

ΠΙΘΑΝΟΤΗΤΕΣ & ΣΤΑΤΙΣΤΙΚΗ – ΤΥΠΟΛΟΓΙΟ Β ΠΡΟΟΔΟΥ

ΣΠΠ: $P(X \in B) = \int_B f_X(x)dx$, $P(a \leq X \leq b) = \int_a^b f_X(x)dx$, $P(-\infty \leq X \leq \infty) = \int_{-\infty}^{\infty} f_X(x)dx = 1$

ΑΣΚ: $F_X(x) = P(X \leq x) = \begin{cases} \sum_{k \leq x} p_X(k), & \text{αν η } X \text{ είναι διακριτή} \\ \int_{-\infty}^x f_X(t)dt, & \text{αν η } X \text{ είναι συνεχή} \end{cases}$

Σχέση ΣΠΠ και ΑΣΚ: $F_X(x) = \int_{-\infty}^x f_X(t)dt$, $f_X(x) = \frac{dF_X}{dx}(x)$

Μέση τιμή: $E[X] = \int_{-\infty}^{\infty} x f_X(x)dx$, $E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)dx$

n-στη ροπή: $E[X^n] = \int_{-\infty}^{\infty} x^n f_X(x)dx$

Διασπορά: $var(X) = E[(X - E[X])^2] = \int_{-\infty}^{\infty} (x - E[X])^2 f_X(x)dx$, $var(X) = E[X^2] - (E[X])^2 \geq 0$

Γραμμικός μετασχηματισμός τ.μ.: $E[Y] = aE[X] + b$, $var(X) = a^2 var(X)$

Ομοιόμορφη τ.μ.: $f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{αν } a \leq x \leq b \\ 0, & \text{αλλιώς} \end{cases}$, $E[X] = \frac{a+b}{2}$, $var(X) = \frac{(b-a)^2}{12}$

Εκθετική τ.μ.: $f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \text{αλλιού} \end{cases}$, $\lambda > 0$, $P(X \geq a) = e^{-\lambda a}$, $F_{exp}(x) = 1 - e^{-\lambda x}$, $E[X] = \frac{1}{\lambda}$, $var(X) = \frac{1}{\lambda^2}$

Κανονική τ.μ.: $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $E[X] = \mu$, $var(X) = \sigma^2$,

Γραμμικός μετ/σμός κανονικής τ.μ.: $Y = aX + b$, $E[Y] = a\mu + b$, $var(Y) = a^2 \sigma^2$

Τυποποιημένη κανονική τ.μ.: $Y = \frac{X-\mu}{\sigma}$, $\mu = 0$, $\sigma^2 = 1$, $\Phi(y) = P(Y \leq y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y e^{-\frac{t^2}{2}} dt$,

$\Phi(-y) = 1 - \Phi(y)$

Κατανομή συνάρτησης τ.μ.: $Y = g(X)$, $F_Y(y) = P(g(X) \leq y) = \int_{\{x|g(x) \leq y\}} f_X(x)dx$, $f_Y(y) = \frac{dF_Y}{dy}(y)$

Τετράγωνο τ.μ.: $Y = X^2$, $f_Y(y) = \frac{1}{2\sqrt{y}} [f_X(\sqrt{y}) + f_X(-\sqrt{y})]$, $y \geq 0$

Γραμμική συνάρτηση τ.μ.: $Y = aX + b$, $a \neq 0$, b σταθερές, $f_Y(y) = \frac{dF_Y}{dy}(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$

Μονότονες συναρτήσεις τ.μ.: $Y = g(X)$, $X = h(Y)$, $f_Y(y) = f_X(h(y)) \left| \frac{dh}{dy}(y) \right|$

Συνδιασπορά: $cov(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$

Ιδιότητες συνδιασποράς: $cov(X, X) = var(X)$, $cov(X, aY + b) = a \cdot cov(X, Y)$

$$cov(X, Y + Z) = cov(X, Y) + cov(X, Z)$$

Αν $E[XY] = E[X]E[Y]$ τότε $cov(X, Y) = 0$

$$var(X + Y) = var(X) + var(Y) + 2cov(X, Y)$$

Συντελεστής συσχέτισης: $\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}, \quad -1 \leq \rho(X, Y) \leq 1, \quad \rho = \pm 1: \quad Y - E[Y] = c(X - E[X])$

