υπολογίσουμε το μέσο αριθμό αυγών σε φωλιές ψαρονιών. Από προκαταρκτική δειγματοληψία αναμένουμε ο μέσο αριθμό αυγών ανά φωλιά να είναι 6.

$$\hat{n} = \frac{z_{a/2}}{h} * \hat{\lambda}_0$$

Θέλουμε η εκτίμησή μας να έχει a = 0.05 ($Z_{a/2}$=1,96) και θέλουμε τα όρια του σφάλματος να είναι h = ±0,20

$$\mathbf{n = \frac{1.96}{0.20} * 6 = 58.8}$$



ΑΡΝΗΤΙΚΗ ΔΙΩΝΥΜΙΚΗ ΚΑΤΑΝΟΜΗ

Δίνει τον αριθμό των αποτυχιών πριν από την χ-οστή επιτυχία σε n ανεξάρτητες δοκιμές Bernoulli

Είναι γενίκευση της **γεωμετρικής κατανομής** *που δίνει τον αριθμό των αποτυχιών μέχρι την πρώτη επιτυχία.*
Ουσιαστικά πρόκειται για την κατανομή **Pascal** *(όταν το χ είναι ασυνεχής κατανομή) η οποία λέγεται αρνητική διωνυμική όταν το χ είναι συνεχής μεταβλητή.*
Πολλοί χρησιμοποιούν τον ίδιο όρο και για τις δύο περιπτώσεις.

$$f(x) = P(X = \chi) = \binom{x-1}{r-1} p^r q^{x-r}$$

Όπου η Χ παίρνει τιμές 0, 1, 2, 3.....n

Η μέση τιμή = μ
Διασπορά = V(x)

$$\mu \approx \overline{X} \qquad \sigma^2 \approx V(x) = \mu + \frac{\mu^2}{K}$$

Όσο μεγαλύτερος είναι ο βαθμός ομαδοποίησης τόσο μικρότερο είναι το K

Οπως και η Διωνυμική περιγράφει δυαδικά δεδομένα (μαύρο/άσπρο, κορώνα/γράμματα) κ.λ.π. τα οποία έχουν την ίδια πιθανότητα επιτυχίας αλλά αντί να μετράμε τον αριθμό των επιτυχιών μετράμε τον αριθμό των προσπαθειών που απαιτούνται μέχρι να φτάσουμε ένα προκαθορισμένο αριθμό επιτυχιών.

Στην Οικολογία είναι σημαντικότερο να κατανοήσουμε ότι η Αρνητική Διωνυμική είναι μια διακριτή κατανομή, όπως η Διωνυμική και η Poisson, αλλά η διασπορά της μπορεί να είναι μεγαλύτερη από τη μέση της τιμή.
Περιγράφει ικανοποιητικά ομαδοποιημένα δεδομένα.

ΠΑΡΑΔΕΙΓΜΑΤΑ

Ο αριθμός των τετραγώνων που πρέπει να εξετάσουμε έως ότου συμπληρώσουμε ένα αριθμό ειδών.
Όταν ο αριθμός των παρασίτων σε ένα ξενιστή (δειγματοληπτική μονάδα: φυτό, η ζώο) είναι ομαδοποιημένη, δηλαδή όταν υπάρχει μεγάλη συγκέντρωση παρασίτων σε κάποιους ξενιστές και πολύ μικρή η καθόλου παρουσία παρασίτων σε κάποιους άλλους.
Η κατανομή των νηματωδών ακολουθεί επίσης της Διωνυμική κατανομή

Θέλουμε ο συντελεστής παραλλακτικότητας CV να έχει κάποια συγκεκριμένη τιμή

$$\mathbf{n} = \frac{\sigma^2}{\mu^2 C^2} \qquad \hat{n} = \frac{\overline{X}_0 + \hat{K}_0}{\overline{X}_0 \hat{K}_0 C^2}$$

Θέλουμε το διάστημα εμπιστοσύνης να είναι ποσοστό **D** *του μέσου όρου*

$$n = \left(\frac{z_{a/2} \cdot \sigma}{D \cdot \mu} \right)^2 \qquad \hat{n} = \frac{Z_{a/2}^2 (\overline{X}_0 + \hat{K}_0)}{\overline{X}_0 \hat{K}_0 D^2}$$

