Εκτιμητική

- **Διαστήματα Εμπιστοσύνης (ακριβή και ασυμπτωτικά)**
  - για τη μέση τιμή κανονικής (με γνωστή ή άγνωστη διασπορά)
  - για τη διασπορά κανονικής
  - για τη μέση τιμή οιασδήποτε κατανομής

    όταν το μέγεθος τους δείγματος είναι μεγάλο (πχ. για ποσοστό)

Εκτιμητική - Διαστήματα Εμπιστοσύνης

Έχουμε κάποιες παρατηρήσεις και προσπαθούμε να προσδιορίσουμε το άγνωστο μοντέλο της κατανομής F η οποία ανήκει σε μια κλάση κατανομών $\mathcal{F}$

- Γνωρίζουμε ότι ο αριθμός των ατυχημάτων σε ένα χρονικό διάστημα ακολουθεί Poisson (λ).

Αν τα δεδομένα που έχω προέρχονται από ατυχήματα είναι λογικό να υποθέσω ότι προέρχονται από την Poisson οικογένεια αλλά δεν ξέρω την τιμή της παραμέτρου λ. Θέλω να εκτιμήσω (να προσεγγίσω όσο το δυνατόν ακριβέστερα) την πραγματική τιμή του λ

Σημειακή Εκτιμήτρια Αντιστοιχούν συγκεκριμένη τιμή στο λ. Μια τέτοια εκτιμήτρια συμβολίζεται $\hat{\lambda}=3$

Διάστημα Εμπιστοσύνης Διάστημα στο οποίο βρίσκεται η πραγματική τιμή του λ με κάποια συγκεκριμένη πιθανότητα $P\left(T_1(\underset{\sim}{X}) \leqslant \lambda \leqslant T_2(\underset{\sim}{X})\right) = 0.95$

Χρησιμοποιώ τις στατιστικές συναρτήσεις (σ.σ.) για να εκτιμήσω τις άγνωστες παραμέτρους της κατανομής την οποία ακολουθεί το χαρακτηριστικό που μελετώ. Συνήθως, εαν $E(X_i)=\mu$ και $Var(X_i)=\sigma^2$, θέλω τουλάχιστον να έχω μια εκτίμηση για την πληθυσμιακή μέση τιμή μ

$\hat{\mu}=\bar{X}$, όπου η σ.σ. $\bar{X}$ είναι ο εκτιμητής της άγνωστης παραμέτρου μ και για οποιοδήποτε δείγμα αυτός θα δώσει μια συγκεκριμένη τιμή π.χ. εαν $\bar{X}=0.94$ τότε $\hat{\mu}=0.94$ αυτή είναι μια εκτίμηση του μ. (αυτή είναι σημειακή εκτίμηση).

$$\widehat{\sigma^2} = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2 \quad \ldots$$

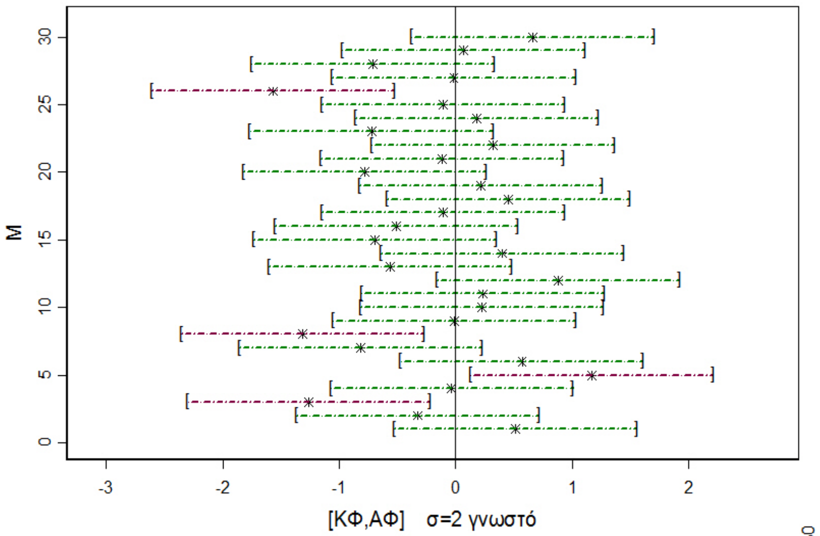
η δειγματική διασπορά S^2 είναι ένας εκτιμητής της πληθυσμιακής διασποράς, σ^2. Μάλιστα αυτός είναι "αμερόληπτος" $E(S^2)=\sigma^2$.
Με διασπορά που όσο μεγαλώνει το μέγεθος του δείγματος n, αυτή μικραίνει. Έτσι οι τιμές του S^2 "συγκεντρώνονται" για μεγάλα δείγματα γύρω από την άγνωστη παράμετρο σ^2.

Διαστήματα Εμπιστοσύνης

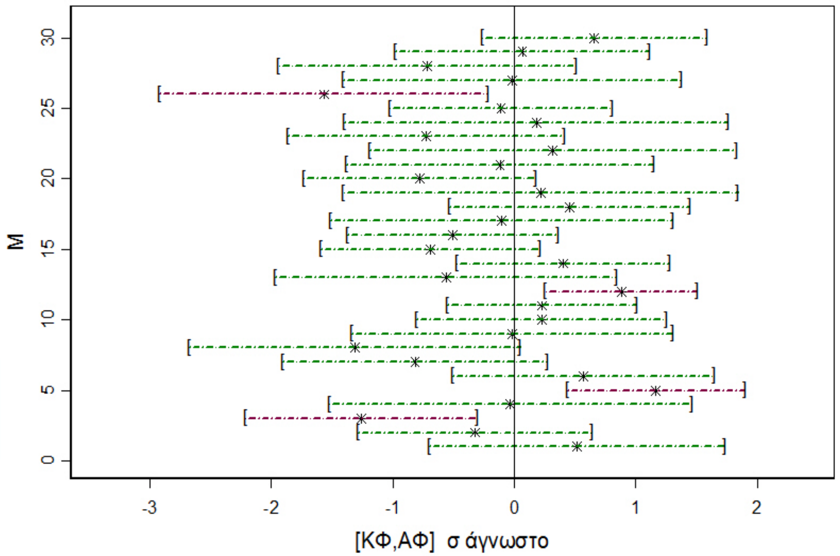
σ.ε. 90%(=1-α, α=0.1, α/2=0.05) για την μέση τιμή με γνωστή (σ=2) και άγνωστη διασπορά

