

2. Vordiplom: Statistik (2 Stunden)

Bemerkungen:

- Lies zuerst alle Aufgaben durch! Verweile nicht zu lange bei einem Aufgabenteil, der Dir grosse Schwierigkeiten bereitet! Für die Note 6 brauchen nicht alle Aufgaben gelöst zu sein!
- Die Benützung des Taschenrechners ist **nicht** erlaubt. Die Berechnungen sind jeweils auf 10% genau verlangt.
- Wenn nicht anders vermerkt, sind die Tests auf dem 5%-Niveau durchzuführen.
- Die nötigen Tabellen befinden sich auf den hintersten Seiten dieser Prüfung.
- Aufgaben 2, 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. Trage die korrekten Antworten der Multiple Choice Aufgaben mit Kreuzchen in das separate Antwortblatt ein.

Viel Erfolg!

1. (10 Punkte)

Corinne und Markus fragen sich, ob Corinne wohl übersinnliche Kräfte hat. Um das zu testen verlangt Markus, dass sie in einer Reihe von n Würfeln einer fairen Münze (unabhängig und gleich verteilt) die Abfolge von Kopf (0) oder Zahl (1) richtig vorher-sagt. Übersinnliche Fähigkeiten liegen dann vor, wenn die Wahrscheinlichkeit p einer korrekten Vorhersage grösser als 0.5 ist.

- a) Angenommen, Corinne's Vorhersagewahrscheinlichkeit beträgt $p = 0.5$, d.h., sie besitzt keine übernatürlichen Kräfte. Wie gross ist der Erwartungswert und die Varianz der richtig geratenen Versuche bei $n = 10$?
- b) (Verwende für die folgenden Teilaufgaben eine geeignete Approximation.) Wenn $n = 100$ Versuche durchgeführt werden, was ist die Wahrscheinlichkeit mehr als 55 Würfe korrekt zu erraten? (Es wird immer noch $p = 0.5$ angenommen.)
- c) Corinne hat 60 von 100 Würfeln korrekt vorhergesagt. Teste die einseitige Hypo- these "Corinne hat keine übernatürlichen Kräfte" auf dem 1% Niveau. Gib auch den Annahmehereich des Tests für die Anzahl der korrekten Vorhersagen an.
- d) In Wahrheit hat Corinne übersinnliche Kräfte und die wahre Vorhersagewahr- scheinlichkeit ist $p = 0.6$. Wie gross ist die Wahrscheinlichkeit, dass sie mit dem eben entworfenen Test (verwenden Sie zur Vereinfachung von Rechnungen als An- nahmehereich $[0, 60]$) als "übersinnlich" eingestuft wird, d.h., was ist die Macht des Tests?

2. (8 Punkte)

Die Gleitzahl eines Gleitschirms ist definiert durch die zurückgelegte horizontale Distanz bei einem Höhenverlust von einem Kilometer. Corinne hat sich einen neuen Gleitschirm gekauft. Sie möchte testen, ob ihr neuer Gleitschirm die bessere Gleitzahl hat, als derjenige von Markus. Zu diesem Zweck machen beide 10 Flüge und notieren sich die Gleitzahl:

Daten:

Corinne	9.3	8.3	10.5	12.0	10.7	10.9	8.5	10.7	9.1	11.7
Markus	9.8	7.6	8.9	10.2	7.3	9.9	8.5	8.7	8.4	11.6
Differenz	-0.5	0.7	1.6	1.8	3.4	1.0	0.0	2.0	0.7	0.1

Kennzahlen:

Corinne	\bar{x}	=	10.17	s_x	=	1.29
Markus	\bar{y}	=	9.09	s_y	=	1.30
Differenz	$\bar{x} - \bar{y}$	=	1.08	s_{x-y}	=	1.15

