Statistika

Tri važne raspodjele

1. Hi-kvadrat raspodjela

Slučajna veličina X ima *gama raspodjelu sa parametrima* α *i* β, $(\alpha, \beta > 0)$, ako je njena funkcija gustine

$$g(x;\alpha,\beta) = \begin{cases} \dfrac{x^{\alpha-1}}{\Gamma(\alpha)}\,\beta^{\alpha}\,e^{-\beta x}, & x > 0 \\ 0, & x \le 0. \end{cases}$$

Koristimo oznaku $X : \Gamma(\alpha, \beta)$.

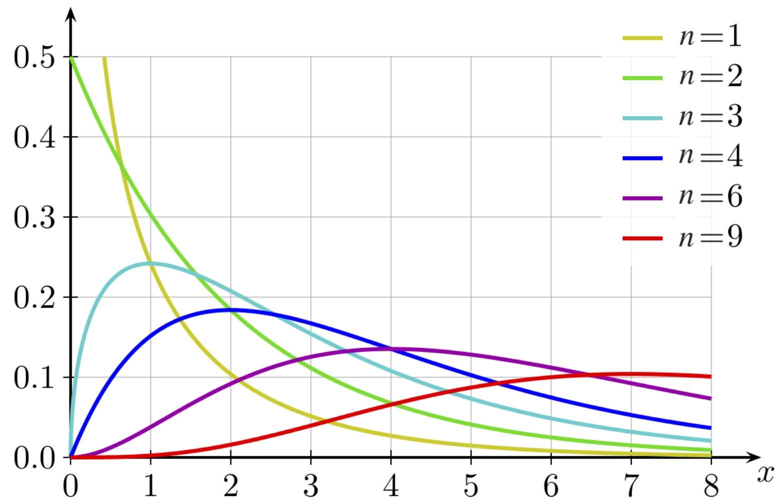
Karakteristična funkcija slučajne veličine X je $f_X(t) = \left(1 - \frac{it}{\beta}\right)^{-\alpha}$.

Neka je n prirodan broj.

Raspodjela $\Gamma(n/2, 1/2)$ naziva se *hi-kvadrat raspodjela sa* n *stepeni slobode* i označava sa χ^2_n.
Njenu funkciju gustine $g(x; n/2, 1/2)$ ćemo zbog jednostavnijeg zapisa označavati sa $h_n(x)$.

Često se slučajna veličina koja ima χ^2_n-raspodjelu označava istim simbolom χ^2_n, a da li se radi o raspodjeli ili slučajnoj veličini jasno je iz konteksta.

Karakteristična funkcija slučajne veličine χ^2_n je $f_{\chi^2_n}(t) = (1 - 2it)^{-n/2}$.



Slika 1. Funkcija gustine $h_n(x)$.

$$\mathsf{E}\chi_n^2 = n, \quad \mathsf{D}\chi_n^2 = 2n.$$

Zaista,

$$f'_{\chi_n^2}(t) = in(1-2it)^{-n/2-1}, \quad f''_{\chi_n^2}(t) = -n(n+2)(1-2it)^{-n/2-2},$$

$$\mathsf{E}\chi_n^2 = \frac{f'_{\chi_n^2}(0)}{i} = n, \quad \mathsf{D}\chi_n^2 = \frac{f''_{\chi_n^2}(0)}{i^2} - (\mathsf{E}\chi_n^2)^2 = 2n.$$

Ako su slučajne veličine χ_m^2 i χ_n^2 nezavisne, onda njihov zbir ima χ_{m+n}^2 raspodjelu.

Ovo direktno slijedi na osnovu Teoreme o jedinstvenosti za karakteristične funkcije, jer važi

$$f_{\chi_m^2+\chi_n^2}(t) = f_{\chi_m^2}(t) f_{\chi_n^2}(t) = (1-2it)^{-m/2}(1-2it)^{-n/2} = (1-2it)^{-(m+n)/2} = f_{\chi_{m+n}^2}(t).$$

Ako su $X_1, \ldots, X_n$ nezavisne slučajne veličine sa $\mathcal{N}(0,1)$ raspodjelom, onda

$$\sum_{i=1}^n X_i^2 : \chi_n^2.$$

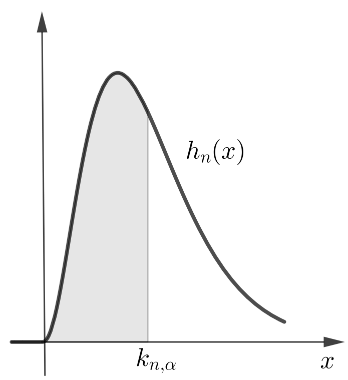
Zaista, funkcija gustine za X_1^2 je (dokazati!)

$$\varphi(x) = \begin{cases} \frac{1}{\sqrt{2\pi x}} e^{-x/2}, & x > 0 \\ 0, & x \leq 0. \end{cases}$$

pa je $X_1^2 : \Gamma(1/2, 1/2)$. Slijedi, karakteristična funkcija za X_1^2 je $(1-2it)^{-1/2}$, a odatle direktno dobijamo da važi tvrđenje.