Χρησιμοποιήθηκαν 30 δείγματα, μεγέθους n=10 έκαστον, από κανονική κατανομή $N(\mu=0,\sigma^2=2^2=4)$

και τα ποσοστιαία σημεία: από την τυπική κανονική $z_{0.05}=1.645$ και από την Student (με n-1=9 β.ε.) $t_{9,0.05}=t_9(0.05)=1.833$



$$\left[\overline{X}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \overline{X}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$



$$\left[\overline{X}-t_{n-1}(\alpha/2)\frac{S}{\sqrt{n}},\ \overline{X}+t_{n-1}(\alpha/2)\frac{S}{\sqrt{n}}\right]$$

Διαστήματα εμπιστοσύνης

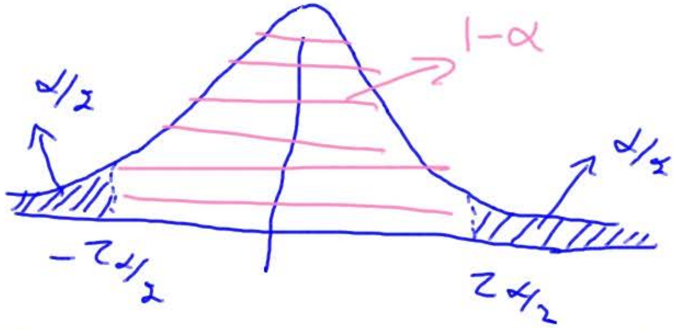
	Παράμετρος πληθυσμού	Προϋποθέσεις	Μέγεθος δείγματος	$100(1-a)\%$ δ.ε.
Κανονική: $N(\mu,\ \sigma^2)$	μ	σ^2 γνωστό	οτιδήποτε	$\overline{x} \pm \frac{\sigma}{\sqrt{n}} z_{a/2}$
		σ^2 άγνωστο	$n \leq 30$	$\overline{x} \pm \frac{s}{\sqrt{n}} t_{n-1;\ a/2}$
τυχαία κατανομή $EX = \mu \quad VarX = \sigma^2$		σ^2 άγνωστο	$n \geq 30$	$\overline{x} \pm \frac{s}{\sqrt{n}} z_{a/2}$
Κανονική: $N(\mu,\ \sigma^2)$	σ^2			$\left(\frac{(n-1)s^2}{\mathcal{X}^2_{n-1;\ a/2}}, \frac{(n-1)s^2}{\mathcal{X}^2_{n-1;\ 1-a/2}} \right)$
Bernoulli: Bin(1,p)	p		$n \geq 30$	$\hat{p} \pm z_{a/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ όπου $\hat{p} = \overline{x}$
			$n < 30$	άβακες

Διαστήματα εμπιστοσύνης (για το $\mu := E(X_i)$ πληθυσμιακός μέσος)

α. $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ και σ^2 γνωστό

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \qquad \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

ποσότητα οδηγό (περιέχει την άγνωστη παράμετρο μ) αλλά η κατανομή της δεν εξαρτάται από το άγνωστο μ.



Ορίζω ένα σημείο z_α τέτοιο ώστε η $P(Z > z_\alpha) = \alpha$ όπου $Z \sim N(0,1)$.

(είναι το άνω α ποσοστιαίο σημείο της $N(0,1)$)

$$P(Z \leq z_\alpha) = 1-\alpha \Rightarrow \Phi(z_\alpha) = 1-\alpha \Rightarrow z_\alpha = \Phi^{-1}(1-\alpha)$$

$$\Phi(z_{\alpha/2}) = 1-\alpha/2 \Rightarrow z_{\alpha/2} = \Phi^{-1}(1-\alpha/2)$$

$$P\left(-z_{\alpha/2} \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right) = 1-\alpha \Rightarrow P\left(-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \bar{X}-\mu \leq z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

$$P\left(\bar{X} - z_{\alpha/2}\,\sigma/\sqrt{n} \leq \mu \leq \bar{X} + z_{\alpha/2}\,\sigma/\sqrt{n}\right) = 1-\alpha$$

$\left[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \ , \ \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$ διάστημα εμπιστοσύνης συντελεστού εμπιστοσύνης $1-\alpha$.

ο οποίος καθορίζει το εύρος ($2z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$) του διαστήματος

σημειακή εκτίμηση για το μ.

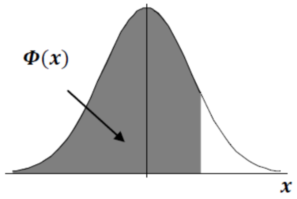
Το εύρος εξαρτάται από το σ^2 και το μέγεθος του δείγματος n.

Συνήθως χρησιμοποιούμε ως $1-\alpha = 0.95 \Rightarrow \alpha = 0.05$ (σ.ε. 95%)

Κατασκευάσαμε ένα δ.ε. ίσων ουρών.