- 1) Nehmen Sie an die beiden machen ihre Flüge völlig unabhängig voneinander und vergleichen danach ihre Liste. Der angemessene t-Test ist ...
 - a) gepaart
 - b) ungepaart
- 2) Nehmen Sie an die beiden machen jeden der 10 Flüge gemeinsam und fliegen nebeneinander ins Tal. Zudem fliegen die beiden genau gleich gut. Der angemessene t-Test ist ...
 - a) gepaart
 - b) ungepaart
- 3) Was ist die korrekte Nullhypothese?
 - a) Die Gleitzahlen unterscheiden sich nicht.
 - b) Corinne's Gleitschirm hat eine grössere Gleitzahl.
 - c) Die Gleitzahlen unterscheiden sich.
- 4) Was ist die korrekte Alternativhypothese?
 - a) Die Gleitzahlen unterschieden sich nicht.
 - b) Die Gleitzahlen unterscheiden sich
 - c) Corinne's Gleitschirm hat eine grössere Gleitzahl.
- 5) Der Wert des gepaarten t-Tests beträgt ...
 - a) 2.97
 - b) 0.65
 - c) 12.18
- 6) Der Wert des ungepaarten t-Tests beträgt ...
 - a) 6.18
 - b) 1.86
 - c) 10.07
- 7) Welche Schlüsse könnten die beiden im Falle eines gepaarten t-Tests ziehen?
 - a) Die Gleitzahlen unterscheiden sich weder signifikant auf dem 5%-, noch auf dem 1%-Niveau.
 - b) Die Gleitzahlen unterscheiden sich signifikant auf dem 5%-Niveau, jedoch nicht auf dem 1%-Niveau.
 - c) signifikant auf dem 1%-Niveau, jedoch nicht auf dem 5%-Niveau.
 - d) Die Gleitzahlen unterscheiden sich sowohl auf dem 1%-Niveau, als auch auf dem 5%-Niveau signifikant.
- 8) Welche Schlüsse könnten die beiden im Falle eines ungepaarten t-Tests ziehen?
 - a) Die Gleitzahlen unterscheiden sich weder signifikant auf dem 5%-, noch auf dem 1%-Niveau.

- b) Die Gleitzahlen unterscheiden sich signifikant auf dem 5%-Niveau, jedoch nicht auf dem 1%-Niveau.
- c) signifikant auf dem 1%-Niveau, jedoch nicht auf dem 5%-Niveau.
- d) Die Gleitzahlen unterscheiden sich sowohl auf dem 1%-Niveau, als auch auf dem 5%-Niveau signifikant.

3. (12 Punkte)

Markus möchte ins Casino gehen und einige Runden Roulette spielen. Seine Strategie wird sein, den Einsatz von 3 Franken auf Rot und Weiss zu verteilen: 2 Franken setzt er auf Rot, 1 Franken auf Schwarz.

Die Zufallsvariable X bezeichnet den Gewinn (Zur Erklärung des Gewinns: Wenn 1 Franken auf Rot gesetzt wird und Rot gewinnt, ist der Gewinn 1 Franken; falls Rot verliert ist der Gewinn -1 Franken.) von Markus in einer einzelnen Runde. Setzt man auf die richtige Farbe, erhält man, wie im Roulette üblich, den Einsatz plus einen Bonus in der Höhe des Einsatzes. Setzt man falsch, ist der Einsatz verloren.

Es handelt sich hier um ein Roulette ohne die Null. Das heisst Sie dürfen annehmen, dass die Roulettekugel mit der gleichen Wahrscheinlichkeit $p = 1/2$ auf Schwarz, wie auf Rot zu liegen kommt.

- a) Berechnen Sie $\mathbf{E}[X]$
- b) Berechnen Sie $\sigma^2 := \text{Var}(X)$.
- c) Markus möchte mehrere Runden Roulette spielen. Bezeichnen wir mit X_i den Gewinn in der i . Runde. Wenn alles mit rechten Dingen zugeht, dürfen wir annehmen, dass die X_i *iid* sind.

