

# Tentamen (Wiskundige) Statistiek Module 5 Statistiek en Analyse

Bachelor 2 Technische Wiskunde

Modulecode: 201400218  
Datum: vrijdag 28 oktober 2016  
Tijd: van 8:45 tot 11:45 uur  
Blokcoördinator: Dr. P.K. Mandal  
Docent: Dr. K. Poortema

Toegestane middelen: alleen gewone rekenmachine (geen GR)

Bijlagen: standaardnormale tabel,  
t-tabel,  
chi-kwadraat-tabel,  
Tabel F verdeling (1 pagina,  $\alpha = 0.05$ )  
formuleblad Wiskundige Statistiek (2 pagina's)

Normering:

1a	1b	1c	2a	2b	3a	3b	3c	3d	4a	4b	5	6a	6b	7a	7b
2	2	2	4	2	1	2	2	2	3	3	2	1	2	4	2

Totaal: 36 punten

### Opgave 1

Een groep van 850 patiënten die recent een bypass heeft gekregen wordt onderverdeeld in twee therapiegroepen. Groep A ( $n = 450$ ) krijgt Marcoumar en groep B ( $n = 400$ ) krijgt Sintrom om te voorkomen dat de bypass dichtslibt. Na 1 jaar is van 70 patiënten in groep A de bypass dichtgeslibd en in groep B is de bypass dichtgeslibd van 45 patiënten.

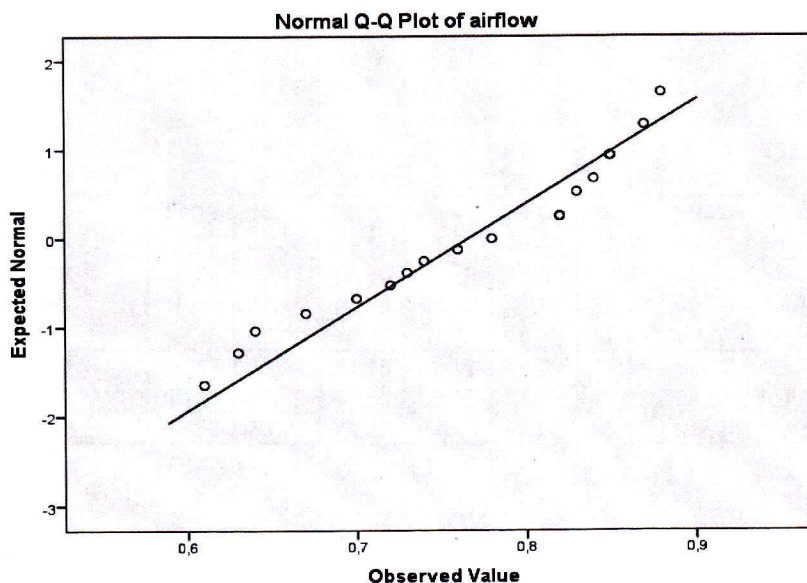
- Bepaal het 95%-betrouwbaarheidsinterval voor het verschil in kans op dichtslibben van de bypass (na 1 jaar) voor de twee therapiegroepen ("A minus B").
- Leg uit hoe je het berekende betrouwbaarheidsinterval moet interpreteren. Geef daarbij ook aan wat "betrouwbaarheid van 95%" betekent.
- Stel we willen een groter (vervolg)onderzoek gaan doen dat een 95%-betrouwbaarheidsinterval voor het verschil in kans op dichtslibben van de bypass oplevert dat (naar verwachting) een lengte heeft (verschil tussen rechter en linker grens) van ten hoogste 1%. Hoeveel patiënten hebben we voor dit nieuwe onderzoek nodig als we uitgaan twee even grote groepen?

