Seminar 7

Tabelle mit Quantilen

x	0.02	0.025	0.04	0.05	0.95	0.96	0.975	0.995
t.ppf$(x, 99)$	-2.08	-1.98	-1.77	-1.66	1.66	1.77	1.98	2.63
t.ppf$(x, 48)$	-2.11	-2.01	-1.79	-1.68	1.68	1.79	2.01	2.68
chi2.ppf$(x, 99)$	72.3	73.36	75.8	77	123.2	125	128.42	139
chi2.ppf$(x, 24)$	11.992	12.4	13.35	13.85	36.42	37.39	39.36	45.6
norm.ppf$(x, 0, 1)$	-2.05	-1.96	-1.75	-1.64	1.64	1.75	1.96	2.58

A1. Ein Internetprovider versichert seine Kunden, dass die durchschnittliche Internet-Verbindungsgeschwindigkeit zwischen 20:00 und 22:00 Uhr 250 Mbps ist mit einer Standardabweichung von 40 Mbps. Anhand einer Stichprobe von 100 Kunden wurde festgestellt, dass die durchschnittliche Verbindungsgeschwindigkeit (zwischen 20:00 und 22:00 Uhr) 242 Mbps beträgt.
(a) Man konstruiere ein 95% zweiseitiges Konfidenzintervall für die durchschnittliche Internet-Verbindungsgeschwindigkeit.
(b) Mit einem Signifikanzniveau von 5% teste man, ob die Behauptung des Internetproviders stimmt.
Lsg.: (a) $n = 100, \bar{x}_{100} = 242, \sigma = 40,\ \alpha = 0.05$;

$$\Big(\bar{x}_n - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}\ ,\ \bar{x}_n + \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}\Big),$$

wobei $\sqrt{n} = 10, z_{1-\frac{\alpha}{2}} = \texttt{norm.ppf}(1 - \frac{\alpha}{2}, 0, 1) = 1.96$. Der Wert des zweiseitigen Konfidenzintervalls ist $(234.16,\ 249.84)$.
(b) Methode 1: $H_0 : \mu = 250,\ H_1 : \mu \neq 250$, man wendet den Z-Test an (Test für Erwartungswert, wenn die Varianz $\sigma^2 = 40^2$ bekannt ist); $\mu_0 = 250 \notin (234.16,\ 249.84)$ H_0 wird angelehnt, die Behauptung des Internetproviders *stimmt nicht.*
Methode 2: $H_0 : \mu = 250,\ H_1 : \mu \neq 250$, man wendet den Z-Test an (Test für Erwartungswert, wenn die Varianz $\sigma^2 = 40^2$ bekannt ist); $\mu_0 = 250, n = 100, \sigma = 40, \bar{x}_{100} = 242,\ z_{1-\frac{\alpha}{2}} = \texttt{norm.ppf}(0.975, 0, 1) = 1.96$, $z = \frac{\bar{x}_{100}-250}{\frac{\sigma}{\sqrt{100}}} = \frac{242-250}{\frac{40}{10}} = -2,\ |z| \geq z_{1-\frac{\alpha}{2}} \implies$ (anhand der statistischen Daten) H_0 wird abgelehnt, *die Behauptung des Internetproviders stimmt nicht.*