Θέλουμε το διάστημα εμπιστοσύνης να έχει μια συγκεκριμένη τιμή ***h***

$$n = \left(\frac{z_{a/2} \cdot \sigma}{h} \right)^2 \qquad \hat{n} = \frac{Z_{a/2}^2 \overline{X}_0 (\hat{K}_0 + \overline{X}_0)}{\hat{K}_0 h^2}$$

Θέλουμε να βρούμε το βέλτιστο μέγεθος δείγματος για την εκτίμηση του μέσου αριθμού αφίδων σε βλαστούς φυτών φασολιού.

Αριθμός αφίδων ανα βλαστό	Αριθμός βλαστών	Ποσοστό
0	6	0.12
1	8	0.16
2	9	0.18
3	6	0.12
4	6	0.12
5	2	0.04
6	5	0.1
7	3	0.06
8	1	0.02
9	4	0.08
	50	

$$\Sigma XF_x = 173$$

$$\overline{X} = 173/50 = 3.46$$

$$s^2 = 7.356$$

$$\sigma^2 \approx V(x) = \mu + \frac{\mu^2}{K}$$

$$K \approx \frac{\overline{x^2}}{s^2 - \overline{x}} = \frac{3.46^2}{7.356 - 3.46} = 3.07$$

Θέλουμε να βρούμε το βέλτιστο μέγεθος δείγματος για την εκτίμηση του μέσου αριθμού αφίδων σε βλαστούς φυτών φασολιού.

Αριθμός αφίδων ανα βλαστό	Αριθμός βλαστών	Ποσοστό
0	6	0.12
1	8	0.16
2	9	0.18
3	6	0.12
4	6	0.12
5	2	0.04
6	5	0.1
7	3	0.06
8	1	0.02
9	4	0.08
	50	

$$\overline{X} = 3.46$$

$$s^2 = 7.356$$

$$K \approx 3.07$$

$$C = 0.15$$

$$\hat{n} = \frac{\overline{X}_0 + \hat{K}_0}{\overline{X}_0 \hat{K}_0 C^2} = \frac{3.46 + 3.07}{3.46 * 3.07 * 0.15^2} = \frac{10.44}{0.23} = 45$$

Θέλουμε να βρούμε το βέλτιστο μέγεθος δείγματος για την εκτίμηση του μέσου αριθμού αφίδων σε βλαστούς φυτών φασολιού.

Αριθμός αφίδων ανα βλαστό	Αριθμός βλαστών	Ποσοστό
0	6	0.12
1	8	0.16
2	9	0.18
3	6	0.12
4	6	0.12
5	2	0.04
6	5	0.1
7	3	0.06
8	1	0.02
9	4	0.08
	50	

$$\overline{X} = 3.46$$

$$s^2 = 7.356$$

$$K \approx 3.07$$

$$D = 0.20$$

$$\hat{n} = \frac{Z_{a/2}^2(\overline{X}_0 + \hat{K}_0)}{\overline{X}_0 \hat{K}_0 D^2} = \frac{1.96^2 * (3.46 + 3.07)}{3.46 * 3.07 * 0.2^2} = 89$$

Θέλουμε να βρούμε το βέλτιστο μέγεθος δείγματος για την εκτίμηση του μέσου αριθμού αφίδων σε βλαστούς φυτών φασολιού.

Αριθμός αφίδων ανα βλαστό	Αριθμός βλαστών	Ποσοστό
0	6	0.12
1	8	0.16
2	9	0.18
3	6	0.12
4	6	0.12
5	2	0.04
6	5	0.1
7	3	0.06
8	1	0.02
9	4	0.08
	50	

$$\overline{X} = 3.46$$

$$s^2 = 7.356$$

$$K \approx 3.07$$

$$h = 0.20$$

$$\hat{n} = \frac{Z^2_{a/2}\overline{X}_0(\hat{K}_0 + \overline{X}_0)}{\hat{K}_0 h^2} = \frac{1.96^2 * 3.46 * (3.07 + 3.46)}{3.07 * 0.20^2}$$