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. (7 Punkte) Dieses Jahr waren die Organisatoren des Tennisturniers von Basel wegen der höheren Geschwindigkeiten der Aufschläge enttäuscht. Es war für die Zuschauer schwierig das Spielgeschehen mitzuverfolgen. Das Ziel für nächstes Jahr besteht darin, die Geschwindigkeit der Aufschläge zu verringern. Die Organisatoren glauben, dass das Problem von den Bällen verursacht wird. Die zuständige Ball-Firma schlägt einen neuen Typ von Ball vor. Um nun den alten Typ (X), welcher dieses Jahr benutzt wurde, mit dem neuen Typ (Y) zu vergleichen, haben sich die Organisatoren für einen Test entschieden. Um diesen Test durchzuführen, haben sie die besten 12 Tennisspieler der Welt ausgewählt: jeder dieser Athleten hat zuerst einen Aufschlag mit dem einen (zufällig gewählten) Balltyp und dann einen mit dem anderen gemacht. Beide Male wird die Geschwindigkeit gemessen. Dies sind die Resultate (dabei bezeichnen X die Werte für den alten Ball, Y die Werte für den neuen Ball und $D = X - Y$ die Differenz in Stundenkilometer):

X	$\hat{\mu}_X$	$=$	194.95	$\hat{\sigma}_X$	$=$	5.18
Y	$\hat{\mu}_Y$	$=$	194.68	$\hat{\sigma}_Y$	$=$	4.74
D	$\hat{\mu}_X - \hat{\mu}_Y$	$=$	0.27	$\hat{\sigma}_{X-Y}$	$=$	0.72

Sie dürfen davon ausgehen, dass die Geschwindigkeiten durch unabhängige $\mathcal{N}(\mu_X, \sigma_X^2)$- resp. $\mathcal{N}(\mu_Y, \sigma_Y^2)$-verteilte Zufallsvariablen beschrieben werden können.

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

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

c) Führen Sie den geeigneten t-Test auf dem 5%-Niveau durch. Bestimmen Sie den Wert der Teststatistik, den Verwerfungsbereich und den Testentscheid.

d) Geben Sie ein einseitiges 95%-Vertrauensintervall an für die Differenz $\mu_X - \mu_Y$. (Einseitigkeit entsprechend der interessierenden Fragestellung.)

e) Wie ändert sich das einseitige 95%-Vertrauensintervall in d), wenn wir annehmen, dass σ_X, σ_Y und σ_{X-Y} bekannt sind? Geben Sie die allgemeine Formel an.

2. (8 Punkte)

Ein renommierter Ameisenforscher ist auf eine noch unbekannte Ameisenart gestossen. Er stellt fest, dass bei dieser Art die Ameisen entweder rot oder schwarz sind. Auf den ersten Blick nimmt er an, dass jede zehnte Ameise rot ist.

a) Wie gross ist die Wahrscheinlichkeit, dass von 5 zufällig ausgewählten Ameisen genau drei rot sind?

Der Forscher möchte jetzt die Ameisenart genauer untersuchen. Dazu nimmt er eine Stichprobe von 150 Ameisen.

b) Welche Verteilung besitzt die Anzahl der roten Ameisen? Berechnen Sie den Erwartungswert und die Varianz dieser Verteilung.

c) Verwenden Sie die Normal-Approximation zur Berechnung der Wahrscheinlichkeit, dass die Anzahl roter Ameisen zwischen 15 und 20 ist.

Der Forscher untersucht die 150 Ameisen. Er stellt fest, dass 20 Ameisen rot sind . Ist diese Häufigkeit vereinbar mit der bisher angenommenen Hypothese, dass durchschnittlich 10% der Ameisen rot sind?

d) Führen Sie dazu einen zweiseitigen Test auf dem 5% Niveau durch. Geben Sie die Null- und Alternativhypothese, die Teststatistik sowie den Testentscheid an. Schreiben Sie die Normal-Approximation hin und benutzen Sie diese.