U tablici za χ_n^2 raspodjelu nalaze se vrijednosti $k_{n,\alpha}$ za koje je

$$\int_0^{k_{n,\alpha}} h_n(x)dx = \alpha.$$



Slika 2. Funkcija gustine $h_n(x)$, za $n > 2$.

2. Studentova raspodjela

Neka su $X : \mathcal{N}(0,1)$ i $Y : \chi_n^2$ nezavisne slučajne veličine. Raspodjela slučajne veličine

$$\frac{X}{\sqrt{Y/n}}$$

naziva se *Studentova raspodjela sa n stepeni slobode.* Koristimo oznaku t_n.

Slučajna veličina koja ima t_n-raspodjelu često se označava istim simbolom t_n.

Funkcija gustine slučajne veličine t_n je

$$v_n(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma(\frac{n}{2})\sqrt{n\pi}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad x \in \mathbf{R}.$$

Specijalno za $n = 1$, Studentova raspodjela se poklapa sa Košijevom $\mathcal{C}(0, 1)$ raspodjelom.

Važi

$$\mathsf{E}t_n = 0, \quad (n > 1), \qquad \mathsf{D}t_n = \frac{n}{n-2} \quad (n > 2).$$

Neka je $n > 1$. Tada važi

$$\mathsf{E}t_n = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{n\pi}} \int_{-\infty}^{+\infty} x \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} dx = 0,$$

zato što je dati integral konvergentan za $n > 1$ (dokazati!), a funkcija $x(1 + \frac{x^2}{n})^{-\frac{n+1}{2}}$ neparna.

Uraditi za vježbu sljedeći zadatak.

Zadatak 1. Dokazati da je integral

$$\int_{-\infty}^{+\infty} x^2 \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} dx$$

konvergira samo ako je $n > 2$.

Neka je $n > 2$ i neka je $c_n = \dfrac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{n\pi}}$. Dobijamo

$$\begin{aligned}
\mathsf{D}t_n = \mathsf{E}X^2 &= c_n \int_{-\infty}^{+\infty} x^2 \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} dx \\
&= 2c_n \int_{0}^{+\infty} x^2 \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} dx \\
&= 2c_n \cdot n \int_{0}^{+\infty} \left(\frac{x^2}{n} + 1 - 1\right) \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} dx
\end{aligned}$$

$$
\begin{aligned}
&= 2c_n \cdot n \int_0^{+\infty} \left(1 + \frac{x^2}{n}\right)^{-\frac{n-2+1}{2}} dx - 2c_n \cdot n \int_0^{+\infty} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} dx \\
&= 2c_n \cdot n \cdot \frac{\sqrt{n}}{\sqrt{n-2}} \cdot \frac{c_{n-2}}{c_{n-2}} \int_0^{+\infty} x^2 \left(1 + \frac{\left(\frac{x\sqrt{n-2}}{\sqrt{n}}\right)^2}{n-2}\right)^{-\frac{n-2+1}{2}} d\left(\frac{x\sqrt{n-2}}{\sqrt{n}}\right) - n \\
&= \frac{c_n}{c_{n-2}} \cdot n \cdot \frac{\sqrt{n}}{\sqrt{n-2}} - n \\
&= \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{n\pi}} \cdot \frac{\Gamma(\frac{n-2}{2})\sqrt{(n-2)\pi}}{\Gamma(\frac{n-1}{2})} \cdot n \cdot \frac{\sqrt{n}}{\sqrt{n-2}} \cdot n - n \\
&= \frac{\Gamma(\frac{n-1}{2}+1)}{\Gamma(\frac{n-2}{2}+1)} \cdot \frac{\Gamma(\frac{n-2}{2})}{\Gamma(\frac{n-1}{2})} \cdot n - n \\
&= \frac{\frac{n-1}{2}\Gamma(\frac{n-1}{2})}{\frac{n-2}{2}\Gamma(\frac{n-2}{2})} \cdot \frac{\Gamma(\frac{n-2}{2})}{\Gamma(\frac{n-1}{2})} \cdot n - n \\
&= \frac{n}{n-2}.
\end{aligned}
$$