Πίνακας τιμών της σ.κ. Φ της τυπικής κανονικής κατανομής (x = 0, .01, .02, ..., 3.49)

$$\Phi(x) = P(X \le x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$



x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5674	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.6315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	$\boxed{.9495}$	$\boxed{.9505}$	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	$\boxed{.9750}$	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990

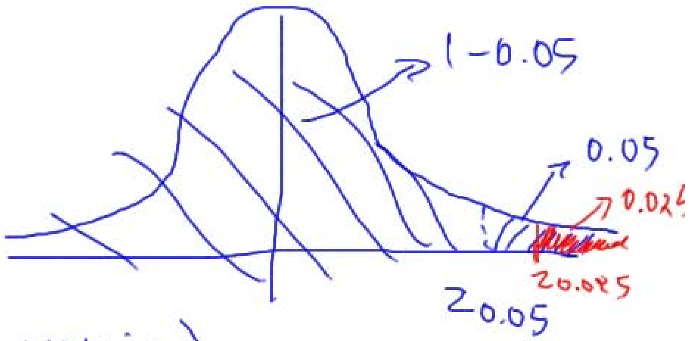
$a = 0.1$ δ.ε. σ.ε 90% $\quad z_{a/2} = z_{0.05}$

περίπου

$\Phi(1.645) = 0.95$

$\Rightarrow z_{0.05} = 1.645$

$\Phi(z_{0.05}) = 1 - 0.05 = 0.95$



($z_{0.25}$ είναι το τρίτο τεταρτημόριο της κατανομής)

$z_{0.75}$ είναι το πρώτο τεταρτημόριο } της κατανομή

$z_{0.5}$ είναι η διάμεσος

$\alpha = 0.05$ δ.ε. σ.ε 95% $\quad z_{a/2} = z_{0.025}$

$z_{0.05} < z_{0.025}$

$\Phi(1.96) = 0.975$

$\Rightarrow z_{0.025} = 1.96$

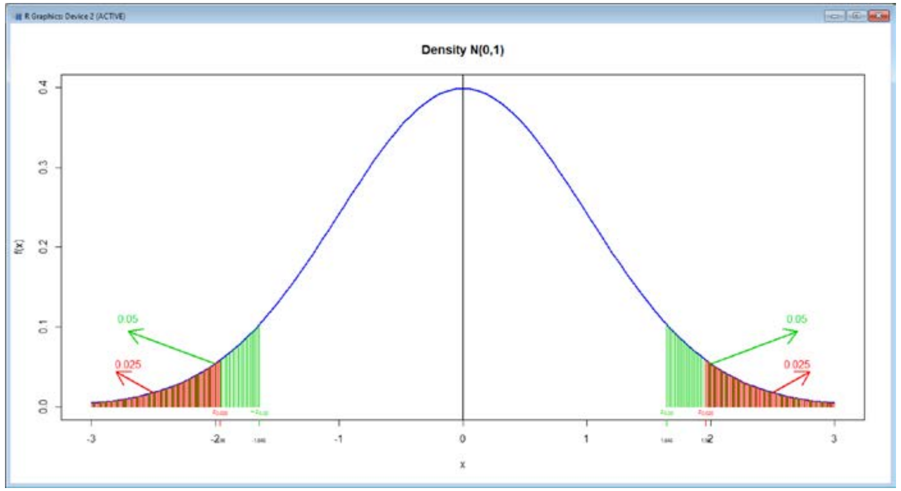
$\Phi(z_{0.025}) = 1 - 0.025 = 0.975$

$\Phi(z_{0.05}) = 1 - 0.05 = 0.95$

$\Phi(1.645) = 0.95$

```
> qnorm(0.95)
[1] 1.644854
```

```
> qnorm(1-0.05)
[1] 1.644854
```



$\Phi(z_{0.025}) = 1 - 0.025 = 0.975$

$\Phi(1.96) = 0.975$

```
> qnorm(0.975)
[1] 1.959964
```

```
> qnorm(1-0.025)
[1] 1.959964
```

$z_{0.25}$ είναι το τρίτο τεταρτημόριο

```
> qnorm(1-0.25)
[1] 0.6744898
```

$z_{0.75}$ είναι το πρώτο τεταρτημόριο

```
> qnorm(1-0.75)
[1] -0.6744898
```

$z_{0.5}$ είναι η διάμεσος

```
> qnorm(1-0.5)
[1] 0
```

Έστω δείγμα μεγέθους $n = 10$ από αντιδραστήρια με βάρη(σε gr) x_1=53, x_2=69, x_3=62, x_4=78, x_5=81, x_6=55, x_7=66, x_8=62, x_9=74, x_{10}=60. Εάν οι X_i, $i = 1 \dots n$ ακολουθούν N(μ, 102), να βρεθεί διάστημα εμπιστοσύνης(δ.ε.) συντελεστού εμπιστοσύνης(σ.ε) 90% για το μ.

$$\hat{\mu} = \frac{x_1 + \dots + x_{10}}{10} = \frac{53 + \dots + 60}{10} = \frac{660}{10} = 66$$

εκτίμηση

$n=10 \qquad 1-\alpha = 90\% \Rightarrow \alpha = 0.1 \qquad z_{0.05} = 1.645$

$$\left[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \ \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] \qquad P(LB \le \mu \le UB) = 1-\alpha$$

LB UB

διασπορά $\bar{X}$

$$\frac{\sigma^2}{n} = \frac{102}{10}$$

$$\left[66 - 1.645 \times \sqrt{\frac{102}{10}}, \ 66 + 1.645 \times \sqrt{\frac{102}{10}}\right]$$

$$[60.74, \ 71.25]$$

$$\frac{\sigma}{\sqrt{n}} = \sqrt{\frac{\sigma^2}{n}}$$

<u>Παρατήρηση</u>: Το διάστημα αυτό $[60.74, 71.25]$ είτε περιέχει είτε δεν περιέχει την πραγματική τιμή του μ (δεν θα το μάθουμε ποτέ)
Ελπίζουμε ότι είναι κάποιο από τα "90 στα 100 δείγματα", λόγω του συντελεστού εμπιστοσύνης 90%, που περιέχουν το μ.

