

## 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 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. (7 Punkte)

Nehmen Sie an, dass von einer sehr grossen Fuchs-Population ein Drittel mit Tollwut infiziert ist.

- a) Man fängt zufällig 18 Tiere ein. Die Anzahl eingefangener Tiere, die mit Tollwut infiziert sind, wird mit  $X$  bezeichnet. Wie gross ist der Erwartungswert  $E(X)$  und die Varianz  $\text{Var}(X)$ ?
- b) Wieviele Tiere müssen mindestens zufällig eingefangen werden, damit die Wahrscheinlichkeit, mindestens ein gesundes Tier darunter zu haben, grösser als 90% ist?
- c) In Wahrheit ist der Anteil  $p$  der Fuchspopulation, der mit Tollwut infiziert ist, unbekannt. Um diesen Anteil zu bestimmen, fängt man 100 Füchse zufällig ein. Es stellt sich heraus, dass 20 von ihnen mit Tollwut infiziert sind. Geben Sie ein 95%-Vertrauensintervall für den Anteil  $p$  der Fuchspopulation an, der mit Tollwut infiziert ist.

*Hinweis:* Der Anteil  $p$  ist sicher zwischen 0 und 1! Die Endpunkte des Vertrauensintervalls müssen somit grösser als 0 und kleiner als 1 sein.

**2. (11 Punkte)**

Die Anzahl Bestellungen  $X$ , die in einer kleinen Firma  $A$  pro Werktag eingehen, kann als Poisson-verteilt angenommen werden. Ein externer Kontrolleur besucht die Firma. Am Tag der Kontrolle stellt er fest, dass nur eine einzige Bestellung eingegangen ist.

- a) Die Angestellten der Firma beteuern, dass im Schnitt pro Tag mindestens 6 Bestellungen eingehen. Kann diese Behauptung statistisch widerlegt werden mit der Tatsache, dass am Tag der Kontrolle nur eine einzige Bestellung eingegangen ist? Führen Sie einen statistischen Test auf dem 5%- Niveau durch. Geben Sie Null- und Alternativhypothese, sowie den p-Wert des Tests und die Testentscheidung an.
- b) Bei einer grösseren Firma  $B$  sind die Bestellungseingänge  $X$  pro Werktag Poisson-verteilt und im Schnitt gehen pro Tag 100 Bestellungen ein. Wie hoch ist die Varianz  $\text{Var}(X)$  der Bestellungen?
- c) Nehmen Sie für die Firma  $B$  aus Teilaufgabe b) an, dass die Bestellungseingänge von Tag zu Tag unabhängig sind. Bezeichnen Sie mit  $Y$  die gesamte Anzahl Bestellungen in 100 Werktagen. Geben Sie die exakte Verteilung von  $Y$  und eine passende Approximation an.
- d) Berechnen Sie mit Hilfe der approximativen Verteilung die Wahrscheinlichkeit, dass in 100 Werktagen bei Firma  $B$  mehr als 10200 Bestellungen eingehen.

*Hinweis:* Näherungswerte der Exponentialfunktion:

$\exp(-1)$	$\exp(-2)$	$\exp(-3)$	$\exp(-4)$	$\exp(-5)$	$\exp(-6)$	$\exp(-7)$	$\exp(-8)$
0.37	0.14	0.05	0.018	0.0067	0.0025	0.00091	0.00034

### 3. (17 Punkte)

Die Geschwindigkeit von einem alten und einem neu gekauften PC soll miteinander verglichen werden. 10 zufällig ausgewählte Arbeiten (computer jobs) werden mit beiden Maschinen einmal durchgeführt und die nötigen Ausführungszeiten in Sekunden gemessen.

