Bachelorprüfung: Mathematik IV – Statistik
(2 Stunden)

Bemerkungen:

- Es sind alle mitgebrachten, schriftlichen Hilfsmittel und der Taschenrechner erlaubt.
- Natels sind auszuschalten!
- Lesen Sie zuerst alle Aufgaben durch! Verweilen Sie nicht zu lange bei einem Aufgabenteil, der Ihnen grosse Schwierigkeiten bereitet! Für die Note 6 brauchen nicht alle Aufgaben gelöst zu sein!
- Wenn nicht anders vermerkt, sind die Tests auf dem 5%-Niveau durchzuführen.
- Die nötigen Tabellen befinden sich auf den hintersten Seiten dieser Prüfung.
- Aufgaben 4 und 5 sind Multiple-Choice Aufgaben. Es ist jeweils genau eine Antwort korrekt. Eine korrekte Antwort gibt 1 **Plus**punkt und eine falsche Antwort $\frac{1}{2}$ **Minus**punkt. Minimal erhält man für eine ganze Multiple-Choice Aufgabe 0 Punkte. Tragen Sie die korrekten Antworten der Multiple Choice Aufgaben mit Kreuzchen in das separate Antwortblatt ein.

Viel Erfolg!

1. (8 Punkte)

Der Sportschuh-Hersteller Hypatia AG, offizieller Ausrüster des syldawischen olympischen Teams, will das beste Material für die bevorstehenden olympischen Spiele zur Verfügung stellen. Dazu lässt der Hersteller die neuesten beiden Schuh-Kreationen, "SpeedShoe" und "Lightning", von zehn syldawischen Athleten auf einer 400m-Bahnrunde testen, wobei jeder Athlet zuerst mit dem einen (zufällig gewählten) Modell läuft und dann mit dem anderen. Die Kennzahlen der gemessenen Zeiten sind wie folgt:

SpeedShoe	$\bar{x}$	$=$	46.02	$\hat{\sigma}_x$	$=$	1.56
Lightning	$\bar{y}$	$=$	46.24	$\hat{\sigma}_y$	$=$	1.52
Differenz	$\bar{x}-\bar{y}$	$=$	-0.22	$\hat{\sigma}_{x-y}$	$=$	0.26

Sie dürfen davon ausgehen, dass die Rundenzeiten durch unabhängige $\mathcal{N}(\mu_x, \sigma_x^2)$- resp. $\mathcal{N}(\mu_y, \sigma_y^2)$-verteilte Zufallsvariablen beschrieben werden können.

Es soll nun getestet werden, ob das Modell "SpeedShoe" zu besseren Leistungen verhilft als das Modell "Lightning".

a) Handelt es sich hier um einen gepaarten oder einen ungepaarten Test? Begründen Sie kurz.

b) Geben Sie die Null- und die Alternativhypothese an.

c) Führen Sie einen t-Test auf dem 5%-Niveau durch. Geben Sie den Verwerfungsbereich und den Wert der Teststatistik an. Wie entscheidet dieser Test?

d) Geben Sie ein einseitiges 95%-Vertrauensintervall an für die Differenz $\mu_x - \mu_y$.

2. (8 Punkte)

Corinne, Markus, Lukas, Nicolas und Christoph treffen sich regelmässig zum Pokern. Markus behauptet, dass er besser bluffe als die restlichen Spieler und daher häufiger gewinne, d.h. dass er mit einer Wahrscheinlichkeit π gewinnt, die grösser als 0.2 ist. Es werden n Runden gepokert um herauszufinden, ob er Recht hat.

a) Mit welcher Wahrscheinlichkeitsverteilung würden Sie die Anzahl von Markus gewonnener Runden beschreiben?

b) Falls Markus sich irrt und alle Spieler gleich gut spielen, wieviele Siege von Markus werden erwartet? Berechnen Sie den Erwartungswert und die Varianz für $n = 20$ Runden.

