

Metodi Matematici e Informatici per la Biologia

a.a. 2024-2025

(Agliari - Vittorio - Presilla - Simonella)

30 giugno 2025

Prova di Verifica I

Nome:

Cognome:

Matricola:

Esercizio 1 (3+3+2=8 punti)

Le temperature medie registrate in 23 diverse località forniscono, in gradi centigradi, i seguenti valori:

6.40, 4.44, 7.49, 9.55, 9.17, 15.15, 14.79, 8.87, 5.12, 2.20, 7.61, 5.91, 15.17, 31.13, 7.04, 22.27, 21.62, 8.84, 19.74, 26.71, 11.16, 43.71, 19.15

1.1 Calcolare il minimo, il massimo, l'indice di variazione (range), la media e la deviazione standard dei dati raccolti, scrivendo esplicitamente le relative definizioni.

Ordiniamo, innanzitutto, il campione di $n = 23$ dati in modo crescente:

$x_1 = 2.20, x_2 = 4.44, x_3 = 5.12, x_4 = 5.91, x_5 = 6.40, x_6 = 7.04, x_7 = 7.49,$
 $x_8 = 7.61, x_9 = 8.84, x_{10} = 8.87, x_{11} = 9.17, x_{12} = 9.55, x_{13} = 11.16, x_{14} = 14.79,$
 $x_{15} = 15.15, x_{16} = 15.17, x_{17} = 19.15, x_{18} = 19.74, x_{19} = 21.62, x_{20} = 22.27, x_{21} = 26.71,$
 $x_{22} = 31.13, x_{23} = 43.71$

Segue immediatamente che

$$x_{\min} = x_1 = 2.20 \text{ }^{\circ}\text{C}, \quad x_{\max} = x_{23} = 43.71 \text{ }^{\circ}\text{C}, \quad \text{range} = x_{\max} - x_{\min} = 41.51 \text{ }^{\circ}\text{C}.$$

Inoltre

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 14.05 \text{ }^{\circ}\text{C}, \quad \sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = 9.99 \text{ }^{\circ}\text{C}.$$

1.2 Calcolare i quartili del campione, la distanza interquartile e mettere in evidenza eventuali dati outlier.

Poiché $23 \cdot 1/4 = 5.75$, $23 \cdot 2/4 = 11.5$ e $23 \cdot 3/4 = 17.25$ non sono numeri interi i quartili Q_1 , Q_2 e Q_3 sono dati da

$$\begin{aligned}Q_1 &= x_{\lceil 5.75 \rceil} = x_6 = 7.04 \text{ }^\circ\text{C}, \\Q_2 &= x_{\lceil 11.5 \rceil} = x_{12} = 9.55 \text{ }^\circ\text{C}, \\Q_3 &= x_{\lceil 17.25 \rceil} = x_{18} = 19.74 \text{ }^\circ\text{C}\end{aligned}$$

La distanza interquartile è $\Delta Q = Q_3 - Q_1 = 19.74 - 7.04 = 12.7 \text{ }^\circ\text{C}$.

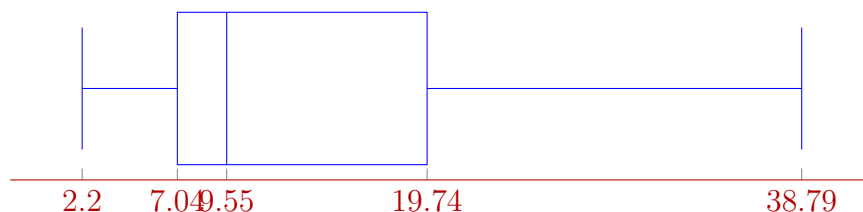
Normalmente si considerano outliers i dati esterni all'intervallo

$$[Q_1 - 1.5\Delta Q, Q_3 + 1.5\Delta Q] = [7.04 - 19.05, 19.74 + 19.05] = [-12.01, 38.79]$$

pertanto nel campione abbiamo un solo dato outlier: $x_{23} = 43.71 \text{ }^\circ\text{C}$.

1.3 Si rappresenti il boxplot del campione di dati, specificando tutti i relativi valori di riferimento.

minimo (esclusi outliers): 2.2, primo quartile: 7.04, mediana: 9.55, terzo quartile: 19.74, massimo (esclusi outliers): 38.79



Esercizio 2 (3+2+2=7 punti)

In primavera la rana Kermit caccia insetti per nutrirsi catturando in media 10 insetti all'ora.

2.1 Si indichi una ragionevole distribuzione di probabilità per calcolare il numero di insetti catturati in un assegnato intervallo di tempo Δt .