Man findet (mit  $A$  werden die Ausführungszeiten des alten PC bezeichnet, mit  $N$  die Ausführungszeiten des neuen PC):

	1	2	3	4	5	6	7	8	9	10
$A$	20	30	25	50	10	42	27	34	14	45
$N$	14	19	16	32	6	28	78	23	9	32
$D = N - A$	-6	-11	-9	-18	-4	-14	51	-11	-5	-13

Die aus den Daten berechneten Mittelwerte und Standardabweichungen sind:

$$\bar{A} = 29.7; \quad \bar{N} = 25.7; \quad \bar{D} = -4$$

$$s_A = 13.22; \quad s_N = 20.47; \quad s_D = 19.80$$

Zusätzlich:  $(\frac{1}{2})^{10} = 0.00098 \approx 0.001$ .

Man möchte nun untersuchen, ob der neue PC generell schneller als der alte ist.

- Es handelt sich hier um 2 gepaarte Stichproben. Wieso? Begründen Sie *kurz*.
- Führen Sie einen t-Test auf dem 5%-Niveau durch. Geben Sie die Modellannahmen, die Nullhypothese und die Alternative, den Wert der Teststatistik, den Verwerfungsbereich und die Entscheidung an.
- Führen Sie einen Vorzeichentest auf dem 5%-Niveau durch. Geben Sie die Nullhypothese und die Alternative, den p-Wert und die Entscheidung an.
- Nehmen Sie an, dass der Wilcoxon-Test einen p-Wert von 0.041 liefert für die gleiche Alternative wie in **b**). Was können Sie daraus schliessen? (Niveau 5%.)
- Würden Sie für diese Daten den Wilcoxon-Test oder den t-Test vorziehen? Begründen Sie *kurz*.

## 4. (10 Punkte)

Bei der Reifung von Käse laufen verschiedene chemische Prozesse ab, welche den Geschmack günstig oder ungünstig beeinflussen können. In einer Studie sollte untersucht werden, ob drei routinemässig gemessene chemische Grössen `Acetic`, `H2S`, `Lactic` einen Einfluss auf den von Testpersonen beurteilten Geschmack (Variable `Taste`) von Cheddar-Käse haben. Als Zielgrösse (`Taste`) wird das Mittel der Geschmacksbeurteilung durch mehrere Tester betrachtet. Die erklärenden Grössen haben folgende Bedeutung:

`Acetic` : Essigsäure-Konzentration (logarithmiert)  
`H2S` : Schwefelwasserstoff-Konzentration (logarithmiert)  
`Lactic` : Milchsäure-Konzentration

Es wird das Modell

$$\text{Taste}_i = \beta_0 + \beta_1 \cdot \text{Acetic}_i + \beta_2 \cdot \text{H2S}_i + \beta_3 \cdot \text{Lactic}_i + \varepsilon_i; \quad \varepsilon_i \text{ i.i.d. } \sim \mathcal{N}(0, \sigma^2)$$

angepasst.

Mit R findet man:

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-17.757	17.625	-1.007	0.32335
Acetic	-2.470	4.004	-0.617	0.54279
H2S	4.039	1.091	3.702	0.00106 **
Lactic	21.458	7.559	??	0.00886 **

---

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

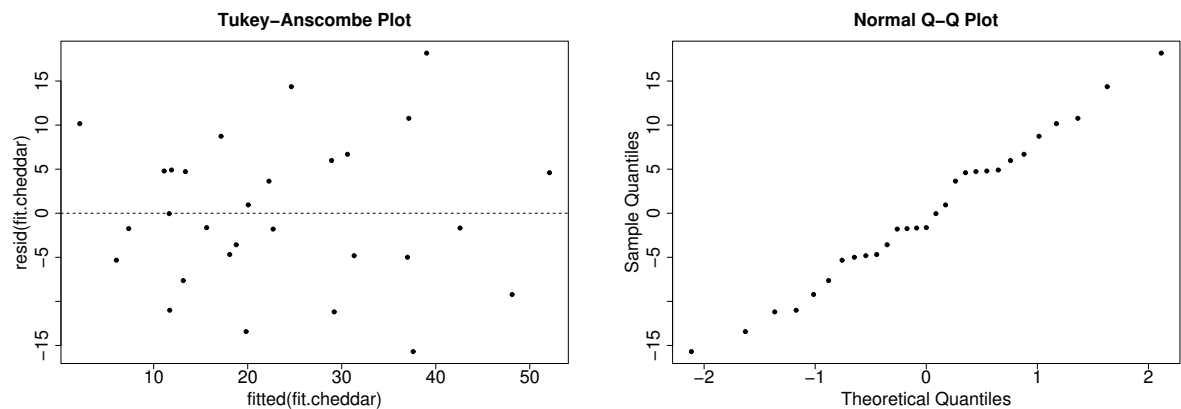
Residual standard error: 8.847 on 25 degrees of freedom

