Schriftliche Prüfung
(2 Stunden)

Bemerkungen:

- Erlaubte Hilfsmittel: 10 hand- oder maschinengeschriebene A4 Seiten (=5 Blätter). Taschenrechner ohne Kommunikationsmöglichkeit.
- Mobiltelefone sind auszuschalten!
- Lesen Sie zuerst alle Aufgaben durch! Verweilen Sie nicht zu lange bei einem Aufgabenteil, der Ihnen grosse Schwierigkeiten bereitet! Es wird nicht erwartet, dass Sie in der vorgegebenen Zeit alle Aufgaben vollständig lösen.
- Die nötigen Tabellen befinden sich auf den hintersten Seiten dieser Prüfung.
- Der Lösungsweg muss klar ersichtlich sein (Ausnahme: Multiple-Choice Aufgaben). Wenn Ihnen für einen Aufgabenteil ein Zwischenresultat fehlt, treffen Sie eine sinnvolle Annahme und markieren Sie diese deutlich.
- Aufgaben 1 bis 4 sind Textaufgaben. Schreiben Sie Ihre Antworten zu diesen Aufgaben auf die dafür vorgesehenen Antwortblätter und nicht auf das Aufgabenblatt.
- Aufgabe 5 ist eine Multiple-Choice Aufgabe. Jede Teilaufgabe besteht aus mehreren Aussagen. Pro Teilaufgabe können keine, eine oder mehrere Aussagen richtig sein. Die Antworten müssen auf dem separaten Multiple Choice Antwortblatt (blau) am Ende des Antwortdossiers angekreuzt werden und dürfen *nicht* auf der Prüfung angekreuzt werden. Für jede Aussage gibt es 1/2 Punkt, wenn sie korrekt markiert wird. 1/2 Punkt wird abgezogen, wenn eine Aussage falsch markiert wird. Wenn eine Aussage nicht markiert wird, gibt es keinen Punkt, es wird aber auch kein Punkt abgezogen. Keine Teilaufgabe kann mit negativer Punktzahl abgeschlossen werden. Minimal erhält man also für eine ganze Teilaufgabe 0 Punkte.

Viel Erfolg!

1. (9 Punkte)
Ruben ist Koch in der Mensa Polyterrasse. Um den beträchtlichen Bedarf an Tomaten zu decken, kauft er bei zwei Lieferanten ein. Mit Wahrscheinlichkeit 0.6 kauft er für eine Lieferung bei Lieferant A ein, andernfalls bei Lieferant B. Leider sind nicht alle gelieferten Tomaten von ausreichender Qualität. Er weiss aus Erfahrung, dass eine Tomate aus einer Lieferung von Lieferant A mit Wahrscheinlichkeit 0.97 verwendbar ist. Wir definieren folgende Ereignisse:

$$\begin{aligned} A &:= \text{die Tomate ist von Lieferant } A \\ B &:= \text{die Tomate ist von Lieferant } B \\ V &:= \text{die Tomate ist verwendbar} \\ V^c &:= \text{die Tomate ist nicht verwendbar.} \end{aligned}$$

a) (1 Punkt) Wie hoch ist die Wahrscheinlichkeit, dass eine gelieferte Tomate von Lieferant A stammt und verwendbar ist?

b) (1 Punkt) Ruben nimmt eine Tomate von Lieferant A entgegen. Wie hoch ist die Wahrscheinlichkeit, dass diese Tomate nicht verwendbar ist? (Ersatzresultat: 0.04)

c) (2 Punkte) Verwenden Sie $\mathbb{P}[V^c|B] = 0.05$. Was ist die Wahrscheinlichkeit, dass eine zufällig ausgewählte Tomate nicht verwendbar ist? In anderen Worten: Berechnen Sie $\mathbb{P}[V^c]$. (Ersatzresultat: 0.06)

d) (1 Punkt) Sind V^c und A unabhängig? Begründen Sie Ihre Antwort.

e) (2 Punkte) Wie hoch ist die Wahrscheinlichkeit, dass eine zufällig ausgewählte nicht verwendbare Tomate von Lieferant A stammt?

f) (2 Punkte) Wieviele Tomaten muss Ruben mindestens bestellen um im Erwartungswert mindestens 1000 verwendbare Tomaten zu erhalten? Nehmen Sie auch hier an, dass er mit Wahrscheinlichkeit 0.6 bei Lieferant A einkauft, ansonsten bei Lieferant B.