Il numero di tentativi di cattura di un insetto in un intervallo di tempo Δt non è noto e il numero di possibili successi non è limitato. Detta $X_{\Delta t}$ la variabile aleatoria discreta numero di insetti catturati nell'intervallo di tempo Δt , è ragionevole assumere per essa una distribuzione di Poisson con PMF

$$P(X_{\Delta t} = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \dots, \quad \lambda > 0.$$

Il parametro λ è determinato dal numero medio di insetti catturati nell'intervallo di tempo Δt in base alla nota relazione

$$E(X_{\Delta t}) = \lambda.$$

2.2 Utilizzando tale distribuzione, si calcoli la probabilità che la rana catturi in 1 minuto al più 1 insetto

Esprimiamo i tempi in minuti. Sappiamo che la rana cattura in media 10 insetti in 60 minuti cioè $10/60 \simeq 0.17$ insetti al minuto. La distribuzione di Poisson che determina $X_{\Delta t=1}$ ha quindi parametro $\lambda = 0.17$. La probabilità richiesta è

$$\begin{aligned} P(X_{\Delta t=1} \leq 1) &= P(X_{\Delta t=1} = 0) + P(X_{\Delta t=1} = 1) \\ &= e^{-0.17}(1 + 0.17) \\ &= 0.987. \end{aligned}$$

2.3 e la probabilità che in 2 minuti catturi più di 2 insetti.

La distribuzione di Poisson che determina $X_{\Delta t=2}$ ha parametro $\lambda = 2 \cdot 10/60 \simeq 0.34$. La probabilità richiesta è

$$\begin{aligned} P(X_{\Delta t=2} > 2) &= 1 - \left(P(X_{\Delta t=2} = 0) + P(X_{\Delta t=2} = 1) + P(X_{\Delta t=2} = 2) \right) \\ &= 1 - e^{-0.34} \left(1 + 0.34 + \frac{(0.34)^2}{2} \right) \\ &= 1 - 0.995 \\ &= 0.005 \end{aligned}$$

Esercizio 3 (3+2+2=7 punti)

Al Dipartimento di Medicina Cardiovascolare dell'Università di Pisa è stata condotta una ricerca per testare fino a che punto l'ipertensione è un fenomeno genetico. A tale scopo sono state esaminate 20 famiglie, prendendo la pressione arteriosa di padre (x) e primogenito (y). Sono stati ottenuti i seguenti risultati:

$$\sum_{i=1}^{20} x_i = 2980 \text{ mmHg}, \quad \sum_{i=1}^{20} y_i = 2030 \text{ mmHg},$$

$$\sum_{i=1}^{20} x_i^2 = 451350 \text{ mmHg}^2, \quad \sum_{i=1}^{20} y_i^2 = 210850 \text{ mmHg}^2, \quad \sum_{i=1}^{20} x_i y_i = 305700 \text{ mmHg}^2.$$

3.1 Calcolare il coefficiente di correlazione tra le variabili x e y ;

Il coefficiente di correlazione è il numero

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

dove, per n dati, abbiamo definito

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2, \quad S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2, \quad S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}).$$

con

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i,$$

Nel nostro caso, in cui $n = 20$, risulta:

$$\bar{x} = 149 \text{ mmHg}, \quad \bar{y} = 101.5 \text{ mmHg}$$

$$S_{xx} = 7330 \text{ mmHg}^2, \quad S_{yy} = 4805 \text{ mmHg}^2, \quad S_{xy} = 3230 \text{ mmHg}^2.$$

Pertanto

$$r = \frac{3230}{\sqrt{7330 \cdot 4805}} \simeq 0.544.$$

Osserviamo che esiste una correlazione debole tra le due variabili.

3.2 Determinare i coefficienti a, b della retta di regressione lineare $y = ax + b$.

Si ha

$$a = \frac{S_{xy}}{S_{xx}} = \frac{3230}{7330} \simeq 0.44,$$

con a numero adimensionale, e

$$b = \bar{y} - a\bar{x} \simeq 101.5 - 0.44 \cdot 149 \simeq 35.94 \text{ mmHg}.$$

3.3 Qual'è la pressione media attesa del figlio se la pressione del padre è 155 mmHg? Viceversa, qual'è la pressione media attesa del padre se la pressione del figlio è 155 mmHg?

Utilizzando la retta di regressione lineare, per $x = 155$ mmHg stimiamo

$$y = ax + b \simeq 0.44 \cdot 155 + 35.94 \simeq 104.14 \text{ mmHg}.$$

Analogamente, per $y = 155$ mmHg stimiamo

$$x = \frac{y - b}{a} \simeq \frac{155 - 35.94}{0.44} \simeq 270.59 \text{ mmHg}.$$

Quest'ultimo risultato poco realistico è frutto della scarsa correlazione tra i dati x e y .

