

Ας υποθέσουμε ότι η τ.μ. X εκφράζει την επιδοση φοιτητών σε κάποιο μάθημα και ότι ο πληθυσμός αποτελείται από 10000 φοιτητές. Με βάση ένα δείγμα 10 φοιτητών πήραμε

10, 5, 6, 9, 7, 8, 4, 3, 2, 5

από όπου προκύπτουν οι σημειακές εκτιμήσεις

$$\bar{X} = 5.9$$

για τον μέσο του πληθυσμού μ

και $S^2 = 6.77$

για την διασπορά του πληθυσμού σ^2 .

Εκτίμηση μέσω διαστημάτων εμπιστοσύνης (Confidence intervals)

Στο παραπάνω δείγμα των 10 φοιτητών, πήραμε $\bar{x} = 5.9$ και $s^2 = 6.77$, όμως εάν επιλέξουμε ένα άλλο δείγμα 10 φοιτητών, πολύ πιθανών να πάρουμε κάποιες άλλες τιμές για τις \bar{x}, s^2 κοκ.

Πόσα διαφορετικά δείγματα μεγέθους 10 υπάρχουν στον πληθυσμό μας, ο οποίος αποτελείται από 10.000 φοιτητές;

Υπάρχουν

$$\binom{10.000}{10} \approx 2743 * 10^{30}$$

διαφορετικά δείγματα μεγέθους 10! Εν δυνάμει $2743 * 10^{30}$ διαφορετικές τιμές για τις \bar{x}, s^2 . Πως είναι δυνατόν να ξέρουμε ποιο δείγμα μας «φέρει» πιο κοντά στα πραγματικά μ και σ^2 ;

Είναι απαραίτητο να βρεθεί ένα τρόπος να γνωρίζουμε πόσο ακριβής είναι η εκτίμηση μας-δηλαδή, πόσο «απέχει» από την πραγματική τιμή.

Μια λύση στο προηγούμενο πρόβλημα έρχεται μέσα από τα διαστήματα εμπιστοσύνης.

Ορισμός

Ένα διάστημα $[L, U]$ το οποίο περιέχει την πραγματική τιμή μιας παραμέτρου θ , με πιθανότητα $1-\alpha$, δηλαδή

$$P(L \leq \theta \leq U) = 1 - \alpha$$

καλείται *διάστημα εμπιστοσύνης* για την παράμετρο θ , με *συντελεστή εμπιστοσύνης* $1-\alpha$ (όπου α είναι ένας αριθμός από το 0 έως το 1).

Τα άκρα του διαστήματος L και U υπολογίζονται από τις τιμές του τυχαίου δείγματος.

Το παραπάνω διάστημα καλείται και **$100(1-\alpha)\%$ διάστημα εμπιστοσύνης**.

Συνήθως το α παίρνει τις τιμές

$$\alpha=0.01 \text{ ή } 0.05 \text{ ή } 0.10$$

Για παράδειγμα, εάν το διάστημα από 40 έως 120, δηλ.

$$[40, 120]$$

είναι ένα διάστημα εμπιστοσύνης, με συντελεστή εμπιστοσύνης

$$1-0.05=0.95$$

(δηλ. $\alpha=0.05$) για τη μέση τιμή του βάρους ενός πληθυσμού, τότε σ' αυτό το διάστημα υπάρχει η πραγματική μέση τιμή, με εμπιστοσύνη 0.95.

Ισοδύναμα, θα μπορούσαμε να πούμε ότι εάν κατασκευάσουμε 100 τέτοια διαστήματα, από διαφορετικά δείγματα ίδιου μεγέθους από ένα πληθυσμό, τότε στα 95 από αυτά θα υπάρχει μέσα η πραγματική μέση τιμή και σε 5 όχι.

Πως όμως κατασκευάζεται ένα τέτοιο διάστημα;

Θα διακρίνουμε μερικές περιπτώσεις.