Aus Erfahrung weiss Markus, dass man im Casino leicht den Überblick über seinen Gesamtgewinn verlieren kann. Da Markus aber den Verlust beschränkt halten möchte, überlegt er sich schon zu Hause, nach wievielen Runden er spätestens aufhören muss, um mit einer Wahrscheinlichkeit von $p = 0.95$ seinen Verlust auf weniger als 10 Franken zu beschränken.

Das heisst er möchte n =Anzahl Runden so wählen, dass folgendes gilt:

$$\mathbf{P} \left[\sum_{i=1}^n X_i \geq -10 \right] = 0.95.$$

Wie muss er n wählen?

Tipp: Benutzen Sie den Zentralen Grenzwertsatz. Weiter dürfen Sie das Resultat $10/1.64 = 6.1$ benutzen.

- d) Corinne hat eine andere Strategie. Sie setzt in jeder Runde den gesamten Einsatz von 3 Franken auf Schwarz. Berechnen Sie den Erwartungswert und die Varianz für den Gewinn von Corinne. Auch sie möchte, dass ihr Verlust mit einer Wahrscheinlichkeit von $p = 0.95$ auf weniger als 10 Franken beschränkt ist. Darf sie länger spielen als Markus? Kurze Begründung reicht.

4. (9 Punkte)

Mit einer gewissen Art der Fleischverarbeitung darf erst begonnen werden, wenn der pH-Wert in den Muskeln des toten Tiers von anfänglich etwa 7.0 auf 6.0 abgefallen ist. Da es aufwändig wäre, bei jeder Verarbeitung den pH-Wert zu überwachen, soll eine Zeit geschätzt werden, innerhalb derer der pH-Wert auf 6.0 abgefallen ist. Dazu wurden mehrere Tierkadaver für eine Messung des pH-Werts y_i zu verschiedenen Zeiten x_i zwischen null und zehn Stunden ausgewählt. Es wurde das Modell

$$y_i = \beta_0 + \beta_1 \log(x_i) + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2) \quad iid$$

angepasst.

Das Computer Programm R lieferte folgenden Ausdruck:

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	6.91318	0.12349	55.982	< 2e-16 ***
log(x)	-0.58968	0.09603	??	1.25e-06 ***

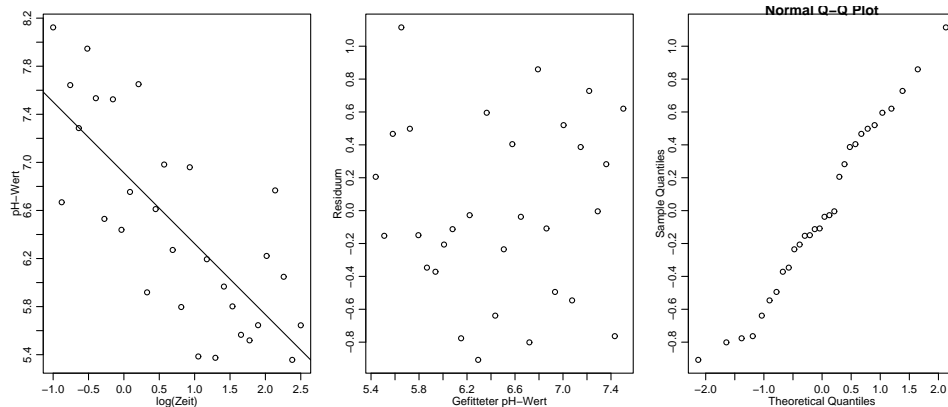
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5494 on 28 degrees of freedom

Multiple R-Squared: 0.5739, Adjusted R-squared: 0.5587

F-statistic: 37.71 on 1 and 28 DF, p-value: 1.254e-06

1. Handelt es sich bei obigem Modell um ein *lineares* Modell?
 - a) Ja, weil die erklärenden Variablen oder Transformationen davon linear eingehen.
 - b) Ja, weil die Koeffizienten linear eingehen.
 - c) Nein, aber es kann dennoch mit den Methoden der linearen Modelle behandelt werden.