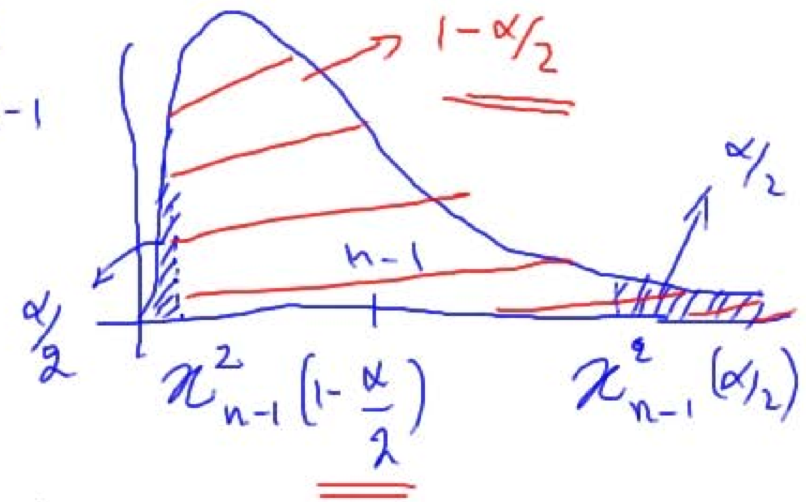
Αν θέλω δ.ε. σ.ε. 95% $\quad z_{0.025} = 1.96$

$$66 \pm 1.96 \times \sqrt{\frac{102}{10}} = 66 \pm 6.25 = [59.74 \;,\; 72.25]$$

έχει μεγαλύτερο εύρος από το δ.ε. σ.ε. 90% $[60.74 \;,\; 71.25]$

Θα ήθελα να έχει μικρότερο εύρος
αυτό επιτυγχάνεται αυξάνοντας το n.

$$S^* = (n-1)\frac{S^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(X_i-\bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}$$



$$E(S^*) = n-1 \qquad var(S^*) = 2(n-1)$$

$$P\left(\chi^2_{n-1}(1-\alpha/2) \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{n-1}(\alpha/2)\right) = 1-\alpha$$

$$\hat{\mu} = \bar{X}$$

$$P\left(\frac{(n-1)}{\chi^2_{n-1}(\alpha/2)} S^2 \le \sigma^2 \le \frac{(n-1)}{\chi^2_{n-1}(1-\alpha/2)} S^2\right) = 1-\alpha$$

$$\hat{\sigma}^2 = S^2$$

Διάστημα εμπιστοσύνης για σ^2 (άγνωστο μ)

$$\left[\frac{(n-1)}{\chi^2_{n-1;\alpha/2}} S^2, \ \frac{(n-1)}{\chi^2_{n-1;1-\alpha/2}} S^2\right]$$

$$S^2 = \frac{(53-66)^2 + \dots + (60-66)^2}{9} = 88.88$$

αν θέλω δ.ε. σ.ε. 90% χρειάζομαι

$\chi^2_9(0.05) = \chi^2_{9;0.05} = 16.91$ $\chi^2_9(0.95) = \chi^2_{9;0.95} = 3.325$

```
> qchisq(1-0.05,df=9)
[1] 16.91898
```

```
> qchisq(1-0.95,df=9)
[1] 3.325113
```

$$\left[\frac{9}{16.91} \times 88.88 \,,\, \frac{9}{3.325} \times 88.88\right] = [47.28 \,,\, 240.59]$$

$\chi^2_{n-1}(\alpha/2)$ $\chi^2_{n-1}(1-\alpha/2)$

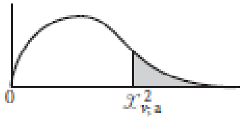
έχει μεγάλο εύρος (το δ.ε. περιέχει το $\sigma^2 = 102$ που μας είχαν δώσει)

$n=100$ [71, 114]

$n=1000$ [82, 95]

(θεωρώ ότι πήρα το ίδιο S^2.)

Πίνακας: Τιμών $\mathcal{X}^2_{v;\,a}$ της $\mathcal{X}^2$ κατανομής για τις οποίες $P(X^2 > \mathcal{X}^2_{v;\,a}) = a$



β.ε.	a=0,995	a=0,990	a=0,975	a=0,950	a=0,900
1	0,0000393	0,0001571	0,0009821	0,0039321	0,0157908
2	0,0100251	0,0201007	0,0506356	0,102587	0,210720
3	0,0717212	0,114832	0,215795	0,351846	0,584375
4	0,206990	0,297110	0,484419	0,710721	1,063623
5	0,411740	0,554300	0,831211	1,145476	1,61031
6	0,675727	0,872085	1,237347	1,63539	2,20413
7	0,989265	1,239043	1,68987	2,16735	2,83311
8	1,344419	1,646482	2,17973	2,73264	3,48954
9	1,734926	2,087912	2,70039	$\boxed{3{,}32511}$	4,16816
10	2,15585	2,55821	3,24697	3,94030	4,86518
11	2,60321	3,05347	3,81575	4,57481	5,57779
12	3,07382	3,57056	4,40379	5,22603	6,30380
13	3,56503	4,10691	5,00874	5,89186	7,04150
14	4,07468	4,66043	5,62872	6,57063	7,78953
15	4,60094	5,22935	6,26214	7,26094	8,54675
16	5,14224	5,81221	6,90766	7,96164	9,31223
17	5,69724	6,40776	7,56418	8,67176	10,0852
18	6,26481	7,01491	8,23075	9,39046	10,8649
19	6,84398	7,63273	8,90655	10,1170	11,6509
20	7,43386	8,26040	9,59083	10,8508	12,4426
21	8,03366	8,89720	10,28293	11,5913	13,2396
22	8,64272	9,54249	10,9823	12,3380	14,0415
23	9,26042	10,19567	11,6885	13,0905	14,8479
24	9,88623	10,8564	12,4011	13,8484	15,6587
25	10,5197	11,5240	13,1197	14,6114	16,4734
26	11,1603	12,1981	13,8439	15,3791	17,2919
27	11,8076	12,8786	14,5733	16,1513	18,1138
28	12,4613	13,5648	15,3079	16,9279	18,9392
29	13,1211	14,2565	16,0471	17,7083	19,7677
30	13,7867	14,9535	16,7908	18,4926	20,5992
40	20,7065	22,1643	24,4331	26,5093	29,0505
50	27,9907	29,7067	32,3574	34,7642	37,6886
60	35,5346	37,4848	40,4817	43,1879	46,4589
70	43,2752	45,4418	48,7576	51,7393	55,3290
80	51,1720	53,5400	57,1532	60,3915	64,2778
90	59,1963	61,7541	65,6466	69,1260	73,2912
100	67,3276	70,0648	74,2219	77,9295	82,3581

