seit 1558

Friedrich-Schiller-Universität Jena

Wirtschaftswissenschaftliche Fakultät
Lehrstuhl für Empirische und Experimentelle Wirtschaftsforschung
Prof. Oliver Kirchkamp

BW24.1 - Basismodul Empirische und Experimentelle Wirtschaftsforschung
Aufgaben für den Arbeitsgruppenwettbewerb

Inhaltsverzeichnis

Bitte bearbeiten Sie alle Aufgabenblätter in einer Arbeitsgruppe. Geben Sie die Lösung auf der Homepage der Vorlesung `http://www.kirchkamp.de/bw241` unter „Arbeitsgruppenwettbewerb“ ein.

Sollte in diesem Arbeitsgruppenwettbewerb nach Zahlen gefragt werden, so ist das Endergebnis jeweils auf 3 Nachkommastellen gerundet anzugeben. Zwischenergebnisse sollten nicht gerundet werden.

Denken Sie daran, dass am Ende des Semesters für jede Arbeitsgruppe die Summe aus allen Wochen zählt – machen Sie also auf jeden Fall schon ab der ersten Woche mit.

Bitte installieren Sie auch zeitnah R auf Ihrem Rechner. Dafür gibt es zwar keine Punkte; Sie werden das Programm aber für den Wettbewerb benötigen. Fragen zu der Installation können Sie am besten im Forum stellen.
Sollten Sie keinen eigenen Computer mit R zur Verfügung haben, so können Sie R auch im Computerpool nutzen.

Am Ende der Aufgabenblätter ist eine Hilfestellung abgedruckt, die Sie so oder so ähnlich auch in der Klausur bekommen werden.

Wettbewerb Aufgabenblatt 1
Abgabe bis Dienstag, 25. Oktober, 12 Uhr

1. In einem Krankenhaus wird die Anzahl der neu eingelieferten Patienten X pro Monat erfasst. Folgende Realisationen für X liegen vor: (612, 507, 450, 495, 509, 533, 480, 531, 657, 569, 636, 634)

   1-i) Wie berechnet sich der empirische Mittelwert der Zufallsvariablen X in R?

   a) `mean(612, 507, 450, 495, 509, 533, 480, 531, 657, 569, 636, 634)`

   b) `1/12 * mean (612, 507, 450, 495, 509, 533, 480, 531, 657, 569, 636, 634)`

   c) `mean((c(612, 507, 450, 495, 509, 533, 480, 531, 657, 569, 636, 634)))`

   d) `1/12 * mean(c(612, 507, 450, 495, 509, 533, 480, 531, 657, 569, 636, 634))`

   e) Keine der obigen Antworten ist richtig.

   1-ii) Wie berechnet sich der Schätzer für die Standardabweichung der Zufallsvariablen X in R?

   a) `sd(612, 507, 450, 495, 509, 533, 480, 531, 657, 569, 636, 634)`

   b) `sqrt(sd(c(612, 507, 450, 495, 509, 533, 480, 531, 657, 569, 636, 634)))`

   c) `1/12*sd(612, 507, 450, 495, 509, 533, 480, 531, 657, 569, 636, 634)`

   d) `sd(c(612, 507, 450, 495, 509, 533, 480, 531, 657, 569, 636, 634))`

   e) Keine der obigen Antworten ist richtig.

2. Ein Arbeitnehmerverband möchte einen neuen Gesetzesvorschlag für die Pendlerpauschale entwickeln. Um sich einen Überblick über die Länge des Arbeitsweges der Arbeitnehmer zu verschaffen, werden einige Interviewer beauftragt, Stichproben zu erheben. Sie sollen zufällig ausgewählte Arbeitnehmer der Stadt Jena nach der Länge ihres Arbeitsweges befragen. Die Zufallsvariable X beschreibt die Länge der Strecke „Wohnung – Arbeitsplatz“ (gemessen in km). Die Interviewer haben die folgenden Stichproben $a, \ldots, g$ erhoben:

Stichprobe a:	37; 5; 17; 68; 75; 113; 1; 92; 12; 21
Stichprobe b:	95; 11; 40; 7; 20; 111; 43; 15; 51; 22
Stichprobe c:	120; 15; 30; 8; 20; 68; 18; 33; 72; 40
Stichprobe d:	10; 30; 93; 2; 25; 160; 25; 17; 19; 45
Stichprobe e:	9; 26; 36; 132; 22; 67; 34; 19; 95; 42
Stichprobe f:	17; 31; 145; 6; 90; 45; 88; 14; 11; 98
Stichprobe g:	12; 15; 70; 20; 89; 37; 4; 42; 150; 25

   2-i) Berechnen Sie die empirischen Mittelwerte der erhobenen Stichproben $a, \ldots, g$.

