

Bachelorprüfung: Mathematik 4 - Statistik (2 Stunden)

Bemerkungen:

- Es sind alle 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!
- 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. (10 Punkte)

In Bakerville gibt es zwei Bäckereien. Beide bieten die bei den Schülern sehr beliebten Pausenbrötchen an. Da die Schüler sehr preisbewusst sind, kann sich keine Bäckerei leisten, teurer als die andere zu sein. Schülervertreterin Corinne hat aber die Vermutung, dass die Bäckereien Geld sparen wollen, indem sie jeweils am Gewicht der Brötchen sparen. Sie hat jedoch noch keinen konkreten Verdacht, welche Bäckerei weniger fürs Geld liefert.

Zunächst will Corinne anhand einer Stichprobe herausfinden, ob überhaupt ein unterschiedliches Gewicht nachgewiesen werden kann, und wie gross dieser Unterschied ist.

Hierzu besorgt sie jeweils 10 Brötchen von jeder Bäckerei und notiert das Gewicht in Gramm. Dann berechnet sie davon folgende Kennzahlen.

$$\begin{array}{l|l} \text{Bäckerei } X & \bar{x} = 36.24 \quad \hat{\sigma}_x = 8.46 \\ \text{Bäckerei } Y & \bar{y} = 41.88 \quad \hat{\sigma}_y = 6.00 \end{array}$$

Als Modell nimmt Corinne an, dass die Gewichte der Brötchen einer Bäckerei sich jeweils gut mit einer Normalverteilung modellieren lassen.

- a) Begründen Sie kurz, wieso es sich hier um ungepaarte Stichproben handelt.
- b) Geben Sie die Null- und die Alternativhypothese an.
- c) Berechnen Sie S_{pool} .
Falls Sie diese Aufgabe nicht lösen können: Verwenden Sie für die folgenden Aufgaben das Ersatzresultat $S_{pool} = 7$.
- d) Führen Sie den geeigneten t-Test auf dem 5%-Niveau durch: Berechnen Sie den Wert der Teststatistik und den Verwerfungsbereich. Wie entscheidet der t-Test?
- e) Ist der p-Wert des Testes aus Aufgabe d) grösser oder kleiner als 5%? Begründen Sie kurz.

2. (10 Punkte)

Wegen niedrigeren Hygienestandards in vielen südlichen Ländern ist Durchfall die weitaus häufigste Reisekrankheit. Man geht davon aus, dass 30% aller Urlauber daran leiden.

a) Wie gross ist die Wahrscheinlichkeit, dass von 10 Urlaubern keiner erkrankt?

In einer Studie werden 300 Urlauber auf Durchfallerkrankung während der Reise befragt. Die Studie ergab, dass 100 der befragten Personen erkrankten.

b) Mit welcher Verteilung kann die Anzahl erkrankter Urlauber approximiert werden? Wie gross ist der Erwartungswert und die Varianz dieser Verteilung?

c) Berechnen Sie mit Hilfe dieser Näherung die Wahrscheinlichkeit, dass mehr als 110 der befragten Urlauber erkrankten.

d) Eine Gruppe von Ärzten vermutet, dass durchschnittlich mehr als 30% aller Reisenden erkrankt. Bestätigt die Studie die Vermutung der Ärzte? Führen Sie dazu einen Test auf dem 5% Niveau durch und geben Sie den entsprechenden Verwerfungsbereich an. Benutzen Sie die Approximation.

Als Approximation der Binomialverteilung wird manchmal auch die Poisson-Verteilung verwendet.

e) Wiederholen Sie die Rechnung aus Teilaufgabe a) mit Hilfe der Poisson-Approximation und vergleichen Sie die beiden Resultate. Was schliessen Sie daraus?

3. (10 Punkte)

Bei einer Wahl stellen drei verschiedene Parteien – nennen wir sie A, B und C – je einen Kandidaten. Es seien im Wahlkreis zudem 100'000 Wahlberechtigte.