Διάστημα εμπιστοσύνης για τη μέση τιμή μιας τ.μ. που ακολουθεί κανονική κατανομή και έχει γνωστή διασπορά (confidence interval of the mean of a normal random variable with known variance)

Ας υποθέσουμε ότι έχουμε ένα τυχαίο δείγμα X_1, X_2, \dots, X_n από μια τ.μ. X που ακολουθεί κανονική κατανομή, με **άγνωστη** μέση τιμή μ και **γνωστή** διασπορά σ^2 (δηλ. $X \sim N(\mu, \sigma^2)$).

Τότε ένα διάστημα εμπιστοσύνης για τη μέση τιμή μ , με συντελεστή εμπιστοσύνης $1-\alpha$ είναι το

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

όπου ο αριθμός $z_{\alpha/2}$ είναι τέτοιος ώστε

$$P(Z > z_{\alpha/2}) = \frac{\alpha}{2}$$

και η $Z \sim N(0,1)$.

Γενικώς, το z_α ονομάζεται και άνω α -ποσοστιαίο σημείο της τυπικής κανονικής κατανομής (για οποιοδήποτε α).

Καθς διάστημα εμπιστοσύνης για μία παράμετρο θ
βασίζεται σε μια οδυσήτρια συνάρτηση.

Η οδυσήτρια συνάρτηση είναι συνάρτηση των X_1, \dots, X_n
αλλά και συνάρτηση της αγνώστης / άγνωστης παραμέτρου
αλλά η κατανομή της δς θέλουμε να εξαρτάται
απο αγνώστες παραμέτρους.

Για την περίπτωση που $X_1, \dots, X_n \sim N(\mu, \sigma^2)$
 με αγνωστή μέση τιμή μ και γνωστή διασπορά σ^2
 το Δ.Ε. για το μ βασίζεται στην οδυγίτρια συνάρτηση

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

η κατανομή δεν εξαρτάται από το μ
 που είναι αγνώστο.

Επόμενο βήμα πάνω στα διαστήματα:

$$P(L \leq \text{οδυγίτρια} \leq U) = 1 - \alpha$$

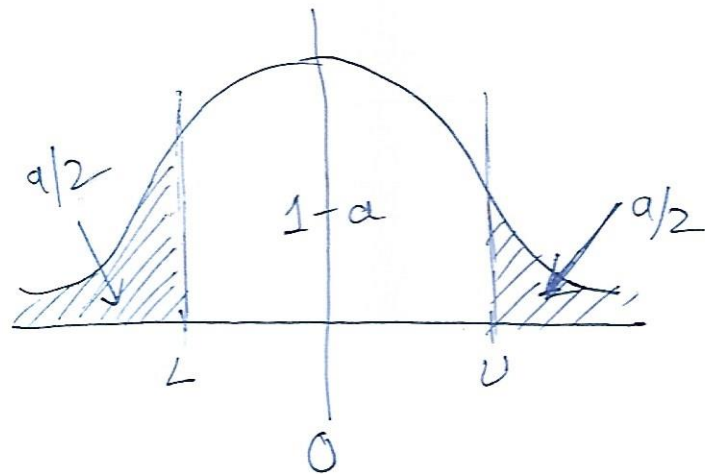
μεγάλη

δυαδική

α μικρό

Τα L και U είναι 'κατάλληλα' σημεία της
κατανομής της οδοντοζίας έτσι ώστε αυτή η
πιθανότητα που μας ενδιαφέρει να είναι
ακριβώς ίση με $1-\alpha$.

Για τη $N(0,1)$ λοιπόν τα σημεία L και U
είναι από το σχήμα:



$\frac{a}{2}$ εμβαδόν σε κάθε άκρη λόγω συμμετρίας και
 εμβαδόν $1-a$ μεταξύ L και U

Επιλέγουμε το συμβολισμό $z_{a/2}$ για το U και
 το L είναι το συμμετρικό του δηλαδή $-z_{a/2}$.

$$z_{a/2} : P(Z > z_{a/2}) = a/2$$

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

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

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

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

Αρα το Διάστημα εμπιστοσύνης με συνέπεια
εμπιστοσύνης $1 - \alpha$ είναι το αχίο διάστημα

$$\left(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right)$$

το οποίο περιέχει το μ με πιθανότητα $1 - \alpha$

Παραδείγματα 17.1

Μια ασφαλιστική εταιρεία γνωρίζει ότι η τυχαία μεταβλητή X που εκφράζει την ετήσια αποζημίωση, για ένα συμβόλαιο πυρασφάλειας, ακολουθεί κανονική κατανομή με διασπορά 9. Παίρνοντας ένα τυχαίο δείγμα 10 συμβολαίων, παρατηρούμε τις παρακάτω αποζημιώσεις (χιλιάδες ευρώ)

10, 20, 15, 8, 50, 30, 25, 35, 40, 12.