2. Betrachte die gezeigten Plots. Welche der nachfolgenden Aussage ist zutreffend?
 - a) Die Normalitätsannahme der Fehler ϵ_i ist klar verletzt.
 - b) Die Annahme der konstanten Varianz der Fehler ϵ_i ist klar verletzt.
 - c) Die Modellannahmen über die Fehler scheinen plausibel.
3. Mit wievielen Beobachtungen wurde die Regression berechnet?
 - a) 25
 - b) 28
 - c) 29
 - d) 30
 - e) 32
 - f) Keine Aussage möglich
4. Wird $\beta_1 = 0$ auf dem 5% Niveau verworfen?
 - a) Ja
 - b) Nein
 - c) Keine Aussage möglich

5. Welches der folgenden Intervalle ist ein exaktes zweiseitiges 95% Vertrauensintervall für β_1 ?
- a) $-0.5897 \pm 1.96 \cdot 0.096$ b) $-0.5897 \pm 2.048 \cdot 0.096$
 c) $-0.5897 \pm 1.96 \cdot \frac{0.096}{\sqrt{28}}$ d) $-0.5897 \pm 2.048 \cdot \frac{0.096}{\sqrt{28}}$
 e) $6.913 \pm 2.048 \cdot 0.123$ f) Keine Aussage möglich
6. Wie gross ist die t-Teststatistik für $H_0 : \beta_1 = 0$?
- a) -6.14 b) 6.14 c) 1.49
 d) -1.49 e) 0 f) Keine Aussage möglich
7. Wie gross ist der mittlere pH-Wert nach 4 Stunden?
- a) $6.913 - 0.589 \cdot \log(4)$ b) $6.913 + 0.589 \cdot \log(4)$
 c) $6.913 - 0.589 \cdot 4$ d) $6.913 + 0.589 \cdot 4$
 e) $6.913 + 0.123 \cdot \log(4)$ f) $6.913 + 0.123 \cdot 4$
8. Wir suchen nun das 95% Vertrauensintervall für den mittleren pH-Wert nach 4 Stunden. Dazu wird das Modell $y_i = \beta_0 + \beta_1(\log(x_i) - \log(4)) + \epsilon_i$ angepasst. Die Computerausgabe liefert für den Achsenabschnitt $\beta_0 = 6.09$ mit Standardfehler $s_{\beta_0} = 0.117$. Wie gross ist also das gesuchte Intervall?
- a) $2.09 \pm 2.048 \cdot 0.117$ b) $6.09 \pm 2.048 \cdot 0.117$
 c) $2.09 \pm 1.96 \cdot 0.117$ d) $6.09 \pm 1.96 \cdot 0.117$
 e) Keine Aussage möglich
9. Zusätzlich zur Zeit wurde auch die Temperatur (Variable *temp*) des Körpers gemessen. Darauf basierend wird folgendes Modell angepasst

$$y_i = \beta_0 + \beta_1 \log(x_i) + temp_i + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2), \quad iid.$$

Folgende Ausgabe wird dabei generiert:

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	6.9198	0.1250	55.348	<2e-16 ***
logTime	-1.3374	1.0863	-1.231	0.229
temp	0.7419	1.0735	0.691	0.495