Ανισότητα Markov: $P(X \geq a) \leq \frac{E[X]}{a}$ για κάθε $a > 0$

Ανισότητα Chebychev: $P(|X - \mu| \geq c) \leq \frac{\sigma^2}{c^2}$ για κάθε $c > 0$, $P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$ για κάθε $k > 0$

Ασθενής νόμος μεγάλων αριθμών: $M_n = \frac{X_1 + \dots + X_n}{n}$ $P(|M_n - \mu| \geq \varepsilon) \rightarrow 0$ όπως $n \rightarrow \infty$

Κεντρικό Οριακό Θεώρημα (ΚΟΘ): X_1, X_2, \dots, X_n ανεξάρτητες τ.μ. με την ίδια κατανομή, ίδια μέση τιμή μ και διασπορά σ^2 , $Z_n = \frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}}$, $\lim_{n \rightarrow \infty} P(Z_n \leq z) = \Phi(z)$ όπου $\Phi(z)$ η ΑΣΚ της τυποποιημένης κανονικής.

Υπολογισμοί πιθανοτήτων με ΚΟΘ: $S_n = X_1 + \dots + X_n$ άθροισμα ανεξάρτητων τ.μ. με ίδια κατανομή, μέση τιμή μ , διασπορά σ^2 : $E[S_n] = n\mu$, $\text{var}(S_n) = n\sigma^2$, $z = \frac{c - n\mu}{\sigma\sqrt{n}}$, $P(S_n \leq c) \approx \Phi(z)$ από τους πίνακες της τυποποιημένης κανονικής ΑΣΚ $\Phi(z)$

Προσέγγιση διωνυμικής με παραμέτρους n, p : $z = \frac{c - np}{\sqrt{np(1-p)}}$, $P(S_n \leq c) \approx \Phi(z)$ από τους πίνακες της τυποποιημένης κανονικής ΑΣΚ $\Phi(z)$

Προσέγγιση De Moivre - Laplace της Διωνυμικής: $P(k \leq S_n \leq l) \approx \Phi\left(\frac{l + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$

$\hat{\theta}$ εκτιμήτρια άγνωστης παραμέτρου θ : συνάρτηση n παρατηρήσεων X_1, \dots, X_n των οποίων η κατανομή εξαρτάται από την θ

Εκτίμηση μέγιστης πιθανοφάνειας: $\hat{\theta}n = \arg \max_{\theta} p_X(x_1, \dots, x_n; \theta)$ ή $\hat{\theta}n = \arg \max_{\theta} f_X(x_1, \dots, x_n; \theta)$

Συνάρτηση πιθανοφάνειας: $p_X(x_1, \dots, x_n; \theta) = \prod_{i=1}^n p_{X_i}(x_i; \theta)$

Λογαριθμική πιθανοφάνεια: $\log p_X(x_1, \dots, x_n; \theta) = \log \prod_{i=1}^n p_{X_i}(x_i; \theta) = \sum_{i=1}^n \log p_{X_i}(x_i; \theta)$ ή

$$\log f_X(x_1, \dots, x_n; \theta) = \log \prod_{i=1}^n f_{X_i}(x_i; \theta) = \sum_{i=1}^n \log f_{X_i}(x_i; \theta)$$

Διάστημα εμπιστοσύνης $(1-\alpha)*100\%$, π.χ. $\alpha=0.05$, 95% δ.ε.: $\Phi(z)$ ΑΣΚ της τυποποιημένης κανονικής και $\Psi_{n-1}(z)$ ΑΣΚ της t κατανομής με $n-1$ βαθμούς ελευθερίας, n μέγεθος δείγματος

(i) Για άγνωστη μέση τιμή κανονικής κατανομής θ , η διασπορά $v = \sigma^2$ είναι γνωστή

$$\left[\hat{\theta}_n - z \sqrt{\frac{v}{n}}, \hat{\theta}_n + z \sqrt{\frac{v}{n}} \right], \text{ όπου } \hat{\theta}_n = \frac{\sum_{i=1}^n X_i}{n} \text{ και } \Phi(z) = 1 - \frac{\alpha}{2}$$