a=0,10	a=0,05	a=0,025	a=0,010	a=0,005	β.ε.
2,70554	3,84146	5,02389	6,63490	7,87944	1
4,60517	5,99147	7,37776	9,21034	10,5966	2
6,25139	7,81473	9,34840	11,3449	12,8381	3
7,77944	9,48773	11,1433	13,2767	14,8602	4
9,23635	11,0705	12,8325	15,0863	16,7496	5
10,6446	12,5916	14,4494	16,8119	18,5476	6
12,0170	14,0671	16,0128	18,4753	20,2777	7
13,3616	15,5073	17,5346	20,0902	21,9550	8
14,6837	$\boxed{16{,}9190}$	19,0228	21,6660	23,5893	9
15,9871	18,3070	20,4831	23,2093	25,1882	10
17,2750	19,6751	21,9200	24,7250	26,7569	11
18,5494	21,0261	23,3367	26,2170	28,2995	12
19,8119	22,3621	24,7356	27,6883	29,8194	13
21,0642	23,6848	26,1190	29,1413	31,3193	14
22,3072	24,9958	27,4884	30,5779	32,8013	15
23,5418	26,2962	28,8454	31,9999	34,2672	16
24,7690	27,5871	30,1910	33,4087	35,7185	17
25,9894	28,8693	31,5264	34,8053	37,1564	18
27,2036	30,1435	32,8523	36,1908	38,5822	19
28,4120	31,4104	34,1696	37,5662	39,9968	20
29,6151	32,6705	35,4789	38,9321	41,4010	21
30,8133	33,9244	36,7807	40,2894	42,7956	22
32,0069	35,1725	38,0757	41,6384	44,1813	23
33,1963	36,4151	39,3641	42,9798	45,5585	24
34,3816	37,6525	40,6465	44,3141	46,9278	25
35,5631	38,8852	41,9232	45,6417	48,2899	26
36,7412	40,1133	43,1944	46,9630	49,6449	27
37,9159	41,3372	44,4607	48,2782	50,9933	28
39,0875	42,5569	45,7222	49,5879	52,3356	29
40,2560	43,7729	46,9792	50,8922	53,6720	30
51,8050	55,7585	59,3417	63,6907	66,7659	40
63,1671	67,5048	71,4202	76,1539	79,4900	50
74,3970	79,0819	83,2976	88,3794	91,9517	60
85,5271	90,5312	95,0231	100,425	104,215	70
96,5782	101,879	106,629	112,329	116,321	80
107,565	113,145	118,136	124,116	128,299	90
118,498	124,342	129,561	135,807	140,169	100

Για $v > 100$, $\mathcal{X}^2_{v;\,a} = \frac{1}{2}\left(z_a + \sqrt{2v-1}\right)^2$

Δ.Ε για μ κανονικής με σ^2 άγνωστο

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \qquad \left(X_i \sim N(\mu, \sigma^2) \text{ ανεξ} \right)$$

$$\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

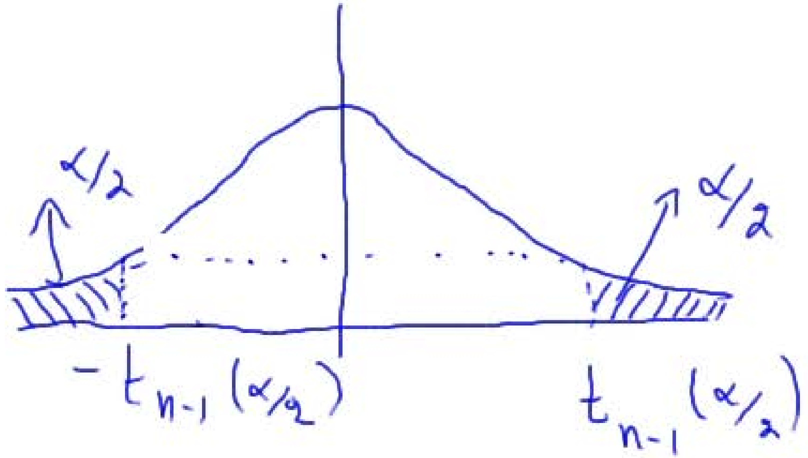
$$(n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{n-1}$$