Esercizio 4 (2+2+2=6 punti)

I seguenti $n = 48$ valori costituiscono un campione di una variabile aleatoria

5.64 10.03 11.19 8.38 10.45 6.65 10.58 8.99
7.62 8.51 6.51 4.78 11.22 11.21 6.91 3.88
9.20 3.89 5.89 10.84 7.97 12.47 8.73 10.79
10.52 5.45 12.84 9.27 12.77 9.54 8.11 9.42
7.57 11.08 9.38 5.80 8.52 7.48 7.46 11.56
6.49 8.61 11.07 7.58 11.10 10.08 12.60 7.42

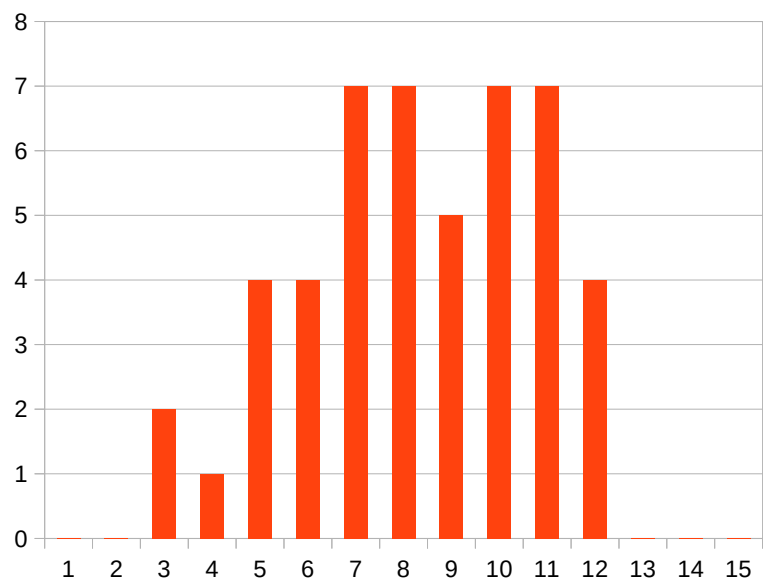
4.1 Determinare valore minimo, valore massimo, valore medio e deviazione standard del campione.

valore minimo = 3.88, valore massimo = 12.84,
valore medio = 8.83, deviazione standard = 2.33.

4.2 Calcolare le frequenze dei dati del campione usando classi di ampiezza 1 e quindi produrre il relativo istogramma.

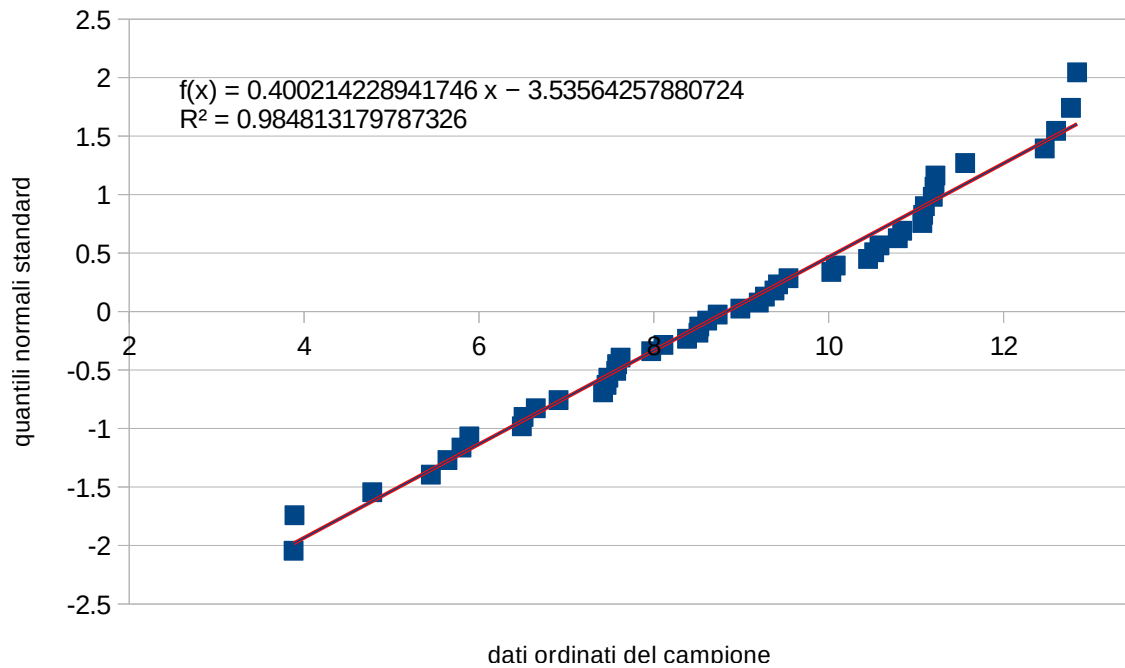
classe frequenza

(2,3]	0
(3,4]	2
(4,5]	1
(5,6]	4
(6,7]	4
(7,8]	7
(8,9]	7
(9,10]	5
(10,11]	7
(11,12]	7
(12,13]	4
(13,14]	0



4.3 Giudicare il carattere normale dei dati del campione realizzando un normal plot e riportando il coefficiente di determinazione della corrispondente retta di regressione lineare.