a) $\hat{\mu}_{x,a} = 40,1$; $\hat{\mu}_{x,b} = 41,5$; $\hat{\mu}_{x,c} = 42,4$; $\hat{\mu}_{x,d} = 42,6$; $\hat{\mu}_{x,e} = 48,2$; $\hat{\mu}_{x,f} = 54,5$; $\hat{\mu}_{x,g} = 46,6$

b) $\hat{\mu}_{x,a} = 44,1$; $\hat{\mu}_{x,b} = 41,5$; $\hat{\mu}_{x,c} = 35,3$; $\hat{\mu}_{x,d} = 42,6$; $\hat{\mu}_{x,e} = 48,2$; $\hat{\mu}_{x,f} = 54,5$; $\hat{\mu}_{x,g} = 46,4$

c) $\hat{\mu}_{x,a} = 44,5$; $\hat{\mu}_{x,b} = 41,5$; $\hat{\mu}_{x,c} = 42,4$; $\hat{\mu}_{x,d} = 42,6$; $\hat{\mu}_{x,e} = 48,2$; $\hat{\mu}_{x,f} = 54,5$; $\hat{\mu}_{x,g} = 46,4$

d) $\hat{\mu}_{x,a} = 44,1$; $\hat{\mu}_{x,b} = 41,5$; $\hat{\mu}_{x,c} = 42,4$; $\hat{\mu}_{x,d} = 42,6$; $\hat{\mu}_{x,e} = 48,2$; $\hat{\mu}_{x,f} = 54,5$; $\hat{\mu}_{x,g} = 51,6$

e) $\hat{\mu}_{x,a} = 44,1$; $\hat{\mu}_{x,b} = 41,5$; $\hat{\mu}_{x,c} = 42,4$; $\hat{\mu}_{x,d} = 42,6$; $\hat{\mu}_{x,e} = 48,2$; $\hat{\mu}_{x,f} = 54,5$; $\hat{\mu}_{x,g} = 46,4$

f) Keine der obigen Antworten ist richtig.

2-ii) Berechnen Sie die Stichprobenvarianzen der Stichproben a, c, e und g. (Hinweis: Die in den Lösungsmöglichkeiten angegebenen Werte für die Stichprobenvarianzen sind gerundet.)

a) $\hat{\sigma}^2_{x,a} = 1401,37$; $\hat{\sigma}^2_{x,c} = 1083,24$; $\hat{\sigma}^2_{x,e} = 1352,35$; $\hat{\sigma}^2_{x,g} = 1833,44$

b) $\hat{\sigma}^2_{x,a} = 1589,21$; $\hat{\sigma}^2_{x,c} = 1203,6$; $\hat{\sigma}^2_{x,e} = 1502,62$; $\hat{\sigma}^2_{x,g} = 2037,16$

c) $\hat{\sigma}^2_{x,a} = 39,86$; $\hat{\sigma}^2_{x,c} = 34,69$; $\hat{\sigma}^2_{x,e} = 38,76$; $\hat{\sigma}^2_{x,g} = 45,13$

d) $\hat{\sigma}^2_{x,a} = 37,43$; $\hat{\sigma}^2_{x,c} = 32,91$; $\hat{\sigma}^2_{x,e} = 36,77$; $\hat{\sigma}^2_{x,g} = 42,82$

e) Keine der obigen Antworten ist richtig.

2-iii) Lassen Sie sich die empirischen Mittelwerte der Stichproben $a, \ldots, g$ mittels R in einem Plot ausgeben ohne diese dabei selbst handschriftlich zu berechnen. Die Variablen für die Stichproben $a, \ldots, g$ wurden bereits in R definiert:

```
a=c(37,5,17,68,75,113,1,92,12,21)
b=c(95,11,40,7,20,111,43,15,51,22)
c=c(120,15,30,8,20,68,18,33,72,40)
d=c(10,30,93,2,25,160,25,17,19,45)
e=c(9,26,36,132,22,67,34,19,95,42)
f=c(17,31,145,6,90,45,88,14,11,98)
g=c(12,15,70,20,89,37,4,42,150,25)
```

Welche der unten angegebenen Lösungsmöglichkeiten enthält die Kommandos, die Sie zur Lösung dieser Teilaufgabe in R korrekt eingeben müssen?

a)
```
v=(mean(a), mean(b), mean(c), mean(d), mean(e), mean(f), mean(g))
plot(v, xlab="Stichproben", ylab="Mittelwerte der Stichproben")
```

b)
```
v=1/7*c(mean(a), mean(b), mean(c), mean(d), mean(e), mean(f), mean(g))
plot(v, xlab="Stichproben", ylab="Mittelwerte der Stichproben")
```

c)
```
v=c(sum(a), sum(b), sum(c), sum(d), sum(e), sum(f), sum(g))
plot(v, xlab="Stichproben", ylab="Mittelwerte der Stichproben")
```

d) `v=c(mean(a), mean(b), mean(c), mean(d), mean(e), mean(f), mean(g))`
`plot(v, xlab="Stichproben", ylab="Mittelwerte der Stichproben")`

e) Keine der obigen Antworten ist richtig.

Wettbewerb Aufgabenblatt 2
Abgabe bis Dienstag, 1. November, 12 Uhr

1.