Verwenden Sie für die folgenden Teilaufgaben c), d) und e) eine geeignete Approximation.

c) Es werden $n = 100$ Runden gepokert. In der folgenden Tabelle sind die Anzahl gewonnener Runden der einzelnen Spieler angegeben:

Markus	Corinne	Lukas	Nicolas	Christoph
28	21	16	18	17

Wenn alle Spieler gleich gut spielen ($\pi = 0.2$), wie gross ist die Wahrscheinlichkeit, dass Markus 28-mal oder häufiger gewinnt?

d) Testen Sie einseitig auf dem 1%-Niveau, ob "Markus gleich gut wie alle anderen Spieler" spielt. Begründen Sie Ihre Antwort.
Schreiben Sie formal Null- und Alternativhypothese, sowie die Teststatistik auf.
Wie oft muss Markus mindestens gewinnen, damit die Nullhypothese verworfen wird? Geben Sie den Verwerfungsbereich des Tests an.

e) Berechnen Sie mit den Angaben aus Aufgabe c) ein approximatives zweiseitiges 95%-Vertrauensintervall für die Gewinnwahrscheinlichkeit π von Markus.

3. (10 Punkte)

Hansi und Fritz sind Sportschützen. Hansi ist ein begnadeter Schütze und trifft sein Ziel mit 95%-iger Wahrscheinlichkeit. Fritz ist weniger treffsicher als Hansi, seine Trefferquote beträgt lediglich 60%.

a) Wie gross ist die Wahrscheinlichkeit, dass Fritz fünfmal hintereinander trifft?

b) Fritz feuert solange auf das Ziel, bis er trifft. Wie gross ist die Wahrscheinlichkeit, dass er genau n Schüsse benötigt? Wie gross ist die Wahrscheinlichkeit, dass er mehr als 3 Schüsse benötigt?
Hinweis: Fritz benötigt genau n Schüsse, wenn er zuerst $(n-1)$-mal verfehlt und beim n-ten Schuss schliesslich trifft.

c) Fritz und Hansi schiessen gleichzeitig auf dasselbe Ziel. Mit welcher Wahrscheinlichkeit wird das Ziel getroffen?
Hinweis: Betrachten Sie das Komplementär-Ereignis.

d) Die beiden machen ein Wettschiessen. Wer mehr Treffer landet gewinnt. Fritz darf dreimal schiessen, Hansi einmal. Mit welcher Wahrscheinlichkeit gewinnt Fritz?
Hinweis: Fritz gewinnt genau dann, wenn Hansi verfehlt und Fritz mindestens einmal trifft, oder wenn Hansi trifft und Fritz mindestens zweimal trifft.

e) Fritz und Hansi schiessen abwechselnd, bis einer von beiden das Ziel trifft. Fritz beginnt. Wie gross ist die Wahrscheinlichkeit, dass Fritz als erster trifft?
Warnung: Dieser Aufgabenteil ist ein bisschen "tricky". Sie sollten sich nur daran versuchen, falls Sie ausreichend Zeit haben.
Hinweis: Betrachten Sie die Ereignisse

$$A_j = \text{"Fritz trifft als erster bei seinem j-ten Schuss"}.$$

Verwenden Sie, dass für alle $|z| < 1$ gilt:

$$\sum_{n=0}^{\infty} z^n = \frac{1}{1-z}.$$

4. (10 Punkte)

Wir betrachten hier Daten, welche in verschiedenen amerikanischen Grossstädten erhoben wurden. In den einzelnen Städten wurde die Arbeitslosigkeit in Prozent (x_1) und die Selbstmordrate pro Million Einwohner (y) gemessen.