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

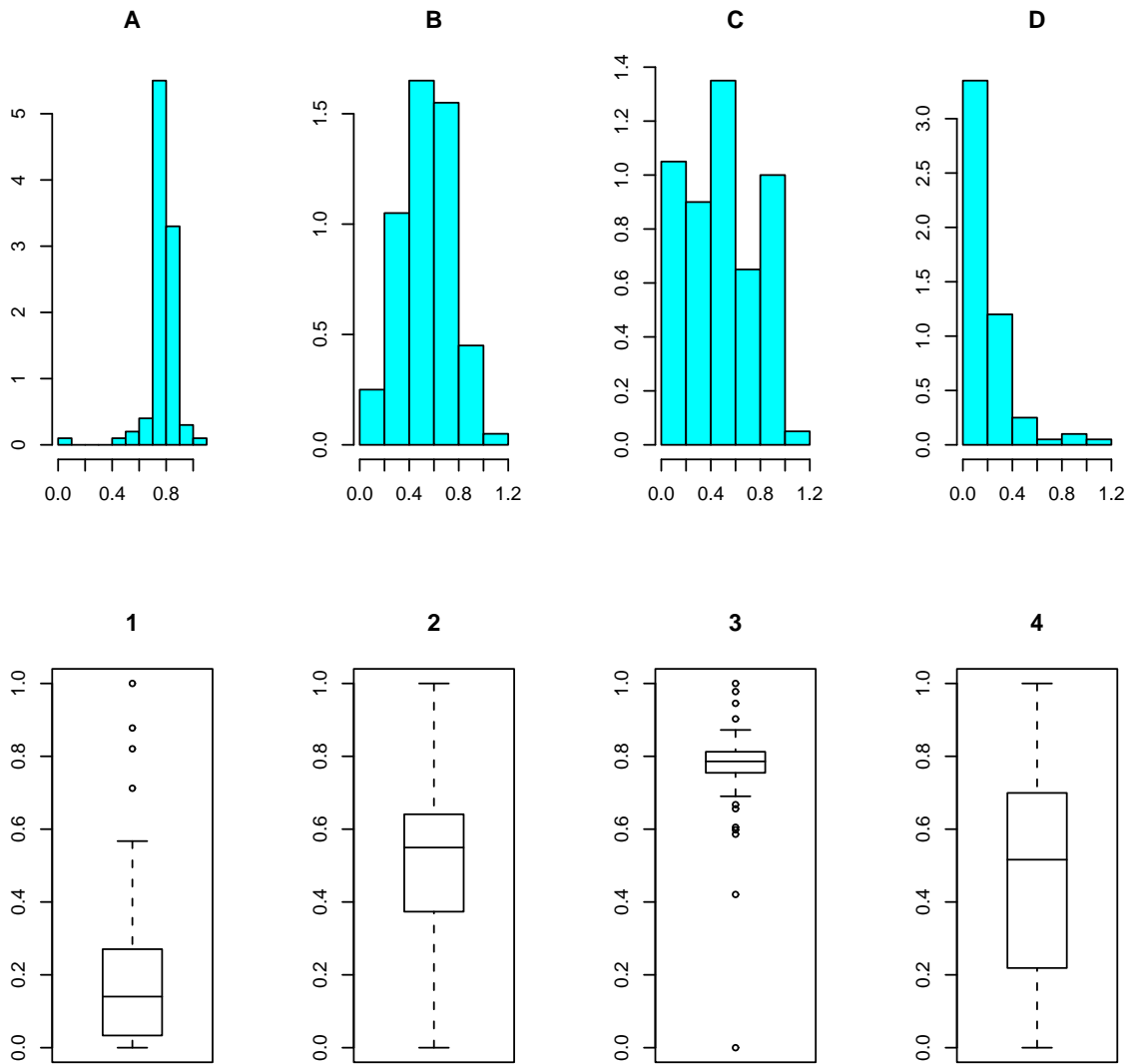
Residual standard error: 0.5546 on 27 degrees of freedom

Multiple R-Squared: 0.5813, Adjusted R-squared: 0.5503

F-statistic: 18.74 on 2 and 27 DF, p-value: 7.868e-06

Wird die Zielgrösse durch die Gesamtheit der erklärenden Variablen signifikant beeinflusst?

- a) Nein, weil weder der Koeffizient von *logTime* noch der von *temp* signifikant von Null verschieden ist.
 b) Ja, weil der F-Test signifikant ist.
 c) Ja, weil der *Intercept*-Term hochsignifikant von null verschieden ist.
 d) Darüber kann ohne weitere Rechnung keine Aussage gemacht werden.