### Opgave 2

De productie van een wasmiddel heeft tot gevolg dat productiemedewerkers langdurig blootgesteld worden aan het enzym Bacillus Subtilis met als mogelijk gevolg een vermindering van de 'airflow rate'. De airflow rate is de verhouding van het maximale volume lucht die een persoon kan uitademen gedurende 1 seconde en het maximale volume lucht die een persoon kan uitademen na zo diep mogelijk te hebben ingeademd. Voor personen zonder longstoornis geldt een norm van 0.80 voor de airflow rate. Van 19 aselect gekozen medewerkers zijn de airflow rates gemeten met als resultaat (de waarden zijn gerangschikt van klein naar groot):

0.61 0.63 0.64 0.67 0.70 0.72 0.73 0.74 0.76 0.78  
0.82 0.82 0.82 0.83 0.84 0.85 0.85 0.87 0.88

Steekproefgemiddelde: 0.7663  
Steekproefstandaardafwijking: 0.08591



### Tests of Normality

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
airflow	,208	19	,030	,924	19	,132

a. Lilliefors Significance Correction

a. Ga uit van een normale verdeling. Ga met een (statistische) toets na of de verwachting van de airflow rate van de medewerkers lager is dan 0.80 .

Neem significantieniveau  $\alpha = 5\%$  en volg het schema van acht stappen (zie formuleblad).

b. In de output van hierboven kun je vinden dat de P-waarde (overschrijdingskans) van de Shaapiro-Wilk toetsingsgrootheid gelijk is aan 0.132. Geef aan wat de correcte conclusie is ten aanzien van de aanname van een normale verdeling op grond van de Shapiro-Wilk toets. Motiveer (kort) je antwoord.

### Opgave 3

We hebben onafhankelijke waarnemingen  $X_1, X_2, \dots, X_n$  die alle exponentieel verdeeld zijn met parameter  $\lambda$ , die onbekend is. We willen twee schatters van  $\mu = 1/\lambda$  vergelijken, te weten:

$$\bar{X} = (X_1 + X_2 + \dots + X_n)/n \quad \text{en} \quad T = n \times U \quad \text{met} \quad U = \min(X_1, X_2, \dots, X_n)$$

a. Toon aan dat de kansdichtheid van  $U$  als volgt is:

$$f_U(u) = n\lambda \exp(-n\lambda u) \text{ als } u > 0, \quad f_U(u) = 0 \text{ elders.}$$

b. Ga voor beide schatters na of ze zuivere schatters (unbiased estimators) van  $\mu$  zijn.

c. Bereken voor beide schatters de verwachte kwadratische fout (Mean Squared Error).

Welke van de twee schatters is beste?

d. Ga voor beide schatters na of ze consistente schatters (consistent estimators) van  $\mu$  zijn.

### Opgave 4

We hebben onafhankelijke waarnemingen  $X_1, X_2, \dots, X_n$  die alle geometrisch verdeeld zijn met parameter  $p$ :

$$P(X_i = x_i) = p(1 - p)^{x_i - 1} \quad \text{voor } x_i = 1, 2, \dots$$

a. Toon aan dat de schatter van de methode van de grootste aannemelijkheid (maximum likelihood estimator) als volgt is :  $\hat{p} = 1/\bar{X}$  met  $\bar{X} = (X_1 + X_2 + \dots + X_n)/n$ .

b. Leid met behulp van het lemma van Neyman-Pearson de MP-toets af voor het toetsen van  $H_0: p = 0.1$  tegen  $H_1: p = 0.2$  : druk de toetsingsgrootheid uit als functie van  $\bar{X}$  en bepaal voor welke waarden van  $\bar{X}$  je de nulhypothese moet verwerpen, kritieke waarde(n) bepalen hoeft niet. Bepaal, indien mogelijk, de UMP-toets voor het toetsen van  $H_0: p = 0.1$  tegen  $H_1: p > 0.1$  (kritieke waarde(n) bepalen hoeft niet).

### Opgave 5

We schatten een (onbekende) parameter  $\vartheta$  door middel van een schatter  $T$ . Op grond van de ongelijkheid van Chebyshev kan de kans  $P(|T - \vartheta| > \varepsilon)$  voor willekeurige positieve getallen  $\varepsilon$  afgeschat worden met behulp van de verwachte kwadratische fout (Mean Squared Error)  $MSE(T)$ .

Geef de bedoelde afschatting en toon deze aan.