(ii) Για άγνωστη μέση τιμή κανονικής κατανομής θ , η διασπορά $v = \sigma^2$ είναι άγνωστη, δείγμα μεγάλο

$$\left[\hat{\theta}_n - z \sqrt{\frac{\hat{S}_n^2}{n}}, \hat{\theta}_n + z \sqrt{\frac{\hat{S}_n^2}{n}} \right] \text{ όπου } \hat{\theta}_n = \frac{\sum_{i=1}^n X_i}{n}, \hat{S}_n^2 = \frac{\sum_{i=1}^n (X_i - \hat{\theta}_n)^2}{n-1} \text{ και } \Phi(z) = 1 - \frac{\alpha}{2}$$

(iii) Για άγνωστη μέση τιμή κανονικής κατανομής θ , η διασπορά $v = \sigma^2$ είναι άγνωστη, δείγμα μικρό

$$\left[\hat{\theta}_n - z \sqrt{\frac{\hat{S}_n^2}{n}}, \hat{\theta}_n + z \sqrt{\frac{\hat{S}_n^2}{n}} \right] \text{ όπου } \hat{\theta}_n = \frac{\sum_{i=1}^n X_i}{n}, \hat{S}_n^2 = \frac{\sum_{i=1}^n (X_i - \hat{\theta}_n)^2}{n-1} \text{ και } \Psi_{n-1}(z) = 1 - \frac{\alpha}{2}$$

(iv) Για άγνωστο ποσοστό θ επιτυχιών:

$$\hat{\theta}_n = \frac{\text{αριθμός επιτυχιών}}{n}, \quad \left[\hat{\theta}_n - z \frac{\sqrt{\hat{\theta}_n(1-\hat{\theta}_n)}}{\sqrt{n}}, \hat{\theta}_n + z \frac{\sqrt{\hat{\theta}_n(1-\hat{\theta}_n)}}{\sqrt{n}} \right], \quad \Phi(z) = 1 - \frac{\alpha}{2}$$

Έλεγχος υπόθεσης:

Με R συμβολίζεται η απορριπτική περιοχή της H_0 .

Το z_α είναι η τιμή τυποποιημένης κανονικής για την οποία $\Phi(z) = 1 - \alpha$,

Το $z_{\alpha/2}$ είναι η τιμή τυποποιημένης κανονικής για την οποία $\Phi(z) = 1 - \frac{\alpha}{2}$

(i) Έλεγχος για **ποσοστό** (αναλογία) χαρακτηριστικού

Εκτίμηση της αναλογίας από δείγμα: $\hat{p} = \frac{x}{n}$
 x = αριθμός εμφανίσεων χαρ/κού & n = μέγεθος δείγματος

$$H_0 : p = p_0$$

$$(\alpha) H_1 : p > p_0$$

$$(\beta) H_1 : p < p_0$$

$$(\gamma) H_1 : p \neq p_0$$

$$R = \{z > z_\alpha\}$$

$$R = \{z < -z_\alpha\}$$

$$R = \{|z| > z_{\alpha/2}\}$$

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

(ii) Έλεγχος για τη **μέση τιμή** μ όταν σ **άγνωστο** και έχουμε **μεγάλο δείγμα**

$n \geq 30$ $H_0 : \mu = \mu_0$ $(\alpha) H_1 : \mu > \mu_0$ $(\beta) H_1 : \mu < \mu_0$ $(\gamma) H_1 : \mu \neq \mu_0$ Στατιστικό μέτρο: $z = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$	$\bar{x} = \frac{\sum_{i=1}^n X_i}{n}, \quad s^2 = \frac{\sum_{i=1}^n (X_i - \bar{x})^2}{n-1}$
$R = \{z > z_\alpha\}$	
$R = \{z < -z_\alpha\}$	
$R = \{ z > z_{\alpha/2}\}$	

(iii) Έλεγχος για τη **μέση τιμή** μ όταν σ **άγνωστο** και έχουμε **μικρό δείγμα**