5. (5 Punkte)

- 1) Wie Sie wissen, bekommen Sie für die multiple choice Aufgaben einen Punkt, falls Sie die richtige Antwort auswählen und einen halben Punkt Abzug, für die falsche. Nehmen Sie an, Sie hätten keine Ahnung, welches die richtige Antwort ist und müssten raten. Es stehen n mögliche Antworten zur Verfügung.

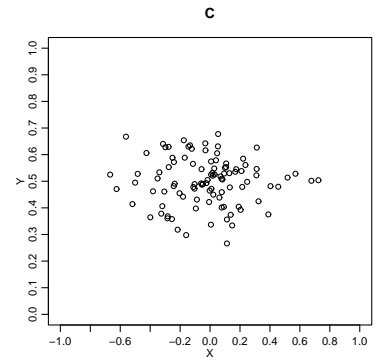
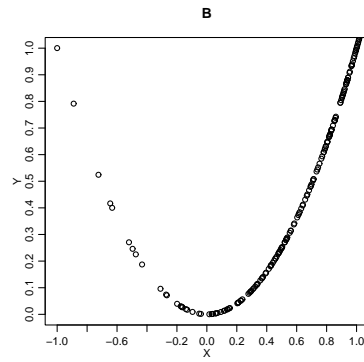
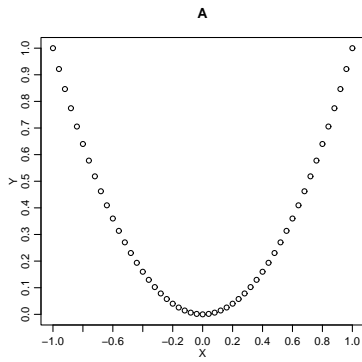
Welches ist das grösste n , sodass es sich noch zu raten lohnt?

- | | |
|----------|----------|
| a) $n=2$ | b) $n=3$ |
| c) $n=4$ | d) $n=5$ |
| e) $n=6$ | f) $n=7$ |

Tipp: Es lohnt sich zu raten, sobald der erwartete Punktergebnis strikt grösser als 0 ist.

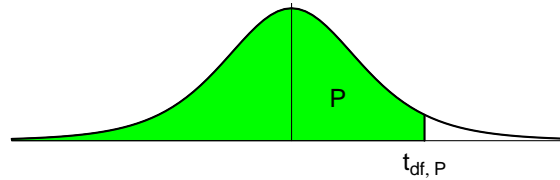
- 2) Betrachten Sie die Darstellungen auf dem oberen Teil dieser Seite. Welche der aufgeführten Zuordnungen von Histogramm und Boxplot ist richtig?
- | | |
|----------------|----------------|
| a) A2,B3,C1,D4 | b) A4,B2,C3,D1 |
| c) A1,B2,C3,D4 | d) A4,B3,C2,D1 |
| e) A3,B2,C4,D1 | f) A3,B4,C2,D1 |

Betrachten Sie nun die drei folgenden Streudiagramme (Anm: Im Folgenden liegt nur dann eine Abhängigkeit vor, wenn sie klar zu erkennen ist.):



- 3) Welche Aussage ist für den Plot A zutreffend?
- X und Y sind unkorreliert und unabhängig.
 - X und Y sind unkorreliert aber abhängig.
 - X und Y sind korreliert und abhängig.
- 4) Welche Aussage ist für den Plot B zutreffend?
- X und Y sind unkorreliert und unabhängig.
 - X und Y sind unkorreliert aber abhängig.
 - X und Y sind korreliert und abhängig.
- 5) Welche Aussage ist für den Plot C zutreffend?
- X und Y sind unkorreliert und unabhängig.
 - X und Y sind unkorreliert aber abhängig.
 - X und Y sind korreliert und abhängig.

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