3. (12.5 Punkte)

Am Ausgang eines Kleidergeschäftes wurde eine neue Alarmanlage installiert. Der Alarm wird in 89% der Fälle eines Diebstahls (wenn jemand mit einem gestohlenen Kleidungsstück das Geschäft verlässt) ausgelöst.

a) Drei Diebe verlassen nacheinander das Geschäft. Wie gross ist die Wahrscheinlichkeit, dass bei mindestens einem der drei der Alarm losgeht? Wie gross ist die Wahrscheinlichkeit, dass genau 2 der Diebe den Alarm auslösen?

b) An einem Aktionstag werden alle Pullis für 70.- verkauft. Insgesamt 20 Diebe haben versucht einen Pulli zu klauen. Wenn ein Dieb auf frischer Tat ertappt wird (durch Auslösen der Alarmanlage) muss er nebst dem Pulli noch eine Busse bezahlen. Wie hoch sollte diese Busse mindestens sein, um den erwarteten Verlust der gestohlenen Pullis (Alarmanlage wurde nicht ausgelöst) zu kompensieren?

c) Die Alarmanlage wird nicht nur durch Diebe ausgelöst. In 5% der Fällen startet der Alarm auch, wenn normale Kunden (keine Diebe) das Geschäft verlassen. Eine Verkäuferin hat beobachtet, dass ein Drittel der Alarme eigentlich nur Fehlalarme sind. Wie gross ist der Anteil an Dieben in der Kundschaft des Geschäftes?

Die Geschäftsbesitzerin verfolgt die Anzahl geschnappter Diebe per Monat schon seit einiger Zeit. Sie teilt dazu die Diebe in 4 Altersklassen auf, und nimmt an, dass die Anzahl Diebe in den verschiedenen Altersklassen normalverteilt sind mit den folgenden Parametern:

Altersklasse	Erwartungswert	Standardabweichung
jünger als 16	100	12
16 bis 20 jährige	150	10
21 bis 30 jährige	125	10
älter als 30	75	15

Die Besitzerin geht davon aus, dass die Anzahl erwischter Diebe in den unterschiedlichen Altersklassen unabhängig sind.

d) Wie gross ist die Wahrscheinlichkeit dafür, dass mehr junge Kunden (20 Jahre alt oder jünger) anstatt älterer Kunden als Diebe geschnappt werden?

e) Mit der neuen Alarmanlage wurden im letzten Monat 600 Diebe geschnappt. Die Besitzerin bemerkte, dass viel mehr Jugendliche unter 16 Jahren gefasst wurden als vorher.
Wie gross ist die Wahrscheinlichkeit, dass die Anzahl der Diebe in der Altersgruppe "jünger als 16" grösser ist als die Anzahl Diebe in allen restlichen Altersgruppen zusammen? (Insgesamt werden 600 Diebe geschnappt. Die Parameter für die drei älteren Altersgruppen sind immer noch sinnvoll.)

Hinweis: Die Summe von normalverteilten Zufallsvariablen ist wiederum normalverteilt

4. (8 Punkte)

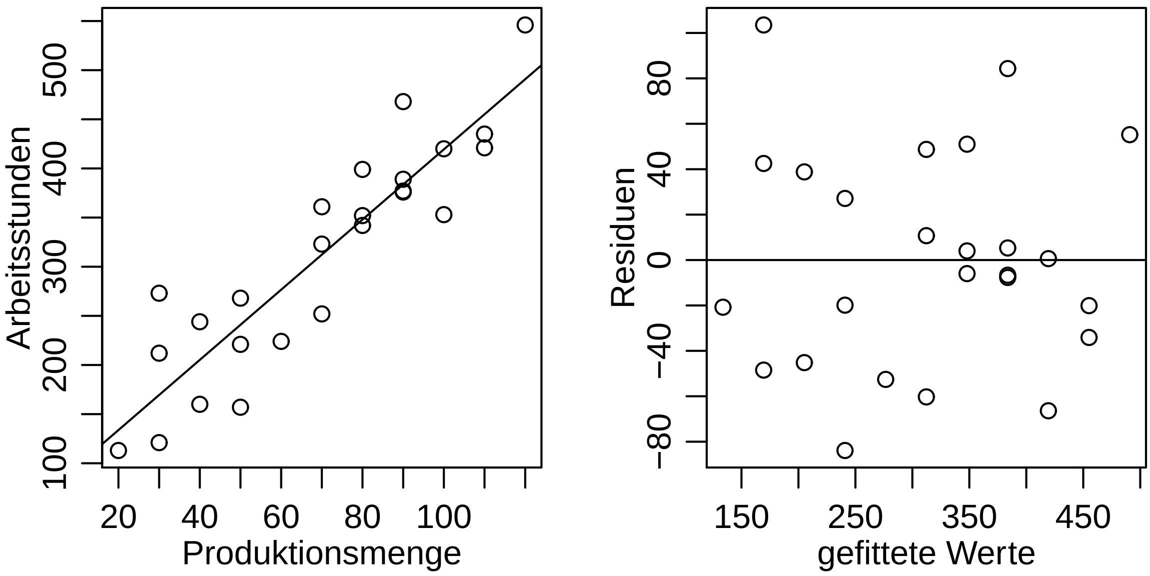
Die Pharmafirma Anex ist ein Grossproduzent für den Schweinegrippeimpfstoff. Anex hat Daten zur Produktionsmenge (in Packungen *à* 10 Einheiten) und den Arbeitsstunden der Mitarbeiter erhoben. Man untersucht jetzt die Arbeitsstunden, indem man ein lineares Regressionmodell der Form