2. (8 Punkte)
Wenn Christina für Prüfungen lernen muss, ist sie nach dem Aufwachen um punkt 7.00 Uhr immer mindestens 2 Stunden wach, wird danach aber schnell wieder müde und macht täglich einen "power nap". Die Zeit (in Stunden) bis zum Beginn des "power naps" sei durch die Zufallsvariable T mit kumulativer Verteilungsfunktion

$$F_T(t) = \begin{cases} 1 - \left(\frac{2}{t}\right)^2 & \text{für } t \geq 2 \\ 0 & \text{für } t < 2. \end{cases}$$

beschrieben.

a) (1 Punkt) Berechnen Sie die Dichte der Zufallsvariable T.

b) (1 Punkt) Wie gross ist die Wahrscheinlichkeit, dass Christina den „power nap" vor 10.00 Uhr beginnt (d.h. $T \leq 3$)?

c) (1 Punkt) Wie gross ist die Wahrscheinlichkeit, dass Christina den „power nap" zwischen 12.00 und 13.00 Uhr beginnt?

d) (2 Punkte) Berechnen Sie den Median m von T.

Christinas Mitbewohnerin Claudia arbeitet in der Gastronomie und kommt abends oft erst spät nach Hause. Daher schläft sie morgens immer länger als Christina. Die Zeit (in Stunden) zwischen Christinas Aufwachen um 7.00 Uhr bis zu Claudias Aufwachen sei durch die Zufallsvariable D mit kumulativer Verteilungsfunktion

$$F_D(t) = \begin{cases} 1 - \exp\left(-\frac{t}{4}\right) & \text{für } t \geq 0 \\ 0 & \text{für } t < 0. \end{cases}$$

beschrieben. Wir nehmen an, dass die Zufallsvariablen T und D unabhängig sind.

e) (1 Punkt) Wie gross ist die Wahrscheinlichkeit, dass an einem Tag Christina den "power nap" nach 12 Uhr beginnt und Claudia vor 12 Uhr aufwacht?

f) (2 Punkte) Wenn Christina einen "power nap" von 30 Minuten macht, wie gross ist dann die Wahrscheinlichkeit, dass um 13.30 Uhr mindestens eine der beiden schläft?

3. (**8 Punkte**)
Jan hat einen Zufallszahlengenerator programmiert. Er sagt, dass dieser (i.i.d.) uniform-verteilte Zufallszahlen generiert auf dem Intervall $[0, c]$. Die obere Schranke c verrät er Ihnen aber nicht. Alles, was wir haben, ist folgende (i.i.d.) Stichprobe:

x_1	x_2	x_3	x_4	x_5
1.3	2.4	5.0	8.5	11.9

a) (3 Punkte) Ermitteln Sie den Momentenschätzer für c. Ist dieser für die konkret vorliegende Stichprobe sinnvoll?

b) (2 Punkte) Bestimmen Sie den Maximum-Likelihood-Schätzer für c (für die konkret vorliegende Stichprobe). *Hinweis:* Überlegen Sie sich, welche Werte von c überhaupt sinnvoll sind. Nichts ableiten.

Die folgenden Teilaufgaben sind unabängig untereinander und unabhängig von von a) und b).

c) (1 Punkt) Die Zufallsvariable X habe die Dichte

$$f(x) = \frac{1}{\lambda\pi}\left(\frac{1}{1+\frac{x^2}{\lambda^2}}\right),\ x \in \mathbb{R}$$

wobei $\lambda > 0$ ein Parameter ist.
Bestimmen Sie die Likelihoodfunktion basierend auf n unabhängigen, identisch verteilten Beobachtungen $x_1, \ldots, x_n$ einer Zufallsvariable mit obiger Dichte.

d) (2 Punkte) Sie schätzen den Modellparameter θ mit dem Schätzer $\widehat{\theta}$. Von der Theorie her weiss man, dass

$$\widehat{\theta} \sim N(\theta, \theta/n),$$

wobei n die Stichprobengrösse ist. Geben Sie damit ein (approximatives) 95%-Vertrauensintervall für θ an, falls der realisierte Wert $\widehat{\theta} = 19.2$ und $n = 16$ ist.