Multiple R-Squared: 0.7083, Adjusted R-squared: 0.6733

F-statistic: 20.24 on 3 and 25 DF, p-value: 7.18e-07

- Mit wievielen Beobachtungen wurde diese Regression gerechnet?
  - 25
  - 26
  - 27
  - 28
  - 29
  - Keine von diesen Werten
- Wird die Nullhypothese  $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$  auf dem 5%-Niveau verworfen?
  - Ja
  - Nein
  - Keine Aussage ist möglich
- Ist es auf dem 5% Niveau statistisch gesichert, dass in diesem Regressionsmodell ein Zusammenhang zwischen `Taste` und `Acetic` besteht (nach Berücksichtigung der Effekte der anderen Variablen)?
  - Ja
  - Nein
  - Keine Aussage ist möglich
- Welches Intervall ist ein exaktes zweiseitiges 95%-Vertrauensintervall für den Koeffizient  $\beta_2$ ?
  - $4.039 \pm 1.64 \cdot 1.091$
  - $4.039 \pm 1.708 \cdot 1.091$
  - $4.039 \pm 1.96 \cdot 1.091$
  - $4.039 \pm 2.060 \cdot 1.091$
  - $4.039 \pm 1.96 \cdot \frac{1.091}{\sqrt{25}}$
  - $4.039 \pm 2.060 \cdot \frac{1.091}{\sqrt{25}}$
- Kann man aufgrund des p-Wertes von `Lactic` entscheiden, ob das zweiseitige 95%-Vertrauensintervall für den Koeffizienten von `Lactic` den Wert 0 enthält?
  - 0 liegt im 95%-Vertrauensintervall
  - 0 liegt nicht im 95%-Vertrauensintervall
  - Kann mit den vorhandenen Angaben nicht entschieden werden

- 6) Wie gross ist der t-Wert der Variable `Lactic`?
- a) 0.352      b) 2.225      c) 2.839  
d) 12.127      e) 14.194      f) 15.287
- 7) Lassen sich Abweichungen von den Modellannahmen feststellen? (Betrachten Sie die gegebenen Plots.)
- a) Die Normalitätsannahme der Fehler  $\varepsilon_i$  ist verletzt.  
b) Die Annahme der konstanten Varianz der Fehler  $\varepsilon_i$  ist verletzt.  
c) Die Annahmen der Normalität und der konstanten Varianz der Fehler  $\varepsilon_i$  sind verletzt.  
d) Es gibt mindestens 4 Ausreisser.  
e) Die idealisierten Annahmen über die Fehler  $\varepsilon_i$  in der kleinste-Quadrate Regression treffen gut zu.  
f) Kann mit den vorhandenen Angaben nicht entschieden werden.