### Opgave 6

In multiple regressie kan de modelvergelijking geschreven worden als  $Y = X\beta + \varepsilon$  en kan de vector van kleinste kwadraten schatters geschreven worden als  $\hat{\beta} = (X^T X)^{-1} X^T Y$ , uitgaande van de gebruikelijke notatie van multiple regressie.

De schatter  $\hat{\beta}$  heeft een multivariate normale verdeling,

- Geef voor deze multivariate normale verdeling de vector van verwachtingen en de variantie-covariantie matrix (variance covariance matrix).
- Toon de multivariate normale verdeling van  $\hat{\beta}$  aan, uitgaande van de gebruikelijke aannames voor de vector van storingen  $\varepsilon$ .

### Opgave 7

Om een ijkpunt te hebben bij het beoordelen van bodemverontreiniging werd in Engeland een voorstudie gehouden naar het voorkomen van PCB's. Op 14 locaties op het platteland en 15 in de stad werd de PCB concentratie gemeten (in 0.0001 g per kg grond).

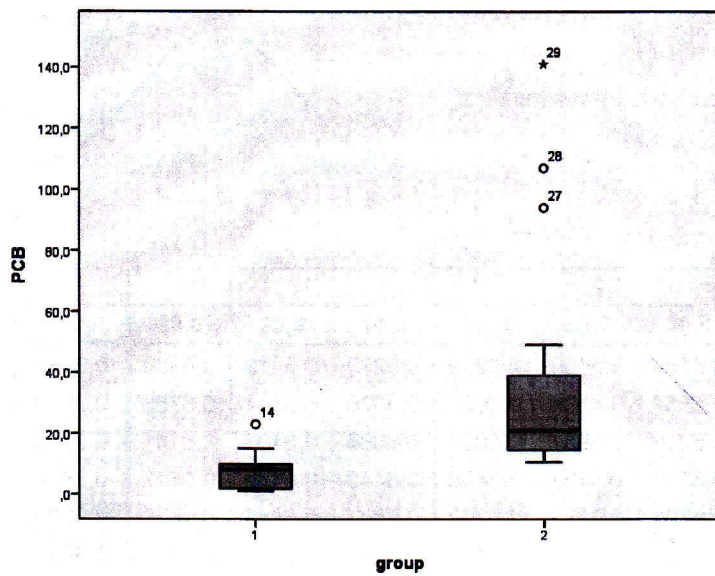
De onderzoekers melden in Chemosphere (Feb. 1986) dat er geen significant verschil is tussen platteland en stad.

De gegevens zijn als volgt (reeds geordend per steekproef):

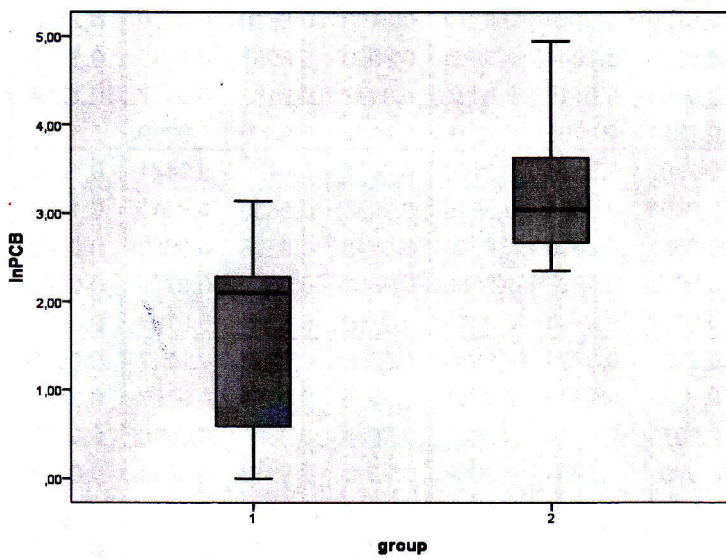
platteland	1.0	1.5	1.6	1.8	3.5	5.3	8.1	8.2	9.0	9.7
	9.8	12.0	15.0	23.0						
stad	10.5	11.0	11.5	13.0	16.0	17.5	18.0	21.0	22.0	24.0
	29.0	49.0	94.0	107.0	141.0					

- Ga met een geschikte verdelingsvrije toets na of je het eens bent met de onderzoekers. Neem onbetrouwbaarheidsdrempel 5% en volg het schema van acht stappen (zie formuleblad).