A2. Ein Messgerät wird im Laufe der Zeit unbrauchbar. Als Kriterium für seine Brauchbarkeit dient die Forderung, dass die Standardabweichung der Messfehler < 0.0007 ist. Kann man bei einer Irrtumswahrscheinlichkeit von $\alpha = 0.05$ auf die Brauchbarkeit des Messgeräts schließen, wenn sich aus 25 Fehlerwerten ein Schätzwert für die Standardabweichung von 0.0009 ergeben hat? Man benutze zweiseitige Konfidenzintervalle für die Standardabweichung der Form $(\hat{\sigma}_1, \hat{\sigma}_2)$. Es wird angenommen, dass die Messfehler normal verteilt sind.
Lsg.: $\left(\sqrt{\frac{n-1}{c_{1-\frac{\alpha}{2}}}} \cdot s_n, \sqrt{\frac{n-1}{c_{\frac{\alpha}{2}}}} \cdot s_n\right)$, d.h. $\hat{\sigma}_1 = \sqrt{\frac{n-1}{c_{1-\frac{\alpha}{2}}}} \cdot s_n$, $\hat{\sigma}_2 = \sqrt{\frac{n-1}{c_{\frac{\alpha}{2}}}} \cdot s_n$ mit $\alpha = 0.05,\ n = 25, s_n = 0.0011, c_{\frac{\alpha}{2}} = \texttt{chi2.ppf}(0.025, 24) = 12.4, c_{1-\frac{\alpha}{2}} = \texttt{chi2.ppf}(0.975, 24) = 39.36$.
Man rechnet $\hat{\sigma}_1 = \sqrt{\frac{n-1}{c_{1-\frac{\alpha}{2}}}} \cdot s_n = \sqrt{\frac{24}{39.36}} \cdot 0.0009 \approx 0.0007027$, $\hat{\sigma}_2 = \sqrt{\frac{n-1}{c_{\frac{\alpha}{2}}}} \cdot s_n = \sqrt{\frac{24}{12.4}} \cdot 0.0009 \approx 0.001252$. Der Wert des zweiseitigen Konfidenzintervalls ist $(0.0007027, 0.001252)$; in diesem Intervall sind alle Werte größer als 0.0007, d.h. anhand der Daten kann man schlußfolgern: *das Messgerät ist nicht brauchbar.*

A3. Eine Firma stellt Steuerungssysteme für Klimaanlagen her und garantiert den Kunden, dass die Ausfallwahrscheinlichkeit weniger (kleiner oder gleich) als 6% beträgt. Es werden 400 Systeme untersucht und man stellt fest, dass 20 defekt waren. Ist die angegebene Garantie der Firma trotzdem gültig? ($\alpha = 0.05$).

Lsg.: Test für Anteilswert: $H_0 : p \leq 0.06, H_1 : p > 0.06$, $p_0 = 0.06$, $n = 400$, $\bar{x}_n = \frac{20}{400} = 0.05$; $\alpha = 0.05$; $np_0(1-p_0) = 400 \cdot 0.06 \cdot 0.94 = 22.56 > 10$

$$z = \frac{\bar{x}_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.05 - 0.06}{\sqrt{\frac{0.06 \cdot 0.94}{400}}} = -0.8422 < \texttt{norm.ppf}(0.95, 0, 1) \approx 1.64 \Rightarrow H_0 \text{ wird akzeptiert },$$

d.h. die angegebene Garantie der Firma ist gültig!

A4. Ein Restaurant möchte die Zufriedenheit seiner Kunden bewerten. Gegeben sind folgende statistische Daten für das Merkmal X = die Zufriedenheit des Kunden auf einer Skala von -2 bis 2:

Meinung der Kunden	Zufriedenheitswert	absolute Häufigkeit
Sehr zufrieden	2	14
Zufrieden	1	8
Neutral	0	4
Unzufrieden	-1	10
Sehr unzufrieden	-2	13

Man berechne (mit Taschenrechner oder mit Python): (a) das Stichprobenmittel; (b) die empirische Standardabweichung;
(c) das zweiseitige 95%-Konfidenzintervall für die durchschnittliche Zufriedenheit der Kunden.

Lsg.: (a) $n = 49, \bar{x}_n = 0$, (b) $s_n \approx 1.62$;
$\alpha = 0.01$, das zweiseitige Konfidenzintervall für die durchschnittliche Zufriedenheit der Kunden (theoretischer Erwartungswert) ist:
$\left(\bar{x}_n - \frac{s_n}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}} \, , \, \bar{x}_n + \frac{s_n}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}}\right) = \left(-\frac{1.62}{7} \cdot 2.01 \, , \, \frac{1.62}{7} \cdot 2.01\right) = (-0.465171, 0.465171)$, wobei $t_{1-\frac{\alpha}{2}} = \texttt{t.ppf}(0.975{,}48) = 2.01$.

Wiederholungsaufgaben:

W1. Eine stetige ZG X hat die Dichtefunktion $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \begin{cases} 0, & x \leq 0 \\ cxe^{-x}, & x > 0. \end{cases}$ Man bestimme $c \in \mathbb{R}$ und
a) die Erwartungswerte $E(X)$ und $E(X^2)$;
b) die Wahrscheinlichkeit des Ereignisses $\{|X - 3| > 2\}$;
c) die Wahrscheinlichkeit des Ereignisses $\{X < 3\}$, bedingt durch das Ereignis $\{X > 1\}$.