Ovdje smo koristili svojstvo gama funkcije $\Gamma(a+1) = a\Gamma(a)$ i da je

$$
2c_n \int_0^{+\infty} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} dx = \int_{-\infty}^{+\infty} c_n \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} dx = \int_{-\infty}^{+\infty} v_n(x) dx = 1.
$$

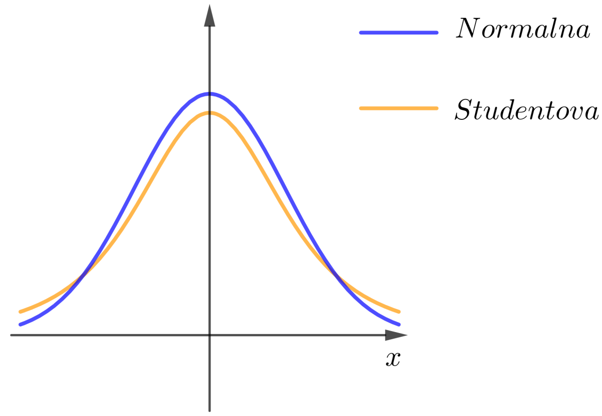
Na isti način je

$$
2c_{n-2} \int_0^{+\infty} \left(1 + \frac{\left(\frac{x\sqrt{n-2}}{\sqrt{n}}\right)^2}{n-2}\right)^{-\frac{n-2+1}{2}} d\left(\frac{x\sqrt{n-2}}{\sqrt{n}}\right) = 1.
$$

Grafik funkcije gustine za Studentovu raspodjelu ima približan oblik grafika funkcije gustine za normalnu raspodjelu i za veliko n normalna raspodjela dobro aproksimira Studentovu.

U tablici za t_n raspodjelu nalaze se vrijednosti $u_{n,\alpha}$ za koje je

$$
\int_0^{u_{n,\alpha}} v_n(x) dx = \alpha.
$$



Slika 3. Grafik funkcije gustine standardne normalne raspodjele i grafik funkcije gustine Studentove raspodjele

Fišerova teorema. Neka je $(X_1, \dots, X_n)$ uzorak obilježja X koje ima $\mathcal{N}(m.\sigma^2)$ raspodjelu. Tada važi

(1) Slučajna veličina $\dfrac{\overline{X}_n - m}{\sigma}\sqrt{n}$ ima $\mathcal{N}(0.1)$ raspodjelu.

(2) Slučajna veličina $\dfrac{n\overline{S}_n^2}{\sigma^2}$ ima χ^2_{n-1} raspodjelu.

(3) Slučajne veličine $\overline{X}_n$ i $\overline{S}_n^2$ su nezavisne.

(4) Slučajna veličina $\dfrac{\overline{X}_n - m}{\overline{S}_n}\sqrt{n-1}$ ima t_{n-1} raspodjelu.

Uraditi za vježbu sljedeći zadatak.

Zadatak 2. Dokazati tvrđenja (1), (2) i (4) Fišerove teoreme.

Zadatak 3. Neka su $X : \mathcal{N}(m_1, \sigma^2)$ i $Y : \mathcal{N}(m_2, \sigma^2)$ nezavisna obilježja, $(X_1, \dots, X_n)$ uzorak obilježja X i $(Y_1, \dots, Y_k)$ uzorak obilježja Y. Tada važi

$$\frac{(\overline{X}_n - m_1) - (\overline{Y}_k - m_2)}{\sqrt{n\overline{S}_n^2(X) + k\overline{S}_k^2(Y)}}\sqrt{\frac{nk}{n+k}(n+k-2)} : t_{n+k-2}.$$

Rješenje: Uočimo da važi sljedeće (obrazložiti detaljno! u kojem koraku se koristi nezavisnost slučajnih veličina i zašto su upravo te slučajne veličine nezavisne?):

$$\frac{\overline{X}_n - m_1}{\sigma} : \mathcal{N}\left(0, \frac{1}{n}\right), \qquad \frac{\overline{Y}_k - m_2}{\sigma} : \mathcal{N}\left(0, \frac{1}{k}\right),$$

$$\frac{\overline{X}_n - m_1}{\sigma} - \frac{\overline{Y}_k - m_2}{\sigma} : \mathcal{N}\left(0, \frac{1}{n} + \frac{1}{k}\right),$$

$$\frac{n\overline{S}_n^2(X)}{\sigma^2} : \chi^2_{n-1}, \qquad \frac{k\overline{S}_k^2(Y)}{\sigma^2} : \chi^2_{k-1},$$

$$\frac{n\overline{S}_n^2(X)}{\sigma^2} + \frac{k\overline{S}_k^2(Y)}{\sigma^2} : \chi^2_{n+k-2},$$

pa sada lako slijedi tvrđenje.