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

anpasst (y_i sind die Arbeitsstunden, x_i sind die Produktionsmengen). Hier sind der R-Output und die Plots:

```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)        62.366          ?   2.382        ?
Produktionsmenge    3.570      0.347       ? 4.45e-10
---

Residual standard error: 48.82 on 23 degrees of freedom
Multiple R-squared: 0.8215,Adjusted R-squared: 0.8138
F-statistic: 105.9 on 1 and 23 DF,  p-value: 4.449e-10
```



1) Betrachte die gezeigten Plots. Welche der nachfolgenden Aussagen ist zutreffend?
   a) Die Normalitätsannahme der Fehler ϵ_i scheint verletzt zu sein.
   b) In der Modellannahme wurde ein quadratischer Term vergessen.
   c) Die Annahme konstanter Fehlervarianz scheint verletzt zu sein.
   d) Keine der Aussagen a) bis c) trifft zu.

2) Wieviele Arbeitsstunden sollte man nach dem linearen Modell erwarten für die Produktionsmenge von 130?
   a) 495.4 b) 591.8 c) 526.5 d) 501.3

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) -0.075 b) -13.36 c) 9.45
d) 10.290 e) 156.8 f) Keine Aussage möglich

5) Mit wievielen Beobachtungen wurde die Regression berechnet?

a) 20 b) 21 c) 22 d) 23 e) 24 f) 25

6) Welches der folgenden Intervalle ist ein exaktes zweiseitiges 99% Vertrauensintervall für β_1?

a) $3.570 \pm 2.797 \cdot 0.347$
b) $3.570 \pm 2.500 \cdot \frac{0.347}{\sqrt{23}}$
c) $3.570 \pm 2.797 \cdot \frac{0.347}{\sqrt{23}}$
d) $3.570 \pm 2.807 \cdot 0.347$
e) $3.570 \pm 2.500 \cdot 0.347$
f) $3.570 \pm 2.492 \cdot 0.347$

7) Enthält das zweiseitige 95% Vertrauensintervall für β_0 den Wert 0?

a) Ja b) Nein c) Keine Aussage möglich

8) Wie gross ist die Schätzung der Standardabweichung σ der Fehler ϵ_i?

a) 81.10 b) 0.17 c) 38.81 d) 14.52 e) 48.82

5. (8 Punkte)

1) Für eine Zufallsvariable X seien $\mathbf{E}[X] = 5$ und $\mathrm{Var}(X) = 2$. Wie gross ist der Erwartungswert der Zufallsvariable $2X(1-X)$?

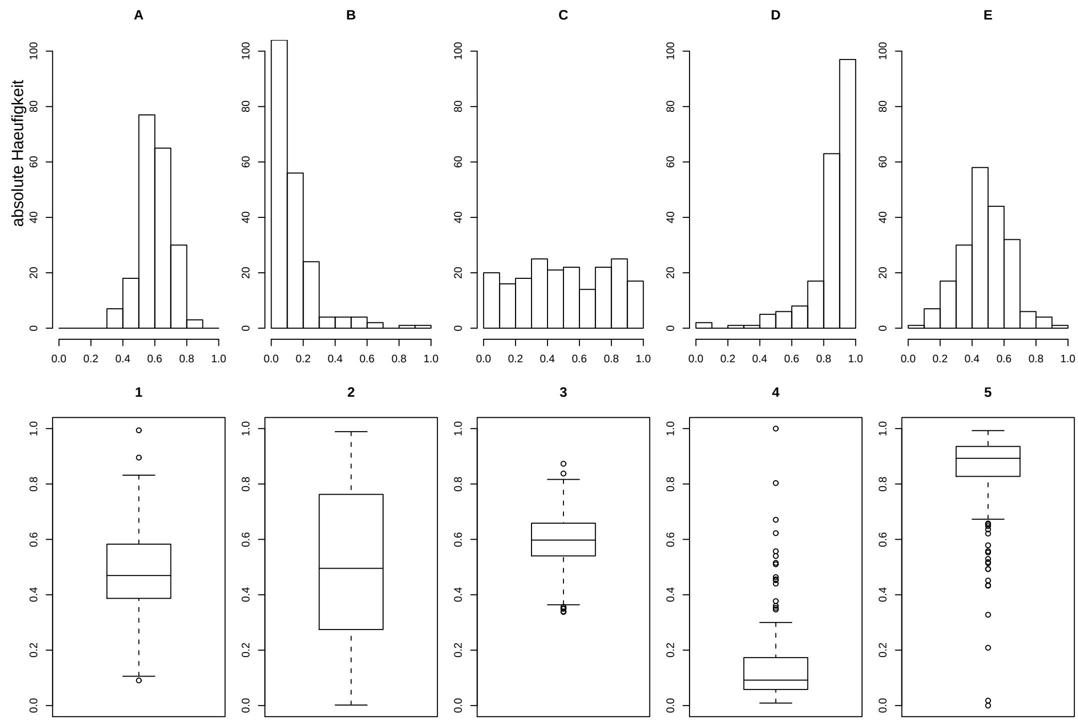
   a) 64   b) 3   c) -36
   d) -44 e) 6   f) keine Aussage möglich