Lsg.: $1 = \int_{-\infty}^{\infty} f(x)dx = c \int_0^{\infty} xe^{-x}dx = c \implies c = 1.$

a) $E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_0^{\infty} x^2e^{-x}dx = 2$, $E(X^2) = \int_{-\infty}^{\infty} x^2f(x)dx = \int_0^{\infty} x^3e^{-x}dx = 6.$

b) $P(|X - 3| > 2) = P((X - 3 < -2) \cup (X - 3 > 2)) = P((X < 1) \cup (X > 5)) = F(1) + (1 - F(5)) = 1 - 2e^{-1} + 6e^{-5} \approx 0.3.$

c) $P(X < 3|X > 1) = \frac{P(1<X<3)}{P(X>1)} = \frac{F(3)-F(1)}{1-F(1)} = \frac{1-4e^{-3}-1+2e^{-1}}{2e^{-1}} = 1 - 2e^{-2} \approx 0.73$

W2. Für welche Konstante $c \in \mathbb{R}$ wird die Funktion $f : \mathbb{R} \to \mathbb{R}$ mit der Gestalt

$$f(x) = \begin{cases} cx & : 0 \leq x \leq 3, \\ c(6 - x) & : 3 < x \leq 6, \\ 0 & : \text{sonst} \end{cases}$$

zur Wahrscheinlichkeitsdichte einer Zufallsgröße X?
Man berechne die Verteilungsfunktion von X. Seien die Ereignisse $A = \{X < 3\}$ und $B = \left\{\frac{3}{2} \leq X \leq \frac{9}{2}\right\}$. Man berechne $P(A \cap B)$.

Lsg.: Es gilt $f \geq 0 \;\Leftrightarrow\; c \geq 0$. Es muss gelten

$$1 = \int_{-\infty}^{\infty} f(x)\,dx = \int_0^3 cx\,dx + \int_3^6 c(6-x)\,dx = \frac{1}{2}cx^2\Big|_0^3 + c\left(6x - \frac{1}{2}x^2\right)\Big|_3^6 = \frac{9}{2}c - 0 + c\left(36 - 18 - 18 + \frac{9}{2}\right) = 9c$$

Die Verteilungsfunktion von X ist $F(x) = \displaystyle\int_{-\infty}^{x} f(t)\,dt$

$$\begin{aligned}
x < 0:\; F(x) &= \int_{-\infty}^{x} 0\,dt = 0,\\
0 \leq x \leq 3:\; F(x) &= \int_0^x \frac{1}{9}t\,dt = \frac{1}{9}\frac{1}{2}t^2\Big|_0^x = \frac{1}{18}x^2,\\
3 < x \leq 6:\; F(x) &= \int_0^3 \frac{1}{9}t\,dt + \int_3^x \frac{1}{9}(6-t)\,dt = \frac{1}{9}\frac{1}{2}t^2\Big|_0^3 + \frac{1}{9}\left(6t - \frac{1}{2}t^2\right)\Big|_3^x = -\frac{1}{18}x^2 + \frac{2}{3}x - 1,\\
x > 6:\; F(x) &= \int_0^3 \frac{1}{9}t\,dt + \int_3^6 \frac{1}{9}(6-t)\,dt + \int_6^x 0\,dt = \frac{1}{9}\frac{1}{2}t^2\Big|_0^3 + \frac{1}{9}\left(6t - \frac{1}{2}t^2\right)\Big|_3^6 + 0 = 1.
\end{aligned}$$

Zusammenfassend gilt also

$$F(x) = \begin{cases} 0 & : x < 0,\\ \frac{1}{18}x^2 & : 0 \leq x \leq 3,\\ -\frac{1}{18}x^2 + \frac{2}{3}x - 1 & : 3 < x \leq 6,\\ 1 & : x > 6. \end{cases}$$

$$\Longrightarrow P(A \cap B) = P\left(\frac{3}{2} \leq X \leq 3\right) = F(3) - F\left(\frac{3}{2}\right) = \frac{1}{2} - \frac{1}{8} = \frac{3}{8}.$$