$n < 30$ $H_0 : \mu = \mu_0$ $(\alpha) H_1 : \mu > \mu_0$ $(\beta) H_1 : \mu < \mu_0$ $(\gamma) H_1 : \mu \neq \mu_0$ Στατιστικό μέτρο: $t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$	$\bar{x} = \frac{\sum_{i=1}^n X_i}{n}, \quad s^2 = \frac{\sum_{i=1}^n (X_i - \bar{x})^2}{n-1}$ $t_{n-1;\alpha}$ η τιμή της t κατανομής για την οποία $\Psi_{n-1}(z) = 1 - \alpha$ $t_{n-1;\alpha/2}$ η τιμή της t κατανομής για την οποία $\Psi_{n-1}(z) = 1 - \frac{\alpha}{2}$
$R = \{t > t_{n-1;\alpha}\}$	
$R = \{t < -t_{n-1;\alpha}\}$	
$R = \{ t > t_{n-1;\alpha/2}\}$	

(iv) Έλεγχος για πιθανότητες κατηγοριών

$$H_0: p_1 = p_{10}, \dots, p_k = p_{k0}, H_1: p_i \neq p_{i0} \text{ για τουλάχιστον ένα } i$$

n_i η παρατηρούμενη συχνότητα της i -κατηγορίας

θ_i η αναμενόμενη συχνότητα της i -κατηγορίας

$$\text{Στατιστικό μέτρο: } X^2 = \sum_{i=1}^k \frac{(n_i - \theta_i)^2}{\theta_i}$$

$$\text{Απορριπτική περιοχή: } R = \{X^2 > X_{k-1, \alpha}^2\}$$

Ο Πίνακας δίνει τα άνω α -ποσοστιαία σημεία της κατανομής χ^2 με n βαθμούς ελευθερίας

n	$\alpha = 0.995$	$\alpha = 0.99$	$\alpha = 0.975$	$\alpha = 0.95$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$
1	0.000	0.000	0.001	0.004	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	11.070	12.832	15.086	16.750
6	0.676	0.872	1.237	1.635	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	14.067	16.013	18.475	20.278
8	1.344	1.647	2.180	2.733	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	31.414	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	36.415	39.364	42.980	45.558
25	10.520	11.524	13.120	14.611	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	38.885	41.923	45.642	48.290
27	11.808	12.878	14.573	16.151	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	41.337	44.461	48.278	50.994
29	13.121	14.256	16.047	17.708	42.557	45.722	49.588	52.335
30	13.787	14.953	16.791	18.493	43.773	46.979	50.892	53.672
40	20.706	22.164	24.4331	26.509	55.756	59.342	63.691	66.766
50	27.991	29.708	32.3574	34.764	67.505	71.420	76.154	79.490
60	35.535	37.485	40.4817	43.188	79.082	83.298	88.379	91.952
70	43.275	45.442	48.7576	51.739	90.531	95.023	100.425	104.215
80	51.172	53.540	57.1532	60.392	101.879	106.629	112.329	116.321
90	59.196	61.754	65.6466	69.126	113.145	118.136	124.116	128.299
100	67.328	70.065	74.2219	77.930	124.342	129.561	135.807	140.169

Τυποποιημένη κανονική κατανομή, τιμές της συνάρτησης ΑΣΚ $\Phi(y) = P(Y \leq y)$

	.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	.5675	.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	.8315	.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
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Πίνακας για την ΑΣΚ της t κατανομής με $n - 1$ β.ε.

- 1^η στήλη: β.ε.
- 1^η γραμμή: Επιθυμητή πιθανότητα β
- Περιεχόμενα: Η τιμή z ώστε $\Psi_{n-1}(z) = 1 - \beta$

	0.100	0.050	0.025	0.010	0.005	0.001
1	3.078	6.314	12.71	31.82	63.66	318.3
2	1.886	2.920	4.303	6.965	9.925	22.33
3	1.638	2.353	3.182	4.541	5.841	10.21
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.893
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.681	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
20	1.325	1.725	2.086	2.528	2.845	3.552
30	1.310	1.697	2.042	2.457	2.750	3.385
60	1.296	1.671	2.000	2.390	2.660	3.232
120	1.289	1.658	1.980	2.358	2.617	3.160
∞	1.282	1.645	1.960	2.326	2.576	3.090