3. Snedekor-Fišerova raspodjela

Neka su $Y_1 : \chi^2_{n_1}$ i $Y_2 : \chi^2_{n_2}$ nezavisne slučajne veličine. Raspodjela slučajne veličine

$$\frac{Y_1/n_1}{Y_2/n_2}$$

naziva se *Snedekor-Fišerova raspodjela sa n_1 i n_2 stepeni slobode.* Koristimo oznaku F_{n_1,n_2}.

Primijetimo da su $F_{1,n}$ i t_n^2 iste raspodjele.

Funkcija gustine za F_{n_1,n_2} je

$$f_{n_1,n_2}(x) = \begin{cases} \dfrac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} \left(\dfrac{n_1}{n_2}\right)^{\frac{n_1}{2}} x^{\frac{n_1}{2}-1} \left(1 + \dfrac{n_1 x}{n_2}\right)^{-\frac{n_1+n_2}{2}}, & x > 0, \\ 0, & x \leqslant 0. \end{cases}$$

Zadatak 4. Neka su $X : \mathcal{N}(m_1, \sigma_1^2)$ i $Y : \mathcal{N}(m_2, \sigma_2^2)$ nezavisna obilježja, $(X_1, \dots, X_n)$ uzorak obilježja X i $(Y_1, \dots, Y_k)$ uzorak obilježja Y. Tada važi

$$\frac{n(k-1)\sigma_2^2\,\overline{S}_n^2(X)}{k(n-1)\sigma_1^2\,\overline{S}_k^2(Y)} : F_{n-1,k-1}.$$

TABLICA χ^2 RASPODJELE

$$\int_0^{k_{n,\alpha}} h_n(x)\ \mathrm{d}x = \alpha$$

$n \backslash \alpha$	0.005	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990	0.995
1	—	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672

Tablica Studentove raspodjele

$$\int_0^{u_{n,\alpha}} v_n(x)\ \mathrm{d}x = \alpha$$

$n \backslash \alpha$	0.10	0.20	0.30	0.40	0.45	0.475	0.49	0.495
1	0.325	0.727	1.367	3.078	6.314	12.706	31.821	63.657
2	0.289	0.617	1.061	1.886	2.920	4.303	6.965	9.925
3	0.277	0.584	0.978	1.638	2.353	3.182	4.541	5.841
4	0.271	0.569	0.941	1.533	2.132	2.776	3.747	4.604
5	0.267	0.559	0.920	1.476	2.015	2.571	3.365	4.032
6	0.265	0.553	0.906	1.440	1.943	2.447	3.143	3.707
7	0.263	0.549	0.896	1.415	1.895	2.365	2.998	3.499
8	0.262	0.546	0.889	1.397	1.860	2.306	2.896	3.355
9	0.261	0.543	0.883	1.383	1.833	2.262	2.821	3.250
10	0.260	0.542	0.879	1.372	1.812	2.228	2.764	3.169
11	0.260	0.540	0.876	1.363	1.796	2.201	2.718	3.106
12	0.259	0.539	0.873	1.356	1.782	2.179	2.681	3.055
13	0.259	0.538	0.870	1.350	1.771	2.160	2.650	3.012
14	0.258	0.537	0.868	1.345	1.761	2.145	2.624	2.977
15	0.258	0.536	0.866	1.341	1.753	2.131	2.602	2.947
16	0.258	0.535	0.865	1.337	1.746	2.120	2.583	2.921
17	0.257	0.534	0.863	1.333	1.740	2.110	2.567	2.898
18	0.257	0.534	0.862	1.330	1.734	2.101	2.552	2.878
19	0.257	0.533	0.861	1.328	1.729	2.093	2.539	2.861
20	0.257	0.533	0.860	1.325	1.725	2.086	2.528	2.845
21	0.257	0.532	0.859	1.323	1.721	2.080	2.518	2.831
22	0.256	0.532	0.858	1.321	1.717	2.074	2.508	2.819
23	0.256	0.532	0.858	1.319	1.714	2.069	2.500	2.807
24	0.256	0.531	0.857	1.316	1.708	2.060	2.485	2.787
25	0.256	0.531	0.856	1.316	1.708	2.060	2.485	2.787
26	0.256	0.531	0.856	1.315	1.706	2.056	2.479	2.779
27	0.256	0.531	0.855	1.314	1.703	2.052	2.473	2.771
28	0.256	0.530	0.855	1.313	1.701	2.048	2.467	2.763
29	0.256	0.530	0.854	1.311	1.699	2.045	2.462	2.756
30	0.256	0.530	0.854	1.310	1.697	2.042	2.457	2.750
40	0.255	0.529	0.851	1.303	1.684	2.021	2.423	2.704
60	0.254	0.527	0.848	1.296	1.671	2.000	2.390	2.660
120	0.254	0.526	0.845	1.289	1.658	1.980	2.358	2.617
∞	0.253	0.524	0.842	1.282	1.645	1.960	2.326	2.576