An diese Daten wurde folgendes Regressionsmodell angepasst:

$$y_i = \beta_0 + \beta_1 x_{1,i} + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}\left(0, \sigma^2\right) \quad \text{i.i.d.}$$

Hier sehen Sie einen Teil des R-Outputs:

```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)             ??      51.25  -1.831  0.10027
Arbeitslosigkeit     59.05      14.24      ??  0.00249

Residual standard error: 36.06 on 9 degrees of freedom
Multiple R-squared: 0.6566,Adjusted R-squared: 0.6185
F-statistic: 17.21 on 1 and 9 DF,  p-value: 0.00249
```

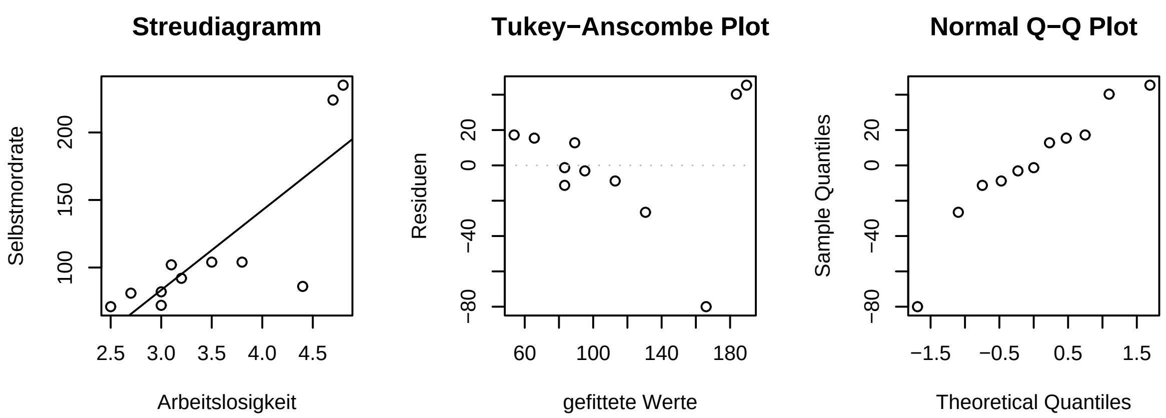
1) Mit wievielen Beobachtungen wurde die Regression berechnet?
   a) 9 b) 12 c) 11 d) 10

2) Wie gross ist der Intercept $\hat{\beta}_0$?
   a) 0 b) -93.8 c) 102.5 d) -75.2

3) Wird $H_0 : \beta_1 = 0$ auf dem 5%-Niveau verworfen (die Alternative ist $H_A : \beta_1 \neq 0$)?
   a) Ja b) Nein c) Keine Aussage möglich

4) Wie gross ist die t-Teststatistik für den Test der Nullhypothese $H_0 : \beta_1 = 0$ (die Alternativhypothese ist $H_A : \beta_1 \neq 0$)?
   a) 1.54 b) 23.21 c) 50.54 d) 4.15

5) Welches der folgenden Intervalle ist ein exaktes zweiseitiges 95%-Vertrauensintervall für β_1?
   a) $59.05 \pm 1.96 \cdot \frac{14.24}{\sqrt{9}}$ b) $59.05 \pm 2.262 \cdot \frac{14.24}{\sqrt{9}}$
   c) $59.05 \pm 1.96 \cdot 14.24$ d) $59.05 \pm 2.262 \cdot 14.24$



6) Betrachten Sie die gezeigten Plots. Welche der nachfolgenden Aussagen ist zutreffend?

a) Die Modellannahmen über die Fehler scheinen plausibel.
b) Die Normalitätsannahme der Fehler ϵ_i scheint grob verletzt zu sein.
c) Die Annahme der konstanten Varianz der Fehler ϵ_i scheint verletzt zu sein.
d) Die Normalitätsannahme der Fehler ϵ_i und die Annahme der konstanten Varianz der Fehler ϵ_i scheinen grob verletzt zu sein.

7) Was passiert mit der Regression, wenn an der Stelle $(2.5, 150)$ noch eine Beobachtung hinzugefügt wird?

a) $\hat{\beta}_1$ wird kleiner, $\hat{\sigma}$ wird grösser.
b) $\hat{\beta}_1$ wird kleiner, $\hat{\sigma}$ wird kleiner.
c) $\hat{\beta}_1$ wird grösser, $\hat{\sigma}$ wird grösser.
d) $\hat{\beta}_1$ wird grösser, $\hat{\sigma}$ wird kleiner.
e) Keine Aussage möglich.