W3. Reagenzgläser sollen bezüglich ihrer Schmelztemperatur untersucht werden. Aus der Produktion wurden zufällig und unabhängig voneinander 100 Reagenzgläser entnommen. Von diesen 100 Gläsern wurden die Schmelztemperaturen (in °C) bestimmt. Der empirische Mittelwert dieser Werte ist 748.2 und die empirische Varianz ist 15.6025. Die zufällige Schmelztemperatur ist normalverteilt.
(a) Testen Sie mit Signifikanzniveau 1% ob die erwartete Schmelztemperatur signifikant verschieden ist von 750 oder nicht.
b) Testen Sie mit Signifikanzniveau 5%, ob die Standardabweichung der Schmelztemperatur signifikant verschieden von 4 ist oder nicht.
Lsg.: $n = 100, \bar{x}_n = 748.2, s_n^2 = 15.6, \mu_0 = 759, \sigma_0^2 = 15$
$H_0:\; \mu = 750,\; H_1:\; \mu \neq 750$ Student Test: Test für Erwatungswert, wenn die Varianz unbekannt ist; $\alpha = 0.01$, $t = \dfrac{\bar{x}_n - \mu_0}{\frac{s_n}{\sqrt{n}}} = \dfrac{748.2 - 750}{\frac{\sqrt{15.6}}{\sqrt{100}}} = -4.5573 \Rightarrow |t| > t_{0.995} = \texttt{t.ppf}(0.995, 99) = 2.62 \Longrightarrow\; H_0$ wird abgelehnt, d.h. anhand der Daten kann behauptet werden: der Erwartungswert der Schmelztemperatur *ist signifikant verschieden* von 750.
b) $H_0:\; \sigma = 4,\; H_1:\; \sigma \neq 4$; man führt den χ^2-Test für Varianz durch; $\alpha = 0.05$; $n = 100, s_n^2 = 15.6035, \sigma_0 = \sqrt{16} = 4$, das Konfidenzintervall für die Standardabweichung ist $\left(\sqrt{\frac{n-1}{c_{1-\frac{\alpha}{2}}}} \cdot s_n, \sqrt{\frac{n-1}{c_{\frac{\alpha}{2}}}} \cdot s_n\right)$, d.h. $\hat{\sigma}_1 = \sqrt{\frac{n-1}{c_{1-\frac{\alpha}{2}}}} \cdot s_n,\; \hat{\sigma}_2 = \sqrt{\frac{n-1}{c_{\frac{\alpha}{2}}}} \cdot s_n$ Man rechnet $\sqrt{15.6025} = 3.95$, $c_{\frac{\alpha}{2}} = \texttt{chi2.ppf}(0.05/2, 99) = 73.36$, $c_{1-\frac{\alpha}{2}} = \texttt{chi2.ppf}(1 - 0.05/2, 99) = 128.42$, $\hat{\sigma}_2 = \sqrt{\frac{n-1}{c_{1-\frac{\alpha}{2}}}} \cdot s_n = \sqrt{\frac{99}{73.36}} \cdot 3.95 \approx 4.588$, $\hat{\sigma}_1 = \sqrt{\frac{n-1}{c_{\frac{\alpha}{2}}}} \cdot s_n = \sqrt{\frac{99}{128.42}} \cdot 3.95 \approx 3.468$. Der Wert des zweiseitigen Konfidenzintervalls ist $(3.468, 4.588)$; $\sigma_0 = 4 \in (3.468, 4.588)$,

d.h. anhand der Daten kann man behaupten: die Standardabweichung der Schmelztemperatur ist *nicht signifikant verschieden* von 4.

W4. Ein Umweltforscher misst die Konzentration von PM2.5 Partikeln in der Luft in einem städtischen Gebiet. Die Daten einer Stichprobe sind in der folgenden Tabelle enthalten.

Konzentration µg/m^3 von PM2.5 Partikeln	absolute Häufigkeit
45	12
48	10
50	30
52	20
55	28

Berechnen Sie (mit Taschenrechner oder mit Python): (a) das Stichprobenmittel; (b) die empirische Standardabweichung;
(c) das zweiseitige 95%-Konfidenzintervall für die durchschnittliche Partikelkonzentration[1].
(d) die beiden einseitigen 95%-Konfidenzintervalle für die durchschnittliche Partikelkonzentration.