We willen ook de toepasbaarheid van de toets van het vorige onderdeel onderzoeken. We hebben boxplots laten maken voor platteland (group 1) en stad (group 2) en steekproefgemiddelden en steekproefstandaardafwijkingen berekend. Dit hebben we herhaald voor een veel gehanteerde transformatie van de data, de (natuurlijke) logaritme (de meting heet na transformatie  $\ln(\text{PCB})$ ).



Steekproefgemiddelden zijn respectievelijk 7.821 en 38.967.  
 Steekproefstandaardafwijkingen zijn respectievelijk 6.1702 en 41.0154.



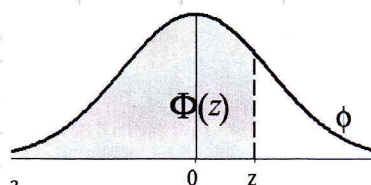
Steekproefgemiddelden zijn respectievelijk 1.6900 en 3.2680.  
 Steekproefstandaardafwijkingen zijn respectievelijk 0.97818 en 0.85520.

b. Geeft deze informatie aanleiding om te zeggen dat we de toets van het vorige onderdeel beter maar niet hadden moeten toepassen? Motiveer je antwoord.

## De standaardnormale verdeling

De tabel geeft de verdelingsfunctie  $\Phi$  voor een  $N(0, 1)$ -verdeelde  $Z$

$$\Phi(z) = P(Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{x^2}{2}} dx$$



Laatste kolom: de dichtheid van de  $N(0,1)$ -verdeling:  $\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$

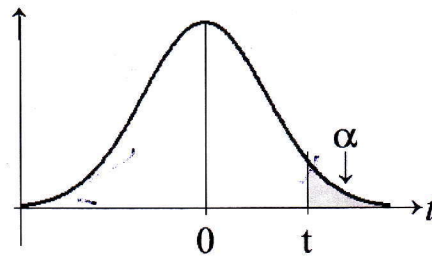
z	Tweede decimaal z										$\varphi(z)$
	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359	0,3989
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753	0,3970
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141	0,3910
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517	0,3814
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879	0,3683
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224	0,3521
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549	0,3332
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852	0,3123
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133	0,2897
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389	0,2661
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621	0,2420
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830	0,2179
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015	0,1942
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177	0,1714
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319	0,1497
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441	0,1295
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545	0,1109
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633	0,0940
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706	0,0790
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767	0,0656
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817	0,0540
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857	0,0440
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890	0,0355
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916	0,0283
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936	0,0224
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952	0,0175
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964	0,0136
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974	0,0104
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981	0,0079
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986	0,0060
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990	0,0044
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993	0,0033
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995	0,0024
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997	0,0017
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998	0,0012
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,0009
3,6	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,0006