4. (8 Punkte)

Emma und Tamara sind ambitionierte 1000m-Läuferinnen. Emma hat sich im letzten Winter eine Knieverletzung zugezogen und möchte nun wissen, ob sie zu ihrer alten Form zurückgefunden hat. Vor ihrer Verletzung war sie über 1000m genauso schnell wie Tamara (d.h. nicht signifikant schneller oder langsamer).

Nach der Verletzungspause haben Emma und Tamara in 10 Trainings ihre Laufzeiten notiert. Zwischen den verschiendenen Trainings herrschten unterschiedliche Laufbedingungen und während eines Trainings liefen Emma und Tamara jeweils gleichzeitig, also bei gleichen Laufbedingungen. Sie möchten nun anhand eines statistischen Tests herausfinden, ob eine der beiden signifikant schneller oder langsamer als die andere war. Die beobachteten Trainingszeiten über 1000m in Sekunden sind wie folgt:

Training Nr.	1	2	3	4	5	6	7	8	9	10
Tamara (x_i)	248.6	234.1	243.3	238.2	237	240.3	248.5	234.5	238.6	251
Ema (y_i)	241.3	236.9	244.3	248.6	240.1	241.8	233.5	243.7	240.2	234.8
Differenz (u_i)	7.3	-2.7	-1.1	-10.5	-3.1	-1.6	15.1	-9.2	-1.7	16.3

Die empirischen Mittwerte und Standardabweichungen sind in folgender Tabelle gegeben:

Kennzahlen	empirischer Mittelwert	empirische Standardabweichung
Tamara (x_i)	241.41	6.14
Ema (y_i)	240.53	4.59
Differenz ($u_i = x_i - y_i$)	0.88	9.16

a) (2 Punkte) Führen Sie einen gepaarten zweiseitigen t-Test auf dem 5%-Niveau durch, um herauszufinden, ob eine der beiden signifikant schneller oder langsamer als die andere war. Nennen Sie explizit Null- und Alternativhypothese, sowie Teststatistik, Verteilung unter H_0, Verwerfungsbereich und den Testentscheid.

b) (1 Punkt) Welche Verteilungsannahme sollte in a) gewährleistet sein, damit der Test adäquat ist (bzw. auch wirklich das 5%-Niveau hat)?

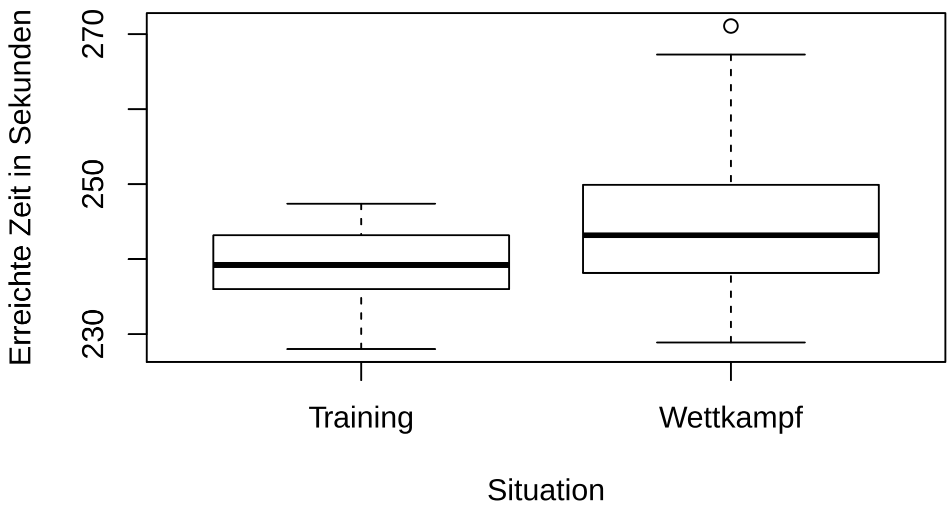
c) (1 Punkt) Berechnen Sie unter den Annahmen von Teilaufgabe a) ein zweiseitiges 99%-Vertrauensintervall für den erwarteten Zeitunterschied.