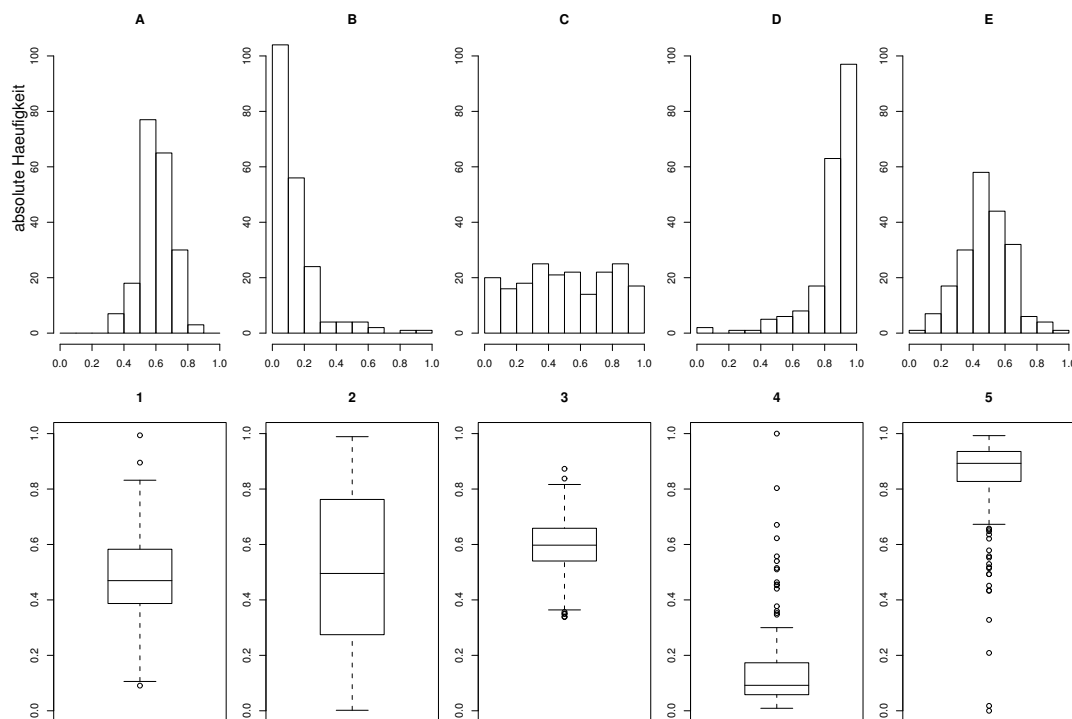
- 8) Wie gross ist die Schätzung von  $\sigma^2$ ?
- a) 0.6733      b) 0.7083      c) 8.847  
d) 17.625      e) 20.24      f) 78.269
- 9) Wie gross ist der prognostizierte Wert für `Taste` falls `Acetic` = 5, `H2S` = 10, `Lactic` = 1?
- a) -0.804      b) 31.741      c) 42.588  
d) 56.114      e) 110.010      f) 188.512
- 10) Lässt sich das Modell mittels “backward elimination” vereinfachen?
- a) Das Modell lässt sich nicht vereinfachen.  
b) Man lässt zuerst die Variable `Acetic` weg, man berechnet das neue Modell und man iteriert so weiter, bis alle Variablen im Modell signifikant sind.  
c) Man lässt zuerst die Variable `H2S` weg, man berechnet das neue Modell und man iteriert so weiter, bis alle Variablen im Modell signifikant sind.  
d) Man lässt zuerst die Variable `Lactic` weg, man berechnet das neue Modell und man iteriert so weiter, bis alle Variablen im Modell signifikant sind.  
e) Man lässt die Variablen `H2S` und `Lactic` weg, man berechnet das neue Modell und man schaut, ob die Variable `Acetic` signifikant ist.  
f) Kann mit den vorhandenen Angaben nicht entschieden werden.

## 5. (7 Punkte)

Verschiedene Aufgaben aus mehreren Gebieten:

- 1) Für fünf Stichproben vom Umfang  $n = 200$  wurden je ein Histogramm und ein Boxplot gezeichnet. Ordne die fünf Boxplots den entsprechenden Histogrammen zu.