Lsg.: (a) $n = 100, \bar{x}_n = 51$, (b) $s_n \approx 3.20983$;
(c) das zweiseitige Konfidenzintervall für die (theoretische) durchschnittliche Partikelkonzentration (Erwartungswert) ist: $(50.36310, 51.63690)$;
(d) die einseitigen Konfidenzintervalle für die (theoretische) durchschnittliche Partikelkonzentration (Erwartungswert) sind: $(-\infty, 51.53296)$; $(50.46704, \infty)$.

```
import numpy
from  scipy.stats import t
daten=12*[45]+10*[48]+30*[50]+20*[52]+28*[55]
m=numpy.mean(daten) # Erwartungswert
st=numpy.std(daten,ddof=1) #Standardabweichung
print("Erwartungswert:",m," Standardabweichung:",st)
n=len(daten)
print("zweiseitiges Konfidenzintervall fur Erwartungswert:")
a=0.05
links=m - (st/numpy.sqrt(n))*t.ppf(1-a/2,n-1)
rechts=m + (st/numpy.sqrt(n))*t.ppf(1-a/2,n-1)
print(f"({links:.5f}, {rechts:.5f})")
obereGrenze=m-(st/numpy.sqrt(n))*t.ppf(a,n-1)
print("einseitiges Konfidenzintervall fur Erwartungswert:")
print(f"({float('-inf')},{obereGrenze:.5f})")
untereGrenze=m-(st/numpy.sqrt(n))*t.ppf(1-a,n-1)
print("einseitiges Konfidenzintervall fur Erwartungswert:")
print(f"({untereGrenze:.5f}, {float('inf')})")
```

W5. In einer Urne befinden sich zwei rote Kugeln und fünf blaue Kugeln. Man zieht ohne Zurücklegen vier Kugeln.
(a) Wie groß ist die Wahrscheinlichkeit, dass man zwei blaue Kugeln (und somit zwei rote Kugel) zieht (in beliebiger Reihenfolge)?
(b) Sei Z die Zufallsgröße die anzeigt wie viele rote Kugeln entnommen wurden. Man schreibe die Wahrscheinlichkeitsverteilung von Z und berechne den Erwartungswert von Z.
Lsg.: (a) $\frac{C_2^2 C_4^2}{C_7^4} = \frac{6}{35}$
(b) $Z \sim \begin{pmatrix} 0 & 1 & 2 \\ \frac{5}{35} & \frac{20}{35} & \frac{10}{35} \end{pmatrix} \Longrightarrow E(Z) = \frac{0\cdot 5+1\cdot 20+2\cdot 10}{35} = \frac{40}{35} \approx 1.1428\,.$

[1]In der Europäischen Union liegt der Jahresgrenzwert für PM2.5 (Partikel mit einem Durchmesser von 2.5 Mikrometern oder kleiner) bei 25 Mikrogramm pro Kubikmeter (µg/m^3).

W6. Ein Programm erzeugt 0 und 1 mit gleicher Wahrscheinlichkeit. Eine zufällige 8-Bit-Zahl wird generiert. Mit welcher Wahrscheinlichkeit:
(a) enthält die Zahl genau 3 mal die 1?
(b) höchstens 3 mal die 1?
(c) ist die Zahl 10101010?
Lsg.: (a) $\frac{C_8^3}{2^8}$; (b) $\frac{C_8^0+C_8^1+C_8^2+C_8^3}{2^8}$; (c) $\frac{1}{2^8}$.

W7. Eine Tankstelle veranschlagt für einen Ölwechsel mindestens α Minuten. Die tatsächlich benötigte Zeit X ist von Kunde zu Kunde verschieden, wobei man davon ausgehen kann, dass $f : \mathbb{R} \to [0, \infty)$ definiert durch

$$f(x) = \begin{cases} e^{-(x-\alpha)} & : x > \alpha \\ 0 & : x \leq \alpha. \end{cases}$$

die Dichtefunktion von X ist. Mit Hilfe der Momentenmethode gebe man einen Schätzwert für α an, wenn folgende statistische Daten (Zeiten in Minuten) bekannt sind: 4.2, 3.1, 3.6, 4.5, 5.1, 3.8, 4.1, 5.2.
Lsg.: $E(X) = \int_\alpha^\infty xe^{-(x-\alpha)}dx = \alpha + 1 \Longrightarrow \alpha + 1 = \frac{1}{8}(4.2 + 3.1 + 3.6 + 4.5 + 5.1 + 3.8 + 4.1 + 5.2) = 4.2$
$\Longrightarrow \hat{\alpha} = 3.2$ ist der Schätzwert.