Να βρεθεί το 90% διάστημα εμπιστοσύνης, για τη μέση τιμή των ετήσιων αποζημιώσεων.

Παράδειγμα 17.1

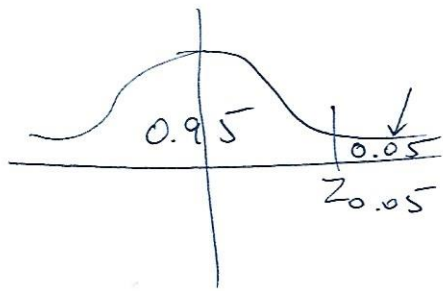
$$\bar{X} = \frac{10+20+15+\dots}{10} = 24.5$$

$$\sigma = \sqrt{9} = 3$$

$$1-\alpha = 0.9 \Rightarrow \alpha = 0.1 \Rightarrow \frac{\alpha}{2} = 0.05$$

90% ΔΕ.

Θέλουμε το σημείο $Z_{\alpha/2} = Z_{0.05}$



0.95 πιθανότητα στα αριστερά του σημείου $Z_{0.05}$

Από πίνακα $N(0,1)$ βρίσκουμε $z_{0.05} = 1.645$

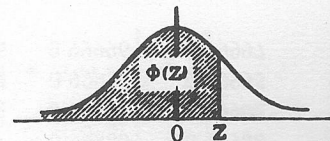
$$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} : \left(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} , \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$

$$\left(24.5 - 1.645 \frac{3}{\sqrt{10}} , 24.5 + 1.645 \frac{3}{\sqrt{10}} \right)$$

$$(22.93 , 26.04)$$

ΠΙΝΑΚΑΣ 2. ΚΑΝΟΝΙΚΗ ΚΑΤΑΝΟΜΗ

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-u^2/2} du$$



Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84850	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92786	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670

Παραδείγματα 17.2

Ας υποθέσουμε ότι η τυχαία μεταβλητή που εκφράζει το χρόνο ζωής ενός εξαρτήματος (σε χιλιάδες ώρες), ακολουθεί κανονική κατανομή με διασπορά 4. Από ένα τυχαίο δείγμα 9 εξαρτημάτων καταγράψαμε τα εξής αποτελέσματα για τον παραπάνω χρόνο

4, 0.5, 5, 3.5, 1, 2, 1.5, 2, 0

Να βρεθεί το 95% διάστημα εμπιστοσύνης, για τη μέση τιμή του χρόνου ζωής αυτού του εξαρτήματος.

Παραδείγματα 17.3

Ας υποθέσουμε ότι η τυχαία μεταβλητή που εκφράζει το χρόνο (σε λεπτά) που ένας μαθητής χρειάζεται για να συμπληρώσει ένα τεστ, ακολουθεί κανονική κατανομή με διασπορά 25. Από ένα τυχαίο δείγμα 10 μαθητών καταγράψαμε τα εξής αποτελέσματα για τον παραπάνω χρόνο

50, 60, 45, 35, 60, 70, 65, 55, 45, 70

Να βρεθεί το 99% διάστημα εμπιστοσύνης, για τη μέση τιμή του χρόνου που χρειάζεται ένας μαθητής για να συμπληρώσει το τεστ.

Επίσης, να βρεθεί και το 90% διάστημα εμπιστοσύνης, για τη μέση τιμή.