- Fišerova raspodela. Vrednosti $F_\alpha(d_1,d_2)$ za koje važi $P(F(d_1,d_2)\geq F_\alpha(d_1,d_2)) = \alpha$

$\alpha = 0.05$

$d_2 \backslash d_1$	1	2	3	4	5	6	8	10	12	15	20	24	30
1	161.45	199.50	215.71	224.58	230.16	233.99	238.88	241.88	243.90	245.95	248.02	249.05	250.10
2	18.51	19.00	19.16	19.25	19.30	19.33	19.37	19.40	19.41	19.43	19.45	19.45	19.46
3	10.13	9.55	9.28	9.12	9.01	8.94	8.85	8.79	8.74	8.70	8.66	8.64	8.62
4	7.71	6.94	6.59	6.39	6.26	6.16	6.04	5.96	5.91	5.86	5.80	5.77	5.75
5	6.61	5.79	5.41	5.19	5.05	4.95	4.82	4.74	4.68	4.62	4.56	4.53	4.50
6	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.06	4.00	3.94	3.87	3.84	3.81
7	5.59	4.74	4.35	4.12	3.97	3.87	3.73	3.64	3.57	3.51	3.44	3.41	3.38
8	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.35	3.28	3.22	3.15	3.12	3.08
9	5.12	4.26	3.86	3.63	3.48	3.37	3.23	3.14	3.07	3.01	2.94	2.90	2.86
10	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.98	2.91	2.85	2.77	2.74	2.70
11	4.84	3.98	3.59	3.36	3.20	3.09	2.95	2.85	2.79	2.72	2.65	2.61	2.57
12	4.75	3.89	3.49	3.26	3.11	3.00	2.85	2.75	2.69	2.62	2.54	2.51	2.47
13	4.67	3.81	3.41	3.18	3.03	2.92	2.77	2.67	2.60	2.53	2.46	2.42	2.38
14	4.60	3.74	3.34	3.11	2.96	2.85	2.70	2.60	2.53	2.46	2.39	2.35	2.31
15	4.54	3.68	3.29	3.06	2.90	2.79	2.64	2.54	2.48	2.40	2.33	2.29	2.25
16	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.49	2.42	2.35	2.28	2.24	2.19
17	4.45	3.59	3.20	2.96	2.81	2.70	2.55	2.45	2.38	2.31	2.23	2.19	2.15
18	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.41	2.34	2.27	2.19	2.15	2.11
19	4.38	3.52	3.13	2.90	2.74	2.63	2.48	2.38	2.31	2.23	2.16	2.11	2.07
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.35	2.28	2.20	2.12	2.08	2.04
21	4.32	3.47	3.07	2.84	2.68	2.57	2.42	2.32	2.25	2.18	2.10	2.05	2.01
22	4.30	3.44	3.05	2.82	2.66	2.55	2.40	2.30	2.23	2.15	2.07	2.03	1.98
23	4.28	3.42	3.03	2.80	2.64	2.53	2.37	2.27	2.20	2.13	2.05	2.01	1.96
24	4.26	3.40	3.01	2.78	2.62	2.51	2.36	2.25	2.18	2.11	2.03	1.98	1.94
25	4.24	3.39	2.99	2.76	2.60	2.49	2.34	2.24	2.16	2.09	2.01	1.96	1.92
30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.16	2.09	2.01	1.93	1.89	1.84
40	4.08	3.23	2.84	2.61	2.45	2.34	2.18	2.08	2.00	1.92	1.84	1.79	1.74
60	4.00	3.15	2.76	2.53	2.37	2.25	2.10	1.99	1.92	1.84	1.75	1.70	1.65