8) Wie gross ist die empirische Korrelation $\hat{\rho}$ zwischen der Arbeitslosigkeit und der Selbstmordrate?

a) $-1 \leq \hat{\rho} < -0.8$
b) $-0.8 \leq \hat{\rho} < -0.5$
c) $-0.5 \leq \hat{\rho} < 0.3$
d) $0.3 \leq \hat{\rho} < 0.6$
e) $0.6 \leq \hat{\rho} < 1$
f) $\hat{\rho} = 1$

Es wird nun eine weitere erklärende Variable hinzugenommen. Für die einzelnen Städte wurde jetzt noch zusätzlich die Kriminalität in Prozent (x_2) gemessen.

Man hat also folgendes Regressionsmodell:

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}\left(0, \sigma^2\right) \quad \text{i.i.d.}$$

Hier der R-Output:

```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)        446.61     277.44   1.610   0.1461
Arbeitslosigkeit  -245.94     155.03  -1.586   0.1513
Kriminalitaet       41.09      20.82   1.974   0.0839

Residual standard error: 31.36 on 8 degrees of freedom
Multiple R-squared: 0.7691,Adjusted R-squared: 0.7113
F-statistic: 13.32 on 2 and 8 DF,  p-value: 0.002844
```

9) Wird die Zielvariable mindestens von einer der beiden erklärenden Variablen (x_1,x_2) signifikant beeinflusst?

a) Ja.
b) Nein.
c) Das kann man nur anhand von zwei einfachen Regressionen entscheiden.
d) Keine Aussage möglich.

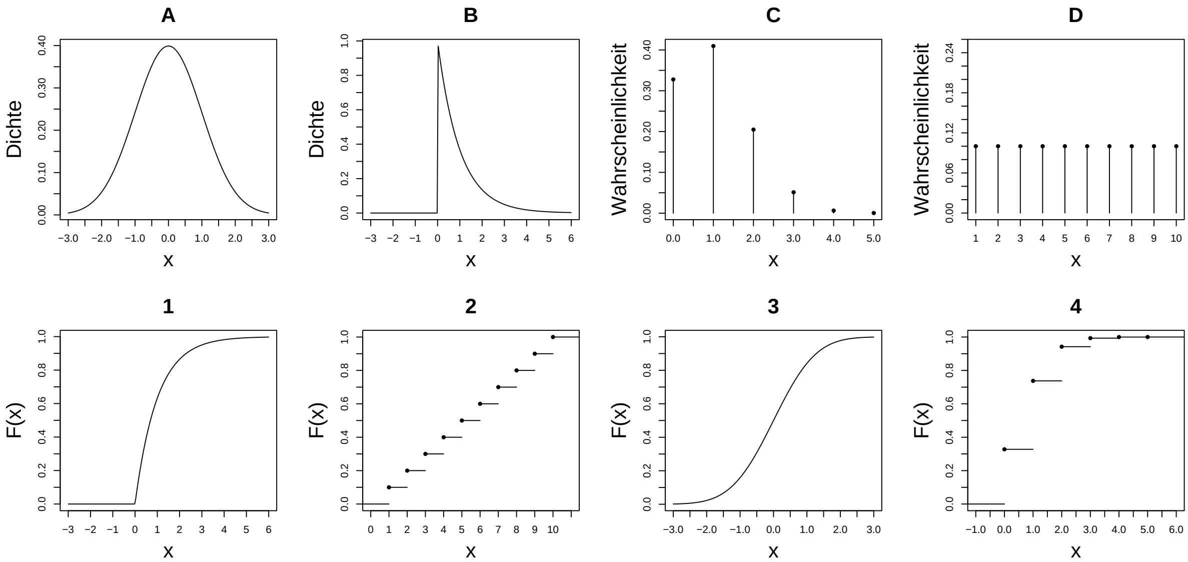
10) Kann man im jetzigen Modell das Bestimmtheitsmass R^2 immer noch als quadratische Stichproben-Korrelation zwischen erklärenden Variablen und Zielvariable interpretieren?