Emma ist insbesondere mit ihrer Leistung an Wettkämpfen nicht zufrieden und möchte nun wissen, ob sie an Wettkämpfen systematisch langsamer gelaufen ist als im Training. Wir nehmen an, dass die Streuung in den Laufzeiten in beiden Situationen (Training vs. Wettkampf) gleich gross ist. Emma liegen folgende Laufzeiten (in Sekunden) von den letzten 8 Wettkämpfen vor (der empirische Mittelwert $\bar{w}_8$ und die empirische Standardabweichung s_8 sind ebenfalls gegeben):

Wettkampf Nr.	1	2	3	4	5	6	7	8	$\bar{w}_8$	s_8
Zeiten (w_i)	247.6	238	255.1	239.1	246	239.2	244.8	256.8	245.8	7.19

d) (3 Punkte) Führen Sie einen geeigneten statistischen Test auf dem 5%-Niveau durch, um zu testen, ob Emma an Wettkämpfen systematisch langsamer war als im Training. Nennen Sie explizit Null- und Alternativhypothese, sowie Teststatistik, Verteilung unter H_0, Verwerfungsbereich und den Testentscheid.

e) (1 Punkt) Nehmen Sie nun an, dass der Boxplot der Trainings- resp. Wettkampfszeiten wie folgt aussieht:



Welchen Einfluss hat dieses Bild auf die Wahl des statistischen Tests, den wir zum Testen unserer Hypothese verwenden?

5. (**10 Punkte**) Entscheiden Sie für jede Aussage, ob sie richtig oder falsch ist. Markieren Sie Ihre Antwort auf dem separaten Multiple Choice Antwortblatt.

a) (2 Punkte) Welche Aussagen sind für zwei beliebige disjunkte Mengen A und B korrekt?

1. $(A^c \cup B)^c = A \cap B^c$
2. $A = (A \cap B) \cup (A \cap B^c)$
3. $\mathbb{P}(B|A) = 0.$
4. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ nur wenn A und B auch unabhängig sind.

b) (2 Punkte) Sei X eine Zufallsvariable mit $\mathbb{E}(X) = 2$ und $\text{Var}(X) = 1$ und sei Y eine Zufallsvariable mit $\mathbb{E}(Y) = 0$ und $\text{Var}(Y) = 9$.

1. Es gilt: $\text{Cov}(X, Y) = \mathbb{E}[XY]$.
2. Wenn $Z = -0.5X + 5$, dann gilt $\text{Var}(Z) = -0.25$.
3. $\text{Var}(X + Y) =$Var(X) + $\text{Var}(Y)$ gilt nur, wenn X und Y unabhängig sind.
4. $\mathbb{E}(a + bX + cY) = a + b\,\mathbb{E}(X) + c\,\mathbb{E}(Y)$ gilt nur, wenn X und Y unabhängig sind.

c) (2 Punkte) Beurteilen Sie folgende Aussagen zu statistischen Tests.

1. Nehmen Sie an, wir haben ein statistisches Testproblem der Form $H_0 : \mu = 0$ und $H_A : \mu \neq 0$. Die vorhanden Daten erfüllen die notwendigen Modellannahmen. Nun berechnen wir ein 95%-Vertrauensintervall und stellen fest, dass der Wert 0 nicht in diesem enthalten ist. Folglich können wir H_0 auf dem 5%-Niveau verwerfen.
2. Wenn ein statistischer Test signifikant ist, dann impliziert dies immer auch fachliche Relevanz.
3. Betrachten Sie einen zweiseitigen t-Test mit Nullhypothese $H_0 : \mu = 0$. Die Macht für die Alternative $\mu = -15$ ist dann grösser als für die Alternative $\mu = 25$.
4. Verwift man eine Nullhypothese bei einem einseitigen t-Test mit dem Signifikanzniveau α, dann würde man immer auch die Nullhypothese des entsprechenden zweiseitigen t-Test (gleiches Signifikanzniveau) verwerfen.