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

$$\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} = \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{(n-1)\frac{S^2}{\sigma^2}/n-1}} \sim \frac{N(0,1)}{\sqrt{\frac{\chi^2_{n-1}}{n-1}}} \quad \text{ανεξ}$$

$$\equiv t_{n-1} \quad (\text{νέα κατανομή})$$

συμμετρική ως προς το 0
και μοιάζει με την $N(0,1)$
για μικρό n είναι πιο
"πλατιά", για $\nu >> 30$ $t_\nu \equiv N(0,1)$



$$P\left(-t_{n-1}(\alpha/2) \le \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \le t_{n-1}(\alpha/2)\right) = 1-\alpha$$

$$P\left(-t_{n-1}(\alpha/2)\frac{S}{\sqrt{n}} \le \bar{X}-\mu \le t_{n-1}(\alpha/2)\frac{S}{\sqrt{n}}\right) = 1-\alpha$$

$$P\left(\bar{X}-t_{n-1}(\alpha/2)\frac{S}{\sqrt{n}} \le \mu \le \bar{X} + t_{n-1}(\alpha/2)\frac{S}{\sqrt{n}}\right) = 1-\alpha$$

$n=10 \qquad 1-\alpha = 0.9$

$t_9(0.05) = 1.833$

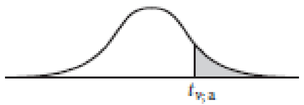
```
> qt(1-0.05,df=9)
[1] 1.833113
```

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

$z_{0.05} = 1.645$

```
> qnorm(1-0.05)
[1] 1.644854
```

Πίνακας: Τιμών $t_{v;\,a}$ της t_v -κατανομής ώστε
$P(t_v > t_{v;\,a}) = a$



β.ε.	a=0,10	a=0,05	a=0,025	a=0,010	a=0,005
1	3,078	6,314	12,706	31,821	63,657
2	1,886	2,920	4,303	6,965	9,925
3	1,638	2,353	3,182	4,541	5,841
4	1,533	2,132	2,776	3,747	4,604
5	1,476	2,015	2,571	3,365	4,032
6	1,440	1,943	2,447	3,143	3,707
7	1,415	1,895	2,365	2,998	3,499
8	1,397	1,860	2,306	2,896	3,355
9	1,383	$\boxed{1,833}$	2,262	2,821	3,250
10	1,372	1,812	2,228	2,764	3,169
11	1,363	1,796	2,201	2,718	3,106
12	1,356	1,782	2,179	2,681	3,055
13	1,350	1,771	2,160	2,650	3,012
14	1,345	1,761	2,145	2,624	2,977
15	1,341	1,753	2,131	2,602	2,947
16	1,337	1,746	2,120	2,583	2,921
17	1,333	1,740	2,110	2,567	2,898
18	1,330	1,734	2,101	2,552	2,878
19	1,328	1,729	2,093	2,539	2,861
20	1,325	1,725	2,086	2,528	2,845
21	1,323	1,721	2,080	2,518	2,831
22	1,321	1,717	2,074	2,508	2,819
23	1,319	1,714	2,069	2,500	2,807
24	1,318	1,711	2,064	2,492	2,797
25	1,316	1,708	2,060	2,485	2,787
26	1,315	1,706	2,056	2,479	2,779
27	1,314	1,703	2,052	2,473	2,771
28	1,313	1,701	2,048	2,467	2,763
29	1,311	1,699	2,045	2,462	2,756
∞	1,282	1,645	1,960	2,326	2,576

$$\left[\bar{x} - t_9(0.05)\frac{s}{\sqrt{n}} \quad , \quad \bar{x} + t_9(0.05)\frac{s}{\sqrt{n}}\right]$$

$$66 - 1.833\sqrt{\frac{88.88}{10}} \; , \; 66 + 1.833\sqrt{\frac{88.88}{10}}$$

$$66 - 1.833\cdot 2.98 \; , \; 66 + 1.833\cdot 2.98$$

$$66 - 5.46 \quad , \quad 66 + 546$$

$[60.53 \quad 71.46]$ είναι

δ.ε. σ.ε. 90% ίσων ουρών
για το μ με άγνωστη
διασπορά.

Διάστημα Εμπιστοσύνης για το μ από άγνωστη κατανομή με <u>n μεγάλο</u>

Θα χρησιμοποιήσω το ΚΟΘ $\dfrac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow[n\to\infty]{\text{ΚΟΘ}} N(0,1)$ 1η μορφή

Επίσης $\dfrac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \xrightarrow[n\to\infty]{\text{ΚΟΘ}} N(0,1)$ 2η μορφή όπου S συνεπής εκτιμητής του σ.

$$P\left(-z_{\alpha/2} \le \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \le z_{\alpha/2}\right) \simeq 1-\alpha \Rightarrow P\left(\bar{X}-z_{\alpha/2}\frac{S}{\sqrt{n}} \le \mu \le \bar{X}+z_{\alpha/2}\frac{S}{\sqrt{n}}\right) \simeq 1-\alpha$$

Ασυμπτωτικό διάστημα σ.ε. 1−α.

<u>Ειδική περίπτωση</u> $X_i \sim \text{Bernoulli}(p)$ $\mu := E(X_i) = p$ $\sigma^2 := \text{var}(X_i) = p(1-p)$

$\bar{X} = \dfrac{\#\text{ από n παρατηρήσεις που προτιμούν}}{n}$ $\hat{p} = \bar{X}$ $S^2 = \hat{p}(1-\hat{p})$

$$P\left(\hat{p}-z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p}+z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) \simeq 1-\alpha$$

Θέμα Ερωτήθηκαν 270 φοιτητές του Τμήματος Βιολογίας (170 αγόρια και 100 κορίτσια) για την κατεύθυνση που θα ακολουθήσουν. Τα αποτελέσματα που πήραμε δίδονται στον πίνακα που ακολουθεί:

	Φύλο	
Κατεύθυνση	Αγόρια	Κορίτσια
Μοριακή Βιολογία	62	36
Γενετική	84	42
Άλλες...	24	22

Εάν συμβολίσουμε με p_M το ποσοστό των φοιτητών που επιλέγουν την κατεύθυνση της Μοριακής Βιολογίας, να δώστε ένα διάστημα εμπιστοσύνης, συντελεστού εμπιστοσύνης 95%, για το p_M. Ποια κατανομή και ποιο θεώρημα χρησιμοποιήσατε για την κατασκευή του διαστήματος αυτού και γιατί;

Εδώ $\widehat{p_M} = \frac{62+36}{270} = 0.3629$. θα κατασκευάσω δ.ε., ίσων ουρών, για το ποσοστό p_M, σ.ε. $1-\alpha = 0.95$.