a) Ja
b) Nein
c) Keine Aussage möglich

5. (7 Punkte)

1) Finden Sie im untenstehenden Plot die richtige Zuordnung zwischen Dichte bzw. Wahrscheinlichkeiten und Verteilungsfunktion.

   a) A1, B3, C2, D4   b) A3, B1, C4, D2
   c) A3, B4, C2, D1   d) A4, B3, C2, D2
   e) A1, B3, C2, D4   f) A1, B4, C3, D2



2) X sei eine Zufallsvariable mit $\mathbf{E}[X] = 2$ und $\mathrm{Var}(X) = 1$. Wir definieren nun $Y = -2 + \frac{1}{4}X$. Wie gross ist $\mathbf{E}[X+Y]$?

   a) 0.125   b) 0   c) 0.25
   d) 1   e) 2   f) Aussagen a) bis e) sind falsch

3) Wir betrachten nochmals die Zufallsvariable $X+Y$ aus Teilaufgabe 2). Wie gross ist $\mathrm{Var}(X+Y)$?

   a) $\frac{17}{16}$ b) $\frac{25}{16}$ c) 1
   d) $\frac{1}{16}$ e) $\frac{15}{16}$ f) Aussagen a) bis e) sind falsch

4) Ein Langstreckenflugzeug der Swiss hat 228 Plätze. Welche Verteilung eignet sich zur Modellierung des Totalgewichts aller Passagiere im vollbesetzten Flugzeug?

   a) Exponentialverteilung   b) Poissonverteilung
   c) Normalverteilung   d) Binomialverteilung

5) Bei einem zweiseitigen t-Test ($H_0 : \mu = \mu_0$, $H_A : \mu \neq \mu_0$) ist der p-Wert 0.02. Zudem gilt für den Wert der Teststatistik t, dass $t > 0$. Wie gross wäre der p-Wert für den einseitigen Test mit der Alternative $H_A : \mu > \mu_0$?

   a) 0.01   b) 0.05
   c) 0.02   d) Keine Aussage möglich

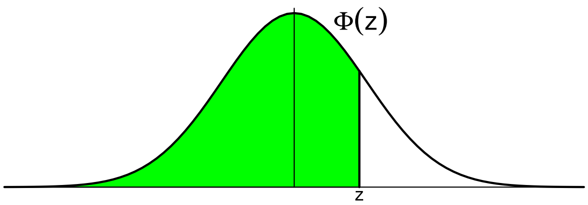
6) X sei eine Zufallsvariable mit $\mathcal{N}(0,2)$ Verteilung. Dann gilt

   a) $\mathrm{P}[-1 \leq X \leq 1] = 0.6826$ b) $\mathrm{P}\left[-\sqrt{2} \leq X \leq \sqrt{2}\right] = 0.6826$
   c) $\mathrm{P}[X \leq 1.96] = 0.975$ d) $\mathrm{P}\left[X \leq 1.96\sqrt{2}\right] = 0.95$

7) Das zu einem einseitigen z-Test gehörende Vertrauensintervall für μ sei $[2.48, \infty)$. Welche Form hatte die Alternativhypothese?

   a) $H_A : \mu < \mu_0$ b) $H_A : \mu > \mu_0$
   c) $H_A : \mu \neq \mu_0$ d) Keine Aussage möglich

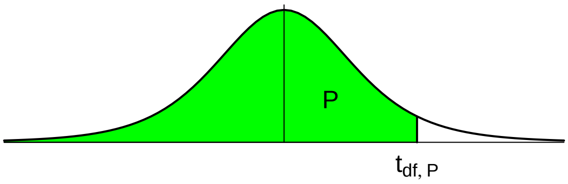
Tabelle der Kumulativen Normalverteilung $\Phi(z) = \mathrm{P}\left[Z \leq z\right], \ Z \sim \mathcal{N}(0,1)$



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

Perzentile 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
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