# Tabel t-verdeling

Gegeven zijn de kritieke waarden  $t$  bij de rechter staartkansen, zodat

$$P(T \geq t) = \alpha$$

$f(t)$

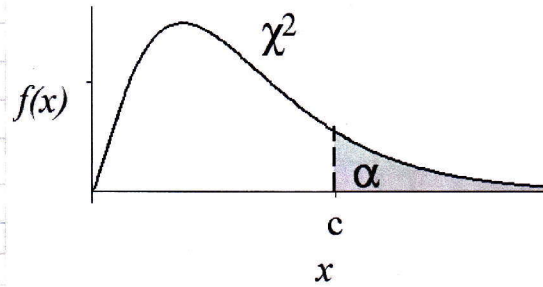


Aantal vrijheidsgraden	$\alpha$							
	0,25	0,10	0,05	0,025	0,01	0,005	0,001	0,0005
1	1,000	3,078	6,314	12,706	31,821	63,657	318,31	636,62
2	0,816	1,886	2,920	4,303	6,965	9,925	22,327	31,599
3	0,765	1,638	2,353	3,182	4,541	5,841	10,215	12,924
4	0,741	1,533	2,132	2,776	3,747	4,604	7,173	8,610
5	0,727	1,476	2,015	2,571	3,365	4,032	5,893	6,869
6	0,718	1,440	1,943	2,447	3,143	3,707	5,208	5,959
7	0,711	1,415	1,895	2,365	2,998	3,499	4,785	5,408
8	0,706	1,397	1,860	2,306	2,896	3,355	4,501	5,041
9	0,703	1,383	1,833	2,262	2,821	3,250	4,297	4,781
10	0,700	1,372	1,812	2,228	2,764	3,169	4,144	4,587
11	0,697	1,363	1,796	2,201	2,718	3,106	4,025	4,437
12	0,695	1,356	1,782	2,179	2,681	3,055	3,930	4,318
13	0,694	1,350	1,771	2,160	2,650	3,012	3,852	4,221
14	0,692	1,345	1,761	2,145	2,624	2,977	3,787	4,140
15	0,691	1,341	1,753	2,131	2,602	2,947	3,733	4,073
16	0,690	1,337	1,746	2,120	2,583	2,921	3,686	4,015
17	0,689	1,333	1,740	2,110	2,567	2,898	3,646	3,965
18	0,688	1,330	1,734	2,101	2,552	2,878	3,610	3,922
19	0,688	1,328	1,729	2,093	2,539	2,861	3,579	3,883
20	0,687	1,325	1,725	2,086	2,528	2,845	3,552	3,850
21	0,686	1,323	1,721	2,080	2,518	2,831	3,527	3,819
22	0,686	1,321	1,717	2,074	2,508	2,819	3,505	3,792
23	0,685	1,319	1,714	2,069	2,500	2,807	3,485	3,768
24	0,685	1,318	1,711	2,064	2,492	2,797	3,467	3,745
25	0,684	1,316	1,708	2,060	2,485	2,787	3,450	3,725
26	0,684	1,315	1,706	2,056	2,479	2,779	3,435	3,707
27	0,684	1,314	1,703	2,052	2,473	2,771	3,421	3,690
28	0,683	1,313	1,701	2,048	2,467	2,763	3,408	3,674
29	0,683	1,311	1,699	2,045	2,462	2,756	3,396	3,659
30	0,683	1,310	1,697	2,042	2,457	2,750	3,385	3,646
35	0,682	1,306	1,690	2,030	2,438	2,724	3,340	3,591
40	0,681	1,303	1,684	2,021	2,423	2,704	3,307	3,551
50	0,679	1,299	1,676	2,009	2,403	2,678	3,261	3,496
60	0,679	1,296	1,671	2,000	2,390	2,660	3,232	3,460
70	0,678	1,294	1,667	1,994	2,381	2,648	3,211	3,435
80	0,678	1,292	1,664	1,990	2,374	2,639	3,195	3,416
100	0,677	1,290	1,660	1,984	2,364	2,626	3,174	3,390
120	0,677	1,289	1,658	1,980	2,358	2,617	3,160	3,373
$\infty$	0,674	1,282	1,645	1,960	2,326	2,576	3,090	3,291