Τι παρατηρείτε για το νέο διάστημα σε σχέση με το 99% δ.ε.;

Παρατηρήσεις

- Το μήκος (πλάτος) του διαστήματος εμπιστοσύνης με συντελεστή εμπιστοσύνης $1-\alpha$, για την περίπτωση που έχουμε μελετήσει, είναι

$$l = 2 \frac{\sigma}{\sqrt{n}} z_{\alpha/2}.$$

- Για συγκεκριμένα δεδομένα, αυξάνοντας το α , το πλάτος του δ.ε. μειώνεται. Αυτό σημαίνει ότι για συγκεκριμένα δεδομένα π.χ. το 90% δ.ε. θα έχει σίγουρα μικρότερο πλάτος από το 95%.

Τι σημαίνει αυτό;

- Τι γίνεται όταν για συγκεκριμένο πρόβλημα, αυξήσω το μέγεθος του δείγματος και διατηρήσω το α σταθερό;

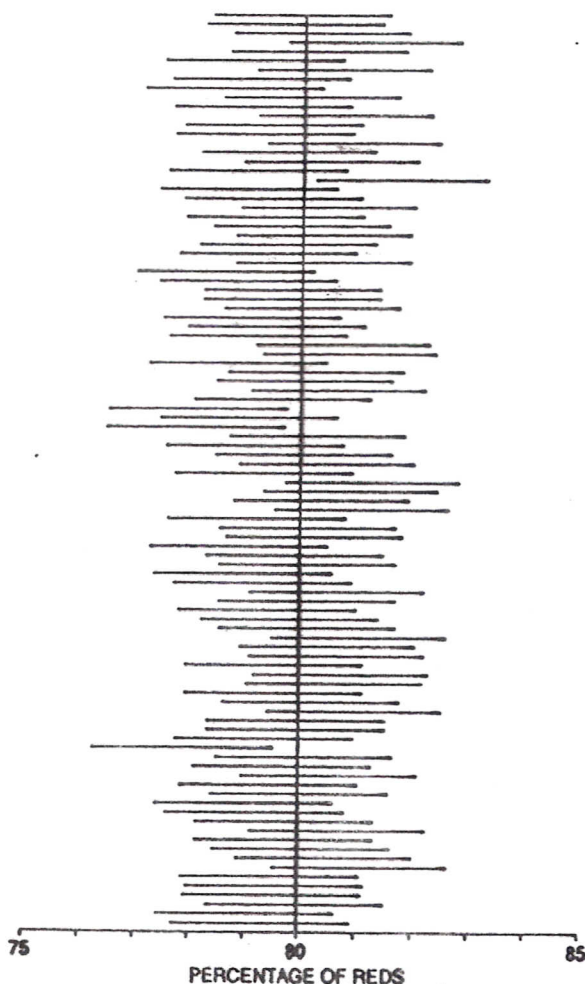
Ποιο πρέπει να είναι το μέγεθος του δείγματος, ώστε το πλάτος να είναι μικρότερο από κάποιον αριθμό $c>0$;

Το Δ.Ε. που έχουμε βρει γενικεύεται για τον μέσο μ οποιασδήποτε κατανομής με γνωστή διασπορά σ^2 αν το n είναι μεγάλο ($n \geq 30$) αφού τότε

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \underset{\text{προσεγγιστικά}}{\sim} N(0,1) \quad \text{από το κ.ο.θ.}$$

$$\Delta E. : \quad \bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Figure 1. Interpreting confidence intervals. The 95%-confidence interval is shown for 100 different samples. The interval changes from sample to sample. For about 95% of the samples, the interval covers the population percentage, marked by a vertical line.



Έστω ζώα X_1, \dots, X_n τ.δ. από την $N(\mu, \sigma^2)$
όπου μ και σ^2 αγνώστα. Να βρεθεί ΔΕ. για το μ .