2) Gleiche Ausgangslage wie in Teilaufgabe 1). Wie gross ist die Varianz der Zufallsvariable $-5X + 12$?

   a) 50   b) 22   c) -12
   d) 2   e) 10   f) keine Aussage möglich

3) 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



4) Die Funktion $F : \mathbb{R} \to \mathbb{R}$ ist definiert durch

$$F(x) = \begin{cases} b - e^{a-x} & x \geq a, \\ 0 & x < a. \end{cases}$$

   Für welche Parameter a und b kann die Funktion F die (kumulative) Verteilungsfunktion einer *positiven* Zufallsvariablen sein?

   a) $a = 1,\ b = 0$ b) $a = -1,\ b = 1$
   c) $a = 0,\ b = -1$ d) $a = 1,\ b = 1$

5) 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

6) Die tägliche Überzeit (in Minuten) einer Angestellten einer Grossbank sei gut mit einer Exponential-Verteilung mit Erwartungswert 10 modellierbar. Welche Verteilung eignet sich (näherungsweise) für die Beschreibung der totalen jährlichen Überzeit, wenn die Überzeit an den einzelnen Tagen als unabhängig betrachtet werden kann?

   a) Uniforme Verteilung
   b) Exponential-Verteilung
   c) Binomial-Verteilung
   d) Normal-Verteilung

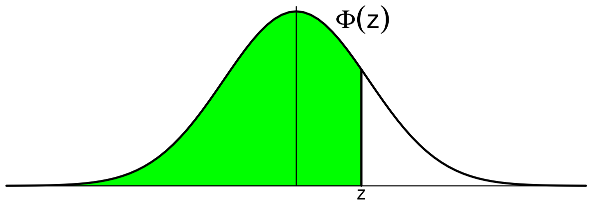
7) Welche der folgenden Aussagen über Vertrauensintervalle ist korrekt?

   a) Wenn das Niveau α eines statistischen Testes kleiner wird, dann wird das zugehörige (1-α) Vertrauensintervall kleiner.
   b) Wenn das Niveau α eines statistischen Testes kleiner wird, dann wird das zugehörige (1-α) Vertrauensintervall grösser.
   c) Das Niveau α und das zugehörige (1-α) Vertrauensintervall haben keine direkte Beziehung zueinander.
   d) Aussagen a) bis c) sind falsch.

8) Welche der folgenden Aussagen über den P-Wert ist zutreffend?

   a) Der P-Wert gibt die Wahrscheinlichkeit an, dass die Nullhypothese richtig ist.
   b) Den P-Wert gibt man sich vor der Durchführung eines statistischen Test vor.
   c) Der P-Wert hängt von den beobachteten Daten ab.
   d) Aussagen a) bis c) sind falsch.

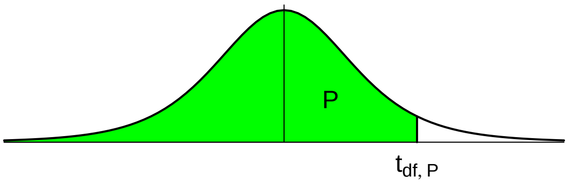
Tabelle der Kumulativen Normalverteilung $\Phi(z) = \mathrm{P}\left[Z \leq z\right], \ Z \sim \mathcal{N}(0,1)$



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