Il normal plot mostra che i dati hanno una distribuzione quasi normale; il coefficiente di determinazione della retta di regressione lineare vale $r^2 \simeq 0.985$.



Esercizio 5 (1+2+3=6 punti)

Viene effettuato lo studio di farmaco antivirale per evitare l'ospedalizzazione. Il farmaco è somministrato a un campione aleatorio di 250 pazienti, mentre ad altri 260 pazienti viene somministrato un placebo. Del primo gruppo vengo ospedalizzati 18 pazienti e del secondo gruppo 33.

5.1 Costruire la tabella di contingenza (frequenze osservate) che riassume i risultati dello studio.

	farmaco	placebo	totale
ospedalizzati	18	33	51
non ospedalizzati	232	227	459
totale	250	260	510

5.2 Costruire la tabella delle frequenze attese in base all'ipotesi nulla H_0 che il farmaco sia inefficace.

	farmaco	placebo	totale
ospedalizzati	$\frac{250 \times 51}{510} = 25$	$\frac{260 \times 51}{510} = 26$	51
non ospedalizzati	$\frac{250 \times 459}{510} = 225$	$\frac{260 \times 459}{510} = 234$	459
totale	250	260	510

5.3 Specificare il tipo di test con cui si intende effettuare la verifica di H_0 e determinare il livello di significatività con cui possiamo ritenere il farmaco inefficace.

Si utilizza il test del χ^2 con $(2 - 1) \times (2 - 1) = 1$ gradi di libertà. Le frequenze osservate sono $F_1 = 18$, $F_2 = 33$, $F_3 = 232$, $F_4 = 227$ mentre quelle attese in base all'ipotesi H_0 risultano $E_1 = 25$, $E_2 = 26$, $E_3 = 225$, $E_4 = 234$. La stima del valore di χ^2 corrispondente al test (statistica del test) è quindi

$$\chi_{\text{test}}^2 = \sum_{i=1}^4 \frac{(E_i - F_i)^2}{E_i} = \frac{(25 - 18)^2}{25} + \frac{(26 - 33)^2}{26} + \frac{(225 - 232)^2}{225} + \frac{(234 - 227)^2}{234} \simeq 4.27.$$

Per 1 grado di libertà risulta $P(\chi^2 \geq \chi_{\text{test}}^2) = \text{CHIDIST}(4.27, 1) \simeq 0.0388$. Possiamo pertanto accettare l'ipotesi H_0 , cioè affermare che il farmaco è inefficace, a un livello di significatività del 3.88%.

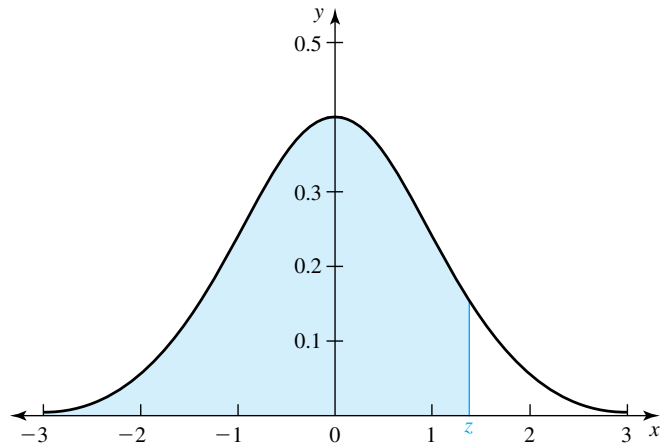
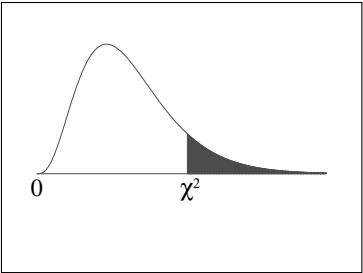


Figure B.1 Areas under the standard normal curve from $-\infty$ to z .

z	0	1	2	3	4	5	6	7	8	9
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5754
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	.7258	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7518	.7549
0.7	.7580	.7612	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7996	.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

Chi-Square Distribution Table



The shaded area is equal to α for $\chi^2 = \chi^2_{\alpha}$.

df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	0.000	0.000	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
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169