Οδυσμεία συνάρτησης

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

αφού το σ είναι αγνώστο
από την εκτίμηση του

αντικαθίσταται
τη δειγματική τυπική απόκλιση

$$\Delta.Ε: \quad \bar{X} \pm t_{n-1, \alpha/2} \frac{S}{\sqrt{n}}$$

Διάστημα εμπιστοσύνης για τη μέση τιμή μιας τ.μ. που ακολουθεί κανονική κατανομή και έχει άγνωστη διασπορά (confidence interval of the mean of a normal random variable with unknown variance)

Ας υποθέσουμε ότι έχουμε ένα τυχαίο δείγμα X_1, X_2, \dots, X_v από μια τ.μ. X που ακολουθεί κανονική κατανομή, με **άγνωστη** μέση τιμή μ και **άγνωστη** διασπορά σ^2 (δηλ. $X \sim N(\mu, \sigma^2)$).

Τότε ένα διάστημα εμπιστοσύνης για τη μέση τιμή μ , με συντελεστή εμπιστοσύνης $1-\alpha$ είναι το

$$\left[\bar{X} - \frac{s}{\sqrt{v}} t_{v-1; \alpha/2}, \bar{X} + \frac{s}{\sqrt{v}} t_{v-1; \alpha/2} \right].$$

Ο αριθμός $t_{v-1; \alpha/2}$ είναι τέτοιος ώστε

$$P(T > t_{v-1; \alpha/2}) = \frac{\alpha}{2}$$

και η τ.μ. T ακολουθεί την **t -κατανομή με $v-1$ βαθμούς ελευθερίας (β.ε.)**.

Η συνάρτηση πυκνότητας πιθανότητας μιας τ.μ. που ακολουθεί t -κατανομή (ή κατανομή Student) με ν βαθμούς ελευθερίας, δίδεται από τη σχέση

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}, -\infty < t < +\infty$$

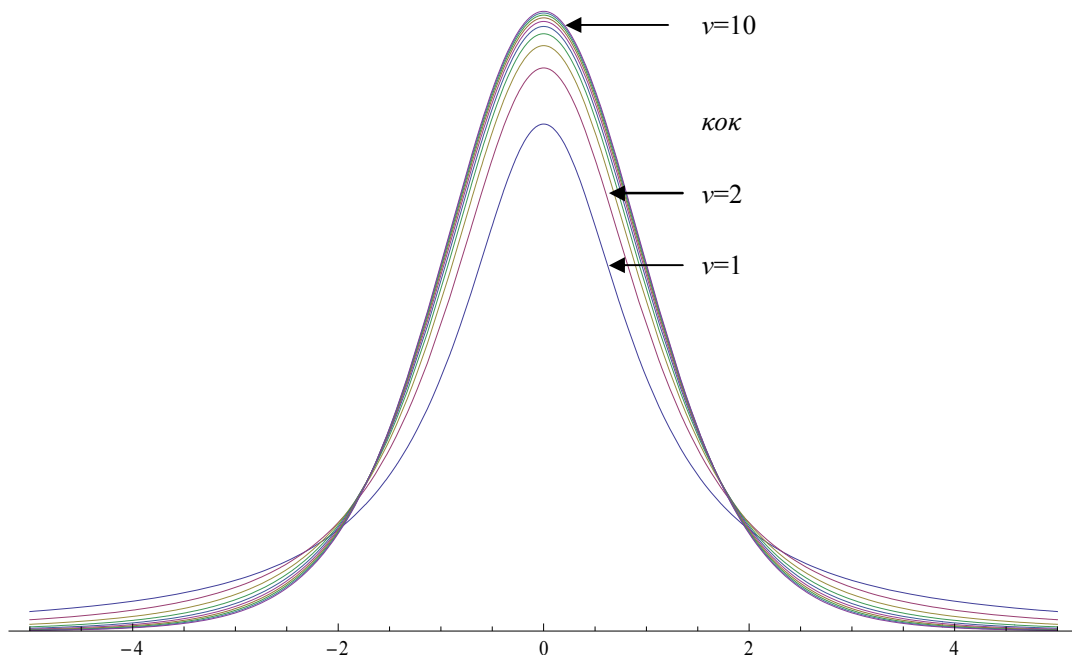
όπου $\Gamma(x)$ είναι η συνάρτηση (Γάμμα)

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt, x > 0$$

Για την συνάρτηση $\Gamma(x)$ ισχύει επιπλέον

$$\Gamma(x+1) = x\Gamma(x), x > 0$$

$$\Gamma(\nu) = (\nu-1)!, \text{ για ακέραιο θετικό } \nu$$



Αποδεικνύεται ότι όταν το $\nu \rightarrow \infty$, η συνάρτηση πυκνότητας πιθανότητας της t -κατανομής, γίνεται $N(0,1)$.