- |                       |                       |
|-----------------------|-----------------------|
| a) A3, B2, C1, D5, E4 | b) A3, B4, C1, D5, E2 |
| c) A3, B4, C2, D5, E1 | d) A1, B4, C3, D5, E2 |
| e) A1, B5, C3, D4, E2 | f) A2, B4, C1, D5, E3 |



- 2) Gegeben seien 10 Multiple-Choice Fragen mit jeweils 4 möglichen Antworten (wobei jeweils genau eine davon richtig ist). Welche Verteilung eignet sich am besten für die Beschreibung der Anzahl richtig beantworteter Fragen, wenn ein(e) StudentIn die Antworten zufällig ankreuzt?

- |                        |                           |
|------------------------|---------------------------|
| a) Uniforme Verteilung | b) Poisson-Verteilung     |
| c) Binomial-Verteilung | d) Exponential-Verteilung |

- 3) Mit welcher Verteilung lässt sich am besten die Körperlänge von termingerechten Neugeborenen in einem Spital beschreiben?

- |                           |                       |
|---------------------------|-----------------------|
| a) Uniforme Verteilung    | b) Poisson-Verteilung |
| c) Exponential-Verteilung | d) Normal-Verteilung  |

- 4) Welche Verteilung eignet sich am besten für die Beschreibung der Anzahl VolleyballspielerInnen in der Schweiz, die im nächsten Jahr einen Bänderriss erleiden werden?

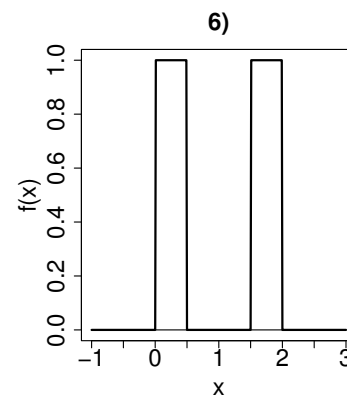
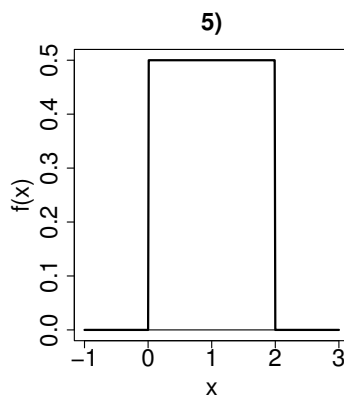
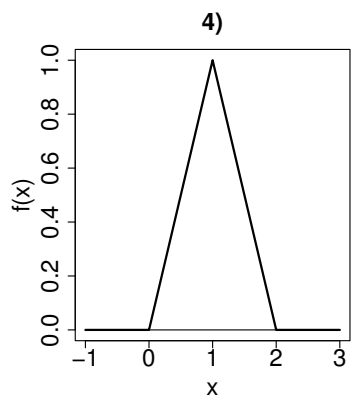
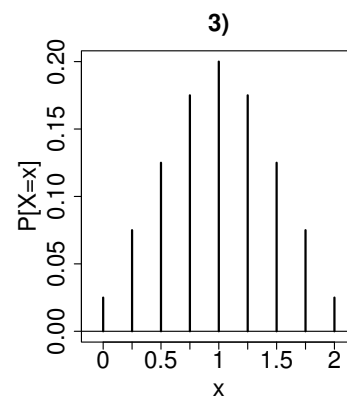
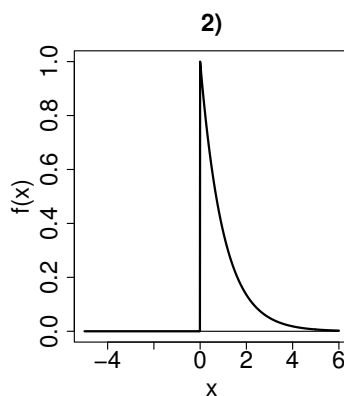
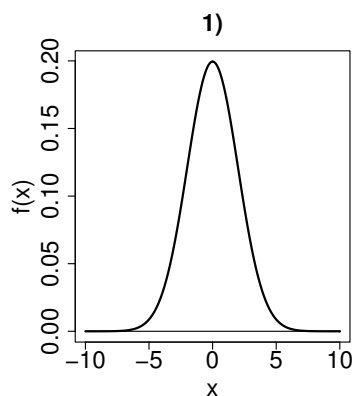
- |                        |                           |
|------------------------|---------------------------|
| a) Uniforme Verteilung | b) Poisson-Verteilung     |
| c) Binomial-Verteilung | d) Exponential-Verteilung |