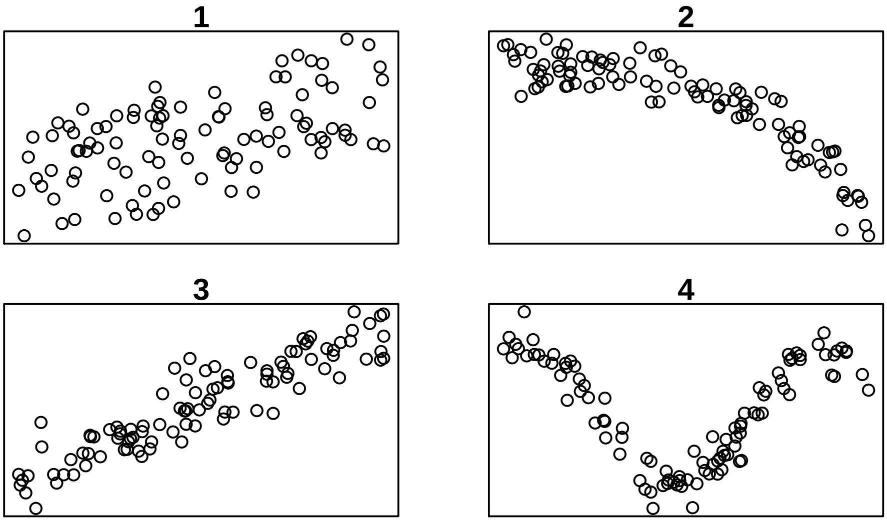
d) (2 Punkte) Nehmen Sie an, dass die Anzahl von Verkehrsunfällen in Zürich pro Tag einem homogenen Poissonprozess folgt. Die durchschnittliche Anzahl von Verkehrsunfällen in Zürich an einem Tag ist 2.

1. Die Wahrscheinlichkeit, dass mehr als zwei Tage zwischen zwei Unfällen vergehen ist 5.15%.
2. T_1 beschreibt den Zeitpunkt des ersten Unfalls. Dann gilt $P(T_1 > t) = \exp(-2t)$.
3. Die Zeiten zwischen zwei aufeinanderfolgenden Unfällen ist Poisson-verteilt.
4. Nehmen Sie an, Sie beobachten jetzt einen Unfall in Zürich. Dann vergehen im Durchschnitt 2 Tage bis der nächste Unfall geschieht.

e) (2 Punkte)
Betrachten Sie die folgenden Korrelationswerte

$$a = -0.04 \qquad b = 0.53 \qquad c = -0.87 \qquad d = 0.92$$

und die untenstehenden Plots. Welche der folgenden Aussagen ist korrekt?



1. Eine korrekte Zuordnung ist: 3b, 2c
2. Eine korrekte Zuordnung ist: 4a, 1b
3. Wenn zwei Zufallsvariablen X und Y unabhängig sind, dann gilt $\mathbb{E}[XY] = \mathbb{E}[X] + \mathbb{E}[Y]$.
4. Wenn eine Abhängigkeit zwischen zwei Zufallsvariablen X und Y besteht, dann gilt immer $Corr(X, Y) \neq 0$.

Tabelle der Standardnormalverteilung

$$\Phi(z) = \mathrm{P}\left[Z \leq z\right], \; Z \sim \mathcal{N}(0,1)$$

Bsp.: $\mathrm{P}\left[Z \leq 1.96\right] = 0.975$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Quantile der t-Verteilung

Bsp.: $t_{9,\,0.975} = 2.262$

df	$t_{0.60}$	$t_{0.70}$	$t_{0.80}$	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$	$t_{0.995}$
1	0.325	0.727	1.376	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	1.318	1.711	2.064	2.492	2.797
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
31	0.255	0.530	0.853	1.309	1.696	2.040	2.452	2.744
32	0.255	0.530	0.853	1.309	1.694	2.037	2.449	2.738
33	0.255	0.530	0.853	1.308	1.693	2.035	2.445	2.733
34	0.255	0.529	0.852	1.307	1.691	2.032	2.441	2.728
35	0.255	0.529	0.852	1.306	1.690	2.030	2.438	2.724
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
90	0.254	0.526	0.846	1.291	1.662	1.987	2.368	2.632
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