Πως θα βρίσκουμε τις τιμές των $t_{\nu-1, \alpha/2}$; Χρησιμοποιώντας τώρα τους πίνακες της t -κατανομής.

Τα άνω α-ποσοστιαία σημεία της t -κατανομής, με ν β.ε. ($t_\nu(\alpha)$)

$\nu \backslash \alpha$	0.4	0.25	0.1	0.05	0.025	0.01	0.005	0.0025	0.001
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785
8	0.262	0.706	1.397	1.86	2.306	2.896	3.355	3.833	4.501
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686
17	0.257	0.689	1.333	1.74	2.110	2.567	2.898	3.222	3.646
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579

συνέχεια...

Τα άνω α -ποσοστιαία σημεία της t -κατανομής, με ν β.ε. ($t_\nu(\alpha)$)

$\nu \backslash \alpha$	0.4	0.25	0.1	0.05	0.025	0.01	0.005	0.0025	0.001
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.45
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396
30	0.256	0.683	1.310	1.697	2.042	2.457	2.75	3.03	3.385
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160

Παράδειγμα 18.1

Ας υποθέσουμε ότι η τυχαία μεταβλητή που εκφράζει το χρόνο ζωής ενός εξαρτήματος (σε χιλιάδες ώρες), ακολουθεί κανονική κατανομή (βλ. και Παράδειγμα 17.2). Από ένα τυχαίο δείγμα 9 εξαρτημάτων καταγράψαμε τα εξής αποτελέσματα για τον παραπάνω χρόνο

4, 0.5, 5, 3.5, 1, 2, 1.5, 2, 0

Να βρεθεί το 95% διάστημα εμπιστοσύνης, για τη μέση τιμή του χρόνου ζωής αυτού του εξαρτήματος.

Έχουμε

$$\bar{X} = 2.17, s = 1.67, t_{8;0.025} = 2.306$$

άρα

$$\left[\bar{X} - \frac{s}{\sqrt{n}} t_{n-1; \alpha/2}, \bar{X} + \frac{s}{\sqrt{n}} t_{n-1; \alpha/2} \right] = [0.878, 3.456]$$

Παράδειγμα 18.2

Ας υποθέσουμε ότι η τυχαία μεταβλητή που εκφράζει το χρόνο (σε λεπτά) που ένας μαθητής χρειάζεται για να συμπληρώσει ένα τεστ, ακολουθεί κανονική κατανομή (βλ. και Παράδειγμα 17.3). Από ένα τυχαίο δείγμα 10 μαθητών καταγράψαμε τα εξής αποτελέσματα για τον παραπάνω χρόνο

50, 60, 45, 35, 60, 70, 65, 55, 45, 70

Να βρεθεί το 99% διάστημα εμπιστοσύνης, για τη μέση τιμή του χρόνου που χρειάζεται ένας μαθητής για να συμπληρώσει το τεστ.

Έχουμε

$$\bar{X} = 55.5, s = 11.66, t_{9;0.005} = 3.25$$

άρα

$$\left[\bar{X} - \frac{s}{\sqrt{v}} t_{v-1; \alpha/2}, \bar{X} + \frac{s}{\sqrt{v}} t_{v-1; \alpha/2} \right] = [43.52, 67.48]$$

Παράδειγμα 18.3

Έστω ότι για ένα τυχαίο δείγμα 81 υλικών καταγράψαμε την αντοχή τους σε μια συγκεκριμένη δοκιμασία. Η δειγματική μέση τιμή βρέθηκε να είναι 74.6 και η τυπική απόκλιση 11.3. Αν υποθέσουμε ότι η τυχαία μεταβλητή που εκφράζει την αντοχή των υλικών ακολουθεί κανονική κατανομή, να βρεθεί ένα 90% διάστημα εμπιστοσύνης για τη μέση τιμή της.