Επειδή $n = 270$ (αρκετά μεγάλο μέγεθος δείγματος) το διάστημα θα δίδεται από τον τύπο:

$$\left(\widehat{p_M} - z_{\frac{\alpha}{2}}\sqrt{\frac{\widehat{p_M}(1-\widehat{p_M})}{n}}, \widehat{p_M} + z_{\frac{\alpha}{2}}\sqrt{\frac{\widehat{p_M}(1-\widehat{p_M})}{n}}\right) \tag{4}$$

όπου $z_{\frac{\alpha}{2}} \overset{\text{εδώ}}{=} z_{0.025} = 1.96$.

(για $Z \sim N(0,1)$, $P(Z > z_{0.025}) = 0.025 \Rightarrow \Phi(z_{0.025}) = P(Z \leq z_{0.025}) = 1 - 0.025 \Rightarrow \Phi(z_{0.025}) = 0.975$, όμως $\Phi(1.96) = 0.975$) Συνεπώς το δ.ε. (4) γίνεται:

$$\left(0.3629 - 1.96 \cdot \sqrt{0.0008},\ \ 0.3629 + 1.96 \cdot \sqrt{0.0008}\right) = (0.3629 - 1.96 \cdot 0.0292,\ \ 0.3629 + 1.96 \cdot 0.0292)$$

$$= (0.3629 - 0.0573,\ \ 0.3629 + 0.0573) = \underline{\underline{(0.3056,\ \ 0.4203)}}$$

Για την κατασκευή του διαστήματος αυτού, επειδή το μέγεθος του δείγματος είναι μεγάλο ($n = 270$), μπορεί να εφαρμοστεί το κεντρικό οριακό θεώρημα και λόγω αυτού η ποσότητα

$$\frac{\widehat{p_M} - p_M}{\sqrt{\frac{\widehat{p_M}(1-\widehat{p_M})}{n}}} \underset{n\to\infty}{\overset{\text{προσεγ.}}{\sim}} N(0,1).$$

Σχέση μεγέθους δείγματος και εύρους διαστήματος εμπιστοσύνης

Θέμα 3ο: **α.** Ερωτήθηκαν 2700φοιτητές του Τμήματος Βιολογίας (1700αγόρια και 1000κορίτσια) για την κατεύθυνση που θα ακολουθήσουν. Τα αποτελέσματα που πήραμε δίδονται στον πίνακα που ακολουθεί:

	Φύλο	
Κατεύθυνση	Αγόρια	Κορίτσια
Μοριακή Βιολογία	620	360
Γενετική	840	420
Άλλες...	240	220

```
> p<-(620+360)/2700
> p
[1] 0.362963
> c(p-qnorm(0.975)*sqrt(p*(1-p)/2700),p+qnorm(0.975)*sqrt(p*(1-p)/2700))
[1] 0.3448254 0.3811006
```

Σχέση μεγέθους δείγματος και εύρους διαστήματος εμπιστοσύνης

	Φύλο	
Κατεύθυνση	Αγόρια	Κορίτσια
Μοριακή Βιολογία	6	4
Γενετική	84	42
Άλλες...	24	22

```
> 6+4+84+42+24+22
[1] 182
>  p<-(6+4)/182
> p
[1] 0.05494505
> c(p-qnorm(0.975)*sqrt(p*(1-p)/182),p+qnorm
[1] 0.02183917 0.08805094
```

	Φύλο	
Κατεύθυνση	Αγόρια	Κορίτσια
Μοριακή Βιολογία	62	36
Γενετική	840	420
Άλλες...	240	220

```
> 62+36+840+420+240+220
[1] 1818
> p<-(62+36)/1818
> p
[1] 0.05390539
> c(p-qnorm(0.975)*sqrt(p*(1-p)/1818),p+qnorm(0.9
[1] 0.04352450 0.06428628
```

τι μπορεί να συμβαίνει όταν το ποσοστό είναι ακόμη πιο «μικρό» (ή πολύ μεγάλο)

	Φύλο	
Κατεύθυνση	Αγόρια	Κορίτσια
Μοριακή Βιολογία	1	1
Γενετική	84	42
Άλλες...	24	22

```
> 1+1+84+42+24+22
[1] 174
>  p<-(1+1)/174
> p
[1] 0.01149425
> c(p-qnorm(0.975)*sqrt(p*(1-p)/174),p+qnorm
[1] -0.00434386  0.02733237
```

Θέμα Για την μελέτη των μεταναστευτικών συνηθειών της ριγωτής πέρκας (Morone saxatilis) αλιεύθηκαν με αγκίστρι 210 από αυτές στον ποταμό Hudson και καταγράφηκε η ηλικία τους (σε έτη). Τα δεδομένα αυτά δίνονται στον πίνακα

Ηλικία (σε έτη)	0-1	1-2	2-3	3-4	4-5	5-6
Πλήθος ψαριών	13	49	96	28	16	8

Για την ηλικία αυτού του είδους της πέρκας, να υπολογιστεί η (δειγματική) μέση τιμή, η διασπορά, το τρίτο τεταρτημόριο και η διάμεσος των παρατηρήσεων.