- 5) Welche Verteilung eignet sich am besten für die Beschreibung der Anzahl VolleyballspielerInnen aus einer Menge von 100 zufällig ausgewählten, die im nächsten Jahr einen Bänderriss erleiden werden?

- |                        |                           |
|------------------------|---------------------------|
| a) Uniforme Verteilung | b) Poisson-Verteilung     |
| c) Binomial-Verteilung | d) Exponential-Verteilung |

- 6) Gegeben seien 2 unabhängige Uniform-(0, 1) verteilte Zufallsvariablen  $X_1$  und  $X_2$ . Welche der folgenden Dichten ist die Dichte der neuen Zufallsvariable  $Z = X_1 + X_2$ ?

- a) 1)      b) 2)      c) 3)      d) 4)      e) 5)      f) 6)



- 7) Gegeben seien die folgenden 6 Funktionen.

- A) für gegebene  $\mu$  und  $\sigma > 0$  sei

$$f(x) = \frac{1}{\sqrt{2\pi}\cdot\sigma} \exp\left\{-0.5 \cdot \left(\frac{x-\mu}{\sigma}\right)^2\right\}$$

- B) für gegebenes  $\lambda > 0$  sei

$$f(x) = \begin{cases} 0 & \text{falls } x \leq 0 \\ \lambda \cdot \exp\{-\lambda \cdot x\} & \text{falls } x > 0 \end{cases}$$

- C) sei

$$f(x) = \begin{cases} 1 & \text{falls } x \in (0, 2) \\ 0 & \text{sonst} \end{cases}$$

- D) für eine gegebene natürliche Zahl  $N$  und  $\lambda > 0$  sei

$$P[X = k] = \begin{cases} \frac{\lambda^k}{k!} \cdot \exp\{-\lambda\} & \text{falls } k = 1, \dots, N. \\ 0 & \text{sonst} \end{cases}$$

E) sei

$$P[X = k] = \begin{cases} 1/9 & \text{falls } k = 0 \\ 3/9 & \text{falls } k = 5 \\ 4/9 & \text{falls } k = 10 \\ 2/9 & \text{falls } k = 20 \\ 0 & \text{sonst} \end{cases}$$

F) für eine gegebene ganze Zahl  $n$  und  $p \in [0, 1]$  sei

$$P[X = k] = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}; \quad k = 0, \dots, n$$

Welche stellen eine Wahrscheinlichkeitsdichte oder eine Wahrscheinlichkeitsverteilung dar?

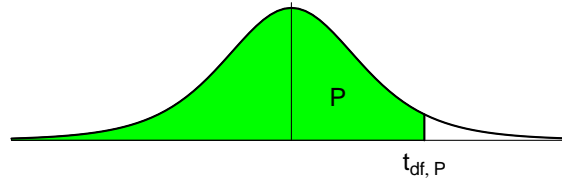
- a) A, B, F
- c) B, C, D, F
- e) alle

- b) A, B, D, F
- d) A, B, D, E, F
- f) Keine von diesen Antworten





## 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
$\infty$	0.253	0.524	0.842	1.282	1.645	1.960	2.326	2.576