Προσεγγιστικό διάστημα εμπιστοσύνης για τη μέση τιμή μιας τ.μ. που ακολουθεί άγνωστη κατανομή και έχει άγνωστή διασπορά (confidence interval of the mean: large random sample, unknown distribution and unknown variance)

Ας υποθέσουμε ότι έχουμε ένα τυχαίο δείγμα X_1, X_2, \dots, X_n από μια τ.μ. X που ακολουθεί μια οποιαδήποτε (άγνωστη) κατανομή, με άγνωστη μέση τιμή μ και άγνωστη διασπορά σ^2 .

Τότε αν το μέγεθος του δείγματος είναι μεγάλο ($n \geq 30$), ένα προσεγγιστικό διάστημα εμπιστοσύνης για τη μέση τιμή μ , με συντελεστή εμπιστοσύνης $1-\alpha$ είναι το

$$\left[\bar{X} - \frac{s}{\sqrt{n}} z_{\alpha/2}, \bar{X} + \frac{s}{\sqrt{n}} z_{\alpha/2} \right].$$

Τι σημαίνει προσεγγιστικό διάστημα εμπιστοσύνης;

Confidence interval for the population mean μ

n large and σ^2 known	n large and σ^2 unknown	n small and σ^2 known	n small and σ^2 unknown
$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$	$\bar{X} \pm z_{\alpha/2} \frac{S}{\sqrt{n}}$	$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$	$\bar{X} \pm t_{n-1, \alpha/2} \frac{S}{\sqrt{n}}$
Any distribution for the population	Any distribution for the population	Population Normally distributed	Population normally distributed

Παράδειγμα 18.4

Έστω ότι για ένα τυχαίο δείγμα 2000 απόφοιτων πανεπιστημίων καταγράψαμε το χρόνο που χρειάστηκαν για να ολοκληρώσουν τις σπουδές τους. Η δειγματική μέση τιμή βρέθηκε να είναι 5.2 έτη και η τυπική απόκλιση 1.2 έτη. Να βρεθεί το 95% και 99% προσεγγιστικό διάστημα εμπιστοσύνης για τη μέση τιμή του πληθυσμού.

Άρα για το 95% έχουμε

$$\bar{X} = 5.2, s = 1.2, n = 2000, \alpha = 0.05, z_{\alpha/2} = z_{0.025} = 1.96$$

και

$$\bar{X} - \frac{s}{\sqrt{n}} z_{\alpha/2} = 5.1474$$

$$\bar{X} + \frac{s}{\sqrt{n}} z_{\alpha/2} = 5.2526$$

Για το 99% έχουμε

$$\alpha = 0.01, z_{\alpha/2} = z_{0.005} = 2.57$$

και

$$\bar{X} - \frac{s}{\sqrt{n}} z_{\alpha/2} = 5.1310$$

$$\bar{X} + \frac{s}{\sqrt{n}} z_{\alpha/2} = 5.2690$$

Προσεγγιστικό διάστημα εμπιστοσύνης για το ποσοστό (Confidence interval of a population proportion)

Ας υποθέσουμε ότι έχουμε ένα τυχαίο δείγμα X_1, X_2, \dots, X_n από μια τ.μ. X η οποία μπορεί να πάρει μόνο δυο διαφορετικές τιμές.

Παραδείγματα τέτοιων μεταβλητών είναι το φύλο ενός παιδιού, η ένδειξη από τη ρίψη ενός νομίσματος, το αποτέλεσμα από ένα έλεγχο για την ύπαρξη μιας ουσίας (υπάρχει ή όχι), το αν μια μέρα έχει βρέξει ή όχι, το εάν είναι κάποιος καπνιστής ή όχι κ.α.

Εάν η τ.μ. X θεωρήσουμε ότι παίρνει τις τιμές 0 ή 1, τότε η πιθανότητα που μας ενδιαφέρει να μελετήσουμε είναι η

$$P(X=1) = p$$

ενώ φυσικά $P(X=0) = 1 - p$.