Aus Meinungsumfragen sowie den Resultaten vergangener Wahlen sei bekannt, dass die Anzahl wählender Anhänger der Parteien alle unabhängig sind, sowie näherungsweise normalverteilt mit folgenden Parametern:

Partei	Erwartungswert	Standardabweichung
A	29000	3000
B	29000	4000
C	23000	9000

- Wenn nur parteitreue Wähler an die Urne gehen, wie gross ist die Wahrscheinlichkeit, dass Partei B mehr Stimmen erhält als Partei C?
- Ein mit Partei B verbandelter Geschäftsmann kauft alle nicht verwendeten Wahlzettel auf und verwendet sie, um für Partei B zu stimmen. Wie gross ist die Wahrscheinlichkeit, dass die Parteien A und C gemeinsam mehr Stimmen erzielen als Partei B?
- Es seien nebst den parteitreuen Wählern noch 3000 unentschlossene Wähler im Wahlkreis, die allesamt für diejenige Partei stimmen, die zum Zeitpunkt der Wahl die besten Schlagzeilen liefert. Die Wahrscheinlichkeit jeder Partei, diese zusätzlichen 3000 Stimmen zu erhalten, betrage unabhängig vom Verhalten der parteitreuen Wähler $1/3$. Wie gross wird nun die Wahrscheinlichkeit aus Aufgabe a)?

Hinweis: Die Summe von normalverteilten Zufallsvariablen ist wiederum normalverteilt

4. (7 Punkte)

1) Welches der folgenden Modelle ist ein lineares Modell?

$$\begin{aligned}\text{Modell 1: } Y &= \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2) \\ \text{Modell 2: } Y &= \beta_0 + \beta_1 X_1 + \beta_2^2 X_2 + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2) \\ \text{Modell 3: } \log Y &= \beta_0 + \beta_1 X^{-1} + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2)\end{aligned}$$

- | | |
|--------------------------|--------------------------|
| a) Nur Modell 1 | b) Nur Modell 2 |
| c) Nur Modell 3 | d) Modell 1 und Modell 2 |
| e) Modell 1 und Modell 3 | f) Modell 2 und Modell 3 |

Die Forschungsgruppe Wahlen in der Stadt Dresden (Deutschland) hat 2007 eine Erhebung gemacht, wie sich die Wählerschaft der rechtsradikalen Partei NPD in der Region zusammensetzt. Dazu hat sie in den einzelnen Gemeinden folgende Daten erhoben:

- den Wähleranteil (`electorate`) der NPD in Prozent,
- den Arbeitslosenanteil (`unemployed`) in Prozent,
- den Anteil Personen aus bildungsfernen Schichten (`uneducated`) in Prozent.

Diese Daten wurden mit dem folgenden Regressionsmodell untersucht:

$$\text{electorate} = \beta_0 + \beta_1 \cdot \text{unemployed} + \beta_2 \cdot \text{uneducated} + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2).$$

Der R-Output sieht folgendermassen aus:

Call:

```
lm(formula = electorate ~ unemployed + uneducated)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-9.82808	-1.48561	0.08649	1.54206	9.84760

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1.7086	2.0756	-0.823	0.4171
unemployed	0.2174	0.5827	0.373	0.7118
uneducated	0.8827	0.5024	1.757	0.0895 .

Residual standard error: 3.944 on 29 degrees of freedom

Multiple R-Squared: 0.5325, Adjusted R-squared: 0.5003

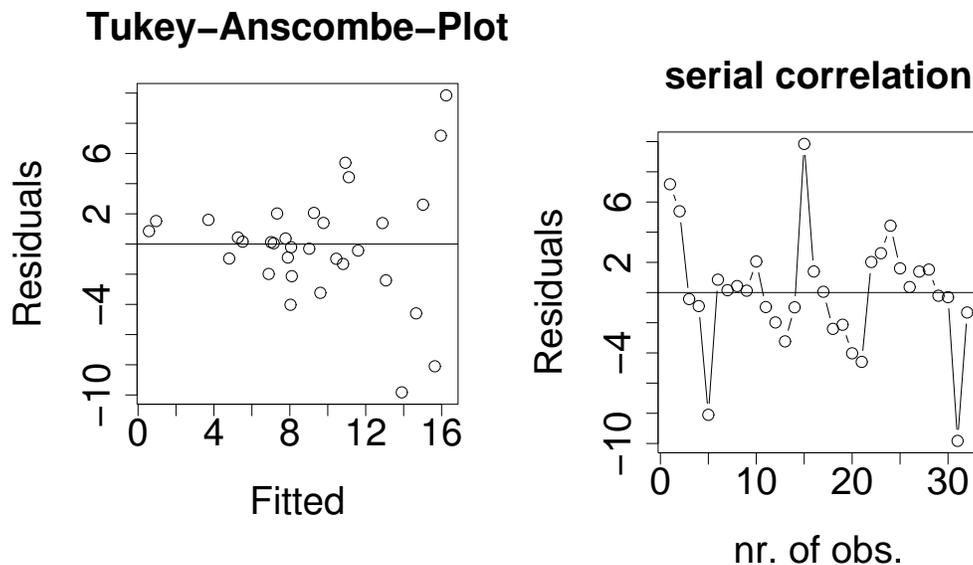
F-statistic: 16.52 on 2 and 29 DF, p-value: 1.627e-05

2) Mit wie vielen Beobachtungen wurde das Modell berechnet?

- | | | |
|-------|-------|-------|
| a) 27 | b) 28 | c) 29 |
| d) 30 | e) 31 | f) 32 |

- 3) Hat mindestens eine der beiden Erklärenden `unemployed` und `uneducated` einen signifikanten Einfluss auf die Zielgrösse?
- Ja.
 - Nein.
 - Man müsste zwei einfache Regressionen machen, um dies zu entscheiden.
- 4) Was ist ein exaktes zweiseitiges 95%-Vertrauensintervall für β_1 ?
- $0.2174 \pm 2.05 \cdot 0.5827$
 - $0.2174 \pm 1.96 \cdot 0.5827$
 - $0.2174 \pm 2.05 \cdot \frac{0.5827}{\sqrt{29}}$
 - $0.2174 \pm 1.96 \cdot \frac{0.5827}{\sqrt{29}}$
- 5) Der Wert $R^2 = 0.5325$ ist relativ klein. Wie verändert sich R^2 , wenn die Forschungsgruppe neben `unemployed` und `uneducated` noch weitere Erklärende ins obige multiple Regressionsmodell aufnimmt?
- R^2 bleibt unverändert.
 - R^2 wird kleiner.
 - R^2 wird grösser.
 - lässt sich nicht generell entscheiden.

Hier sind der Tukey-Anscombe und der Plot der seriellen Korrelation zu obigem Regressionsmodell:



- 6) Welche der untenstehenden Massnahmen ist aufgrund des Tukey-Anscombe-Plots angezeigt?
- Keine, der Plot ist in Ordnung.
 - Logarithmieren von `unemployed`.
 - Logarithmieren von `electorate`.
 - Wurzel ziehen aus `unemployed` und `uneducated`.
 - Logarithmieren von `unemployed` und `uneducated`.
 - Einen quadratischen Term ins Modell aufnehmen.
- 7) Welcher der folgenden Schlüsse ist aufgrund des Plots der seriellen Korrelation zulässig?
- Die Residuen sind nicht mit der Beobachtungsreihenfolge korreliert.
 - Die Residuen sind mit der Beobachtungsreihenfolge korreliert, dies hat einen Einfluss auf Tests und Vertrauensintervalle für die Koeffizienten β_i .

- c) Die Residuen sind mit der Beobachtungsreihenfolge korreliert, dies hat aber keinen Einfluss auf Tests und Vertrauensintervalle für die Koeffizienten β_i .
- d) Die Beobachtungsreihenfolge sollte als Erklärende ins Modell aufgenommen werden.

5. (10 Punkte)

1) Mit welcher Verteilung lässt sich am ehesten die Zeit, bis eine Glühbirne kaputt geht, beschreiben?

- | | |
|----------------|-----------------|