### Tabel Chi kwadraat verdeling

Getabelleerd zijn de kritieke waarden  $c$  voor de staartkansen

$$P(\chi^2 \geq c) = \alpha$$



Aantal vrijheidsgrader	$\alpha$											
	0,995	0,990	0,975	0,95	0,90	0,75	0,25	0,10	0,05	0,025	0,010	0,005
1	0,00	0,00	0,00	0,00	0,02	0,10	1,32	2,71	3,84	5,02	6,63	7,88
2	0,01	0,02	0,05	0,10	0,21	0,58	2,77	4,61	5,99	7,38	9,21	10,60
3	0,07	0,11	0,22	0,35	0,58	1,21	4,11	6,25	7,81	9,35	11,34	12,84
4	0,21	0,30	0,48	0,71	1,06	1,92	5,39	7,78	9,49	11,14	13,28	14,86
5	0,41	0,55	0,83	1,15	1,61	2,67	6,63	9,24	11,07	12,83	15,09	16,75
6	0,68	0,87	1,24	1,64	2,20	3,45	7,84	10,64	12,59	14,45	16,81	18,55
7	0,99	1,24	1,69	2,17	2,83	4,25	9,04	12,02	14,07	16,01	18,48	20,28
8	1,34	1,65	2,18	2,73	3,49	5,07	10,22	13,36	15,51	17,53	20,09	21,95
9	1,73	2,09	2,70	3,33	4,17	5,90	11,39	14,68	16,92	19,02	21,67	23,59
10	2,16	2,56	3,25	3,94	4,87	6,74	12,55	15,99	18,31	20,48	23,21	25,19
11	2,60	3,05	3,82	4,57	5,58	7,58	13,70	17,28	19,68	21,92	24,72	26,76
12	3,07	3,57	4,40	5,23	6,30	8,44	14,85	18,55	21,03	23,34	26,22	28,30
13	3,57	4,11	5,01	5,89	7,04	9,30	15,98	19,81	22,36	24,74	27,69	29,82
14	4,07	4,66	5,63	6,57	7,79	10,17	17,12	21,06	23,68	26,12	29,14	31,32
15	4,60	5,23	6,26	7,26	8,55	11,04	18,25	22,31	25,00	27,49	30,58	32,80
16	5,14	5,81	6,91	7,96	9,31	11,91	19,37	23,54	26,30	28,85	32,00	34,27
17	5,70	6,41	7,56	8,67	10,09	12,79	20,49	24,77	27,59	30,19	33,41	35,72
18	6,26	7,01	8,23	9,39	10,86	13,68	21,60	25,99	28,87	31,53	34,81	37,16
19	6,84	7,63	8,91	10,12	11,65	14,56	22,72	27,20	30,14	32,85	36,19	38,58
20	7,43	8,26	9,59	10,85	12,44	15,45	23,83	28,41	31,41	34,17	37,57	40,00
21	8,03	8,90	10,28	11,59	13,24	16,34	24,93	29,62	32,67	35,48	38,93	41,40
22	8,64	9,54	10,98	12,34	14,04	17,24	26,04	30,81	33,92	36,78	40,29	42,80
23	9,26	10,20	11,69	13,09	14,85	18,14	27,14	32,01	35,17	38,08	41,64	44,18
24	9,89	10,86	12,40	13,85	15,66	19,04	28,24	33,20	36,42	39,36	42,98	45,56
25	10,52	11,52	13,12	14,61	16,47	19,94	29,34	34,38	37,65	40,65	44,31	46,93
26	11,16	12,20	13,84	15,38	17,29	20,84	30,43	35,56	38,89	41,92	45,64	48,29
27	11,81	12,88	14,57	16,15	18,11	21,75	31,53	36,74	40,11	43,19	46,96	49,64
28	12,46	13,56	15,31	16,93	18,94	22,66	32,62	37,92	41,34	44,46	48,28	50,99
29	13,12	14,26	16,05	17,71	19,77	23,57	33,71	39,09	42,56	45,72	49,59	52,34
30	13,79	14,95	16,79	18,49	20,60	24,48	34,80	40,26	43,77	46,98	50,89	53,67
35	17,19	18,51	20,57	22,47	24,80	29,05	40,22	46,06	49,80	53,20	57,34	60,27
40	20,71	22,16	24,43	26,51	29,05	33,66	45,62	51,81	55,76	59,34	63,69	66,77
50	27,99	29,71	32,36	34,76	37,69	42,94	56,33	63,17	67,50	71,42	76,15	79,49
60	35,53	37,48	40,48	43,19	46,46	52,29	66,98	74,40	79,08	83,30	88,38	91,95
70	43,28	45,44	48,76	51,74	55,33	61,70	77,58	85,53	90,53	95,02	100,43	104,21
80	51,17	53,54	57,15	60,39	64,28	71,14	88,13	96,58	101,88	106,63	112,33	116,32
90	59,20	61,75	65,65	69,13	73,29	80,62	98,65	107,57	113,15	118,14	124,12	128,30
100	67,33	70,06	74,22	77,93	82,36	90,13	109,14	118,50	124,34	129,56	135,81	140,17