Υποθέτοντας ότι η ηλικία των ψαριών αυτών ακολουθεί κατανομή με μέσο μ, να δοθεί διάστημα εμπιστοσύνης, συντελεστού εμπιστοσύνης 90%, για το μέσο μ (σε έτη). Ποιας κατανομής τα ποσοστιαία σημεία χρησιμοποιήσατε για την κατασκευή του διαστήματος αυτού και γιατί;

Λύση: Τα δεδομένα είναι συνεχή και παρουσιάζονται στον πίνακα που μας δίδεται ομαδοποιημένα. Έτσι δημιουργούμε τον ακόλουθο πίνακα (για να πραγματοποιήσουμε τις αναγκαίες πράξεις):

$L_i - U_i$	x_i	f_i	F_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
0-1	0.5	13	13	6.5	3.25
1-2	1.5	49	62	73.5	110.25
2-3	2.5	96	158	240.0	600.00
3-4	3.5	28	186	98.0	343.00
4-5	4.5	16	202	72.0	324.00
5-6	5.5	8	210	44.0	242.00
Σύνολο		210		534	1622.5

Έτσι ο δειγματικός μέσος, $\overline{X}$, για την ηλικία της πέρκας δίνεται από τον τύπο:

$$\overline{x} = \frac{\sum_{i=1}^{\kappa} x_i \cdot f_i}{\sum_{i=1}^{\kappa} f_i} \underset{\kappa=6}{\overset{\varepsilon\delta\acute{\omega}}{=}} \frac{0.5 \times 13 + 1.5 \times 49 + \cdots + 5.5 \times 8}{13 + 49 + \cdots + 8} = \frac{534}{210} = 2.542857 \simeq \underline{\underline{2.54 \text{ έτη}}}$$

Εάν συμβολίσουμε με $n = \sum_{i=1}^{6} f_i$ το μέγεθος του δείγματος, η δειγματική διασπορά δίδεται από τον τύπο:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{\kappa} (x_i - \overline{x})^2 \cdot f_i \overset{\acute{\eta}}{=} \frac{1}{n-1} \left\{ \sum_{i=1}^{\kappa} x_i^2 \cdot f_i - \frac{\left(\sum_{i=1}^{\kappa} x_i \cdot f_i\right)^2}{n} \right\} \overset{\varepsilon\delta\acute{\omega}}{=} \frac{1}{209} \left\{ 1622.5 - \frac{534^2}{210} \right\}$$

$$= \frac{1}{209} \{1622.5 - 1357.886\} = \frac{264.6143}{209} = \underline{\underline{1.266097}} = (1.1252)^2,$$

Αναζητώ διάστημα εμπιστοσύνης (δ.ε.) συντελεστού εμπιστοσύνης (σ.ε.) $1-\alpha$, για το μέσο άγνωστης κατανομής, με άγνωστη διασπορά και μεγάλο μέγεθος δείγματος n=210. Γιαυτό, χρησιμοποιώ τον τύπο

$$\left(\overline{X}-z_{\alpha/2}\frac{S}{\sqrt{n}},\ \ \overline{X}+z_{\alpha/2}\frac{S}{\sqrt{n}}\right)\overset{\varepsilon\delta\acute{\omega}}{=\!=}\left(\overline{X}-z_{0.05}\frac{S}{\sqrt{210}},\ \ \overline{X}+z_{0.05}\frac{S}{\sqrt{210}}\right)=$$

$$\left(2.54-1.645\frac{1.1252}{\sqrt{210}},\ \ 2.54+1.645\frac{1.1252}{\sqrt{210}}\right)=(2.54-1.645\cdot 0.07764,\ \ 2.54+1.645\cdot 0.07764)=$$

$$(2.54-0.1277,\ \ 2.54+0.1277)=\underline{\underline{(2.4151,\ \ 2.671)}}$$

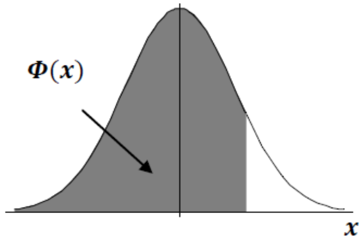
εφόσον, $1-\alpha=0.90\Rightarrow\frac{\alpha}{2}=0.05$ και $z_{0.05}$ είναι τέτοιο ώστε:

$P(Z>z_{0.05})=0.05\Rightarrow\Phi(z_{0.05})=P(Z\leq z_{0.05})=1-P(Z>z_{0.05})=1-0.05\Rightarrow\Phi(z_{0.05})=0.95$, όμως $\Phi(1.645)=0.95\Rightarrow z_{0.05}=1.645$, όπου $Z\sim N(0,1)$.

Τα δεδομένα στην περίπτωση που μελετάμε προέρχονται από άγνωστη κατανομή για την οποία δε γνωρίζουμε τη διασπορά. Συνεπώς, για την κατασκευή του δ.ε. χρησιμοποιούμε τη σ.σ. $\dfrac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}\overset{\text{ΚΟΘ}}{\sim}N(0,1)$ και επειδή το μέγεθος του δείγματος είναι μεγάλο, μπορούμε να εφαρμόσουμε το κεντρικό οριακό θεώρημα για τον δειγματικό μέσο. Έτσι τα ποσοστιαία σημεία είναι αυτά της τυπικής κανονικής κατανομής, $N(0,1)$.

Πίνακας τιμών της σ.κ. Φ της τυπικής κανονικής κατανομής ($x = 0, .01, .02, ..., 3.49$)

$$\Phi(x) = P(X \le x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$



x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5674	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.6315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990

$$\Phi(z_{0.05}) = 0.95,$$

$$\Phi(1.645) = 0.95 \Rightarrow z_{0.05} = 1.645,$$

$$\Phi(z_{0.025}) = 0.975, \text{ όμως}$$

$$\Phi(1.96) = 0.975)$$