| a) Poisson | b) Uniform |
| c) Binomial | d) Normal |
| e) Exponential | f) t-Verteilung |

Seien X und Y unabhängige Zufallsvariablen mit $E[X] = 2$, $E[Y] = 3$, $\text{Var}(X) = 10$ und $\text{Var}(Y) = 1$.

2) Berechnen Sie $\text{Var}(X - 2Y + 3)$.

- | | | |
|-------|-------|-----------------------------------|
| a) 6 | b) 8 | c) 10 |
| d) 12 | e) 14 | f) Aussagen a) bis e) sind falsch |

3) Berechnen Sie $E[X^2]$.

- | | | |
|-------|-------|-----------------------------------|
| a) 6 | b) 8 | c) 10 |
| d) 12 | e) 14 | f) Aussagen a) bis e) sind falsch |

Die Zufallsvariable X hat die kumulative Verteilungsfunktion

$$F(x) = \begin{cases} x^5 & x \in [0, 1] \\ 0 & x \leq 0 \\ 1 & x \geq 1 \end{cases}.$$

4) Die Dichte von X ist auf dem Intervall $[0, 1]$ gleich:

- | | | |
|-----------|---------------------|-----------------------------------|
| a) $4x^4$ | b) $3x^4$ | c) x^4 |
| d) $5x^6$ | e) $\frac{1}{6}x^6$ | f) Aussagen a) bis e) sind falsch |

5) Für den Erwartungswert von X gilt:

- | | | |
|-------------------------|----------------|-----------------------------------|
| a) $E[X] = 0$ | b) $E[X] = 1$ | c) $E[X] = \frac{5}{6}$ |
| d) $E[X] = \frac{3}{4}$ | e) $E[X] = 30$ | f) Aussagen a) bis e) sind falsch |

6) Berechnen Sie $P[F(X) \leq y]$ für $y \in [0, 1]$:

- | | | |
|------------|----------------------|----------|
| a) y | b) $y^{\frac{1}{5}}$ | c) y^5 |
| d) $1 - y$ | e) 0 | f) 1 |

X_1, X_2, \dots, X_{200} sind i.i.d. UNIF[1,13]-verteilte Zufallsvariablen.

7) Wie ist $\sum_{i=1}^{200} X_i$ approximativ verteilt?

- a) $\mathcal{N}(1400, 2400)$
- b) $\mathcal{N}(1400, 200)$
- c) $\mathcal{N}(1400, 28800)$
- d) $\mathcal{N}(700, 200)$
- e) Poisson(700)
- f) Poisson(200)

8) Berechnen Sie approximativ $P[\sum_{i=1}^{200} X_i \geq 1400]$.

- | | | |
|---------|----------|-----------|
| a) 0.25 | b) 0.5 | c) 0.0075 |
| d) 0.75 | e) 0.025 | f) 0.075 |

X_1 und X_2 sind i.i.d. Bernoulli($\frac{2}{3}$)-verteilt.

9) Berechnen Sie die Wahrscheinlichkeit, dass $X_1 + X_2 = 1$ ist.

a) $\frac{1}{9}$
d) $\frac{5}{9}$

b) $\frac{2}{9}$
e) $\frac{4}{9}$

c) $\frac{1}{3}$

f) Aussagen a) bis e) sind falsch

10) Berechnen Sie die Wahrscheinlichkeit, dass $X_1 = X_2$ ist.

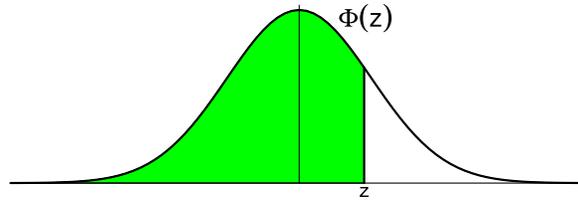
a) $\frac{1}{9}$
d) $\frac{5}{9}$

b) $\frac{2}{9}$
e) $\frac{4}{9}$

c) $\frac{1}{2}$

f) Aussagen a) bis e) sind falsch

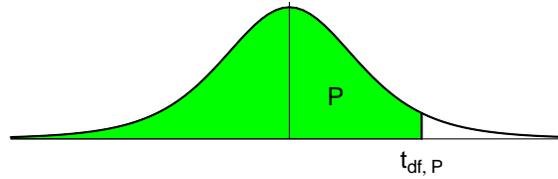
Tabelle der Kumulativen Normalverteilung $\Phi(z) = P[Z \leq z]$, $Z \sim \mathcal{N}(0, 1)$



Bsp.: $P[Z \leq 1.96] = 0.975$

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