Η τ.μ. X λέγεται ότι ακολουθεί την κατανομή Bernoulli.

Τότε αν το μέγεθος του δείγματος είναι μεγάλο ($n \geq 30$), ένα προσεγγιστικό διάστημα εμπιστοσύνης για την άγνωστη πιθανότητα p , με συντελεστή εμπιστοσύνης $1-\alpha$ είναι το

$$\left[\hat{p} - \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} z_{\alpha/2}, \hat{p} + \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} z_{\alpha/2} \right]$$

όπου

$$\hat{p} = \frac{X_1 + X_2 + \dots + X_n}{n} = \bar{X}.$$

Προσοχή: το παραπάνω είναι προσεγγιστικό διάστημα εμπιστοσύνης;

Παράδειγμα 18.5

Σε ένα τυχαίο δείγμα 1000 φοιτητών, οι 700 φοιτητές δήλωσαν ότι είναι καπνιστές. Ποιο είναι το 95% προσεγγιστικό δ.ε. για την πιθανότητα κάποιος να είναι καπνιστής.

Πως μπορούμε να βρούμε το αντίστοιχο προσεγγιστικό δ.ε. για την πιθανότητα κάποιος να μην είναι καπνιστής.

Για το 95% δ.ε. έχουμε

$$\hat{p} = \frac{X_1 + X_2 + \dots + X_v}{v} = \frac{700}{1000} = 0.7, v = 1000, a = 0.05, z_{a/2} = z_{0.025} = 1.96.$$

Άρα

$$\hat{p} - \sqrt{\frac{\hat{p}(1-\hat{p})}{v}} z_{a/2} = 0.6716, \hat{p} + \sqrt{\frac{\hat{p}(1-\hat{p})}{v}} z_{a/2} = 0.7284$$

και το 95% για την p είναι

$$[0.6716, 0.7284]$$

ενώ για το $1-p$, το 95% είναι (γιατί;)

$$[1 - 0.7284, 1 - 0.6716] = [0.2716, 0.3284].$$

Παράδειγμα 18.6

Σε ένα τυχαίο δείγμα 500 ατόμων, οι 200 δήλωσαν ότι συμφωνούν με τις θέσεις ενός κόμματος για την οικονομία και οι υπόλοιποι όχι. Ποιο είναι το 90% προσεγγιστικό δ.ε. για την πιθανότητα κάποιο άτομο στον πληθυσμό, να συμφωνεί με τις θέσεις του παραπάνω κόμματος για την οικονομία;

Για το 90% δ.ε. έχουμε

$$\hat{p} = \frac{X_1 + X_2 + \dots + X_v}{v} = \frac{200}{500} = 0.4, v = 500, a = 0.10, z_{a/2} = z_{0.05} = 1.64.$$

Άρα

$$\hat{p} - \sqrt{\frac{\hat{p}(1-\hat{p})}{v}} z_{a/2} = 0.3641, \hat{p} + \sqrt{\frac{\hat{p}(1-\hat{p})}{v}} z_{a/2} = 0.4359.$$

Παράδειγμα 18.7

Σε 300 όμοιες κατασκευές έγινε ανεξάρτητα ένα τεστ αντοχής και οι 50 από αυτές παρουσίασαν σοβαρές αλλοιώσεις (και οι υπόλοιπες όχι). Ποιο είναι το 99% προσεγγιστικό δ.ε. για την πιθανότητα μια τέτοια κατασκευή να παρουσιάσει σοβαρές αλλοιώσεις κάτω από παρόμοιες επιβαρύνσεις;

Για το 99% δ.ε. έχουμε

$$\hat{p} = \frac{X_1 + X_2 + \dots + X_v}{v} = \frac{50}{300} = 0.167, v = 300, a = 0.01, z_{a/2} = z_{0.005} = 2.57.$$

Άρα

$$\hat{p} - \sqrt{\frac{\hat{p}(1-\hat{p})}{v}} z_{a/2} = 0.1114, \hat{p} + \sqrt{\frac{\hat{p}(1-\hat{p})}{v}} z_{a/2} = 0.2220.$$