Eine Fast-Food-Kette besitzt 100 Filialen in Deutschland. $\bar{x}_1, ..., \bar{x}_{100}$ seien die durchschnittlichen monatlichen Umsätze der einzelnen Filialen im Jahr 2008 (die Umsätze der einzelnen Filialen sind unabhängig voneinander). X sei der monatliche Umsatz einer Filiale. Es gilt: $\mathrm{E}(X) = \theta$ und $\mathrm{var}(X) = \sigma^2$.

1-i) Welche Schätzfunktionen sind erwartungstreu zum Schätzen von θ (mehrere richtige Antworten möglich)?

a) $g_1 = 10 \cdot \bar{x}_1 + 90 \cdot \bar{x}_{99}$

b) $g_2 = \frac{10\bar{x}_1 + 90\bar{x}_{99}}{100}$

c) $g_3 = \frac{1}{100} \sum_{i=1}^{100} \bar{x}_i$

d) $g_4 = \frac{1}{2}\bar{x}_5 + \frac{1}{3}\bar{x}_{20} + \frac{1}{6}\bar{x}_{37}$

e) $g_5 = \sum_{i=1}^{100} \bar{x}_i$

f) Keine der obigen Antworten ist richtig.

1-ii) Welche der Schätzfunktionen ist am wirksamsten?

a) $g_1 = 10 \cdot \bar{x}_1 + 90 \cdot \bar{x}_{99}$

b) $g_2 = \frac{10\bar{x}_1 + 90\bar{x}_{99}}{100}$

c) $g_3 = \frac{1}{100} \sum_{i=1}^{100} \bar{x}_i$

d) $g_4 = \frac{1}{2}\bar{x}_5 + \frac{1}{3}\bar{x}_{20} + \frac{1}{6}\bar{x}_{37}$

e) $g_5 = \sum_{i=1}^{100} \bar{x}_i$.

2.

Gegeben sei die Stichprobe $(X_1, ..., X_i, ..., X_{40})$ mit unabhängigen und identisch verteilten X_i. Die Varianz σ^2 sei bekannt und es soll der unbekannte Erwartungswert θ geschätzt werden. Weiterhin sei $\overline{X}_1$ das arithmetische Mittel der ungeraden i und $\overline{X}_2$ das arithmetische Mittel der geraden i. n bezeichnet die Anzahl der Beobachtungen.

2-i) Welche der folgenden Schätzfunktionen sind erwartungstreu zum Schätzen von θ (mehrere richtige Antworten möglich)?

a) $g_1 = \overline{X}_1$

b) $g_2 = \frac{3}{4} \cdot (\overline{X}_1 + \overline{X}_2)$

c) $g_3 = \left(\frac{n}{4} \cdot \overline{X}_2 - 8 \cdot \overline{X}_1\right) \cdot \frac{1}{2}$

d) $g_4 = \frac{\left(\frac{1}{3}\cdot\overline{X}_1 + \frac{2}{3}\overline{X}_2\right)}{2\cdot n}$

e) Keine der obigen Antworten ist richtig.

2-ii) Berechnen Sie den mittleren quadratischen Fehler (MSE) von allen erwartungstreuen Schätzfunktionen und wählen Sie die richtigen Antworten aus (mehrere richtige Antworten möglich):

a) Die Schätzfunktion $g_5 = \frac{1}{2} \cdot \left(\overline{X}_1 + \overline{X}_2\right)$ ist erwartungstreu und wirksamer als die erwartungstreuen Schätzfunktionen aus 2-i.

b) Kein mittlerer quadratischer Fehler der erwartungstreuen Schätzfunktionen ist größer als $0.4 \cdot \sigma^2$.

c) Die Schätzfunktion g_1 ist nicht besonders wirksam weil nur die Hälfte der Informationen genutzt werden.

d) Keine der obigen Antworten ist richtig.

3.

Ein neu entwickeltes Futtermittel soll die Milchproduktion von Kühen anregen. Die Milchproduktion einer Kuh messen wir als X.
Es gilt $E(X) = \theta$ und $Var(X) = \sigma^2$. Sie testen das Futtermittel an Kühen in vier (identischen) Milchhöfen. Die folgende Tabelle zeigt die Anzahl der Kühe und die durchschnittliche Milchproduktion pro Kuh in diesen vier Milchhöfen:

	Anzahl der Kühe	durchschnittliche Milchproduktion
Milchhof 1	10	$\overline{X}_1$
Milchhof 2	20	$\overline{X}_2$
Milchhof 3	30	$\overline{X}_3$
Milchhof 4	40	$\overline{X}_4$

(Die $\overline{X}_1$, $\overline{X}_2$, $\overline{X}_3$, $\overline{X}_4$ sind Mittelwerte von jeweils unabhängig und identischen verteilten X).

3-i) Welche der folgenden Funktionen sind erwartungstreue Schätzer für den Erwartungswert der Milchproduktion pro Kuh E(X) (mehrere richtige Antworten möglich)?

a) $g_1 = 0.7 \cdot \overline{X}_2 + 0.3 \cdot \overline{X}_4$

b) $g_2 = 0.3 \cdot \overline{X}_1 + 0.9 \cdot \overline{X}_3$

c) $g_3 = 0.25 \cdot \overline{X}_1 + 0.25 \cdot \overline{X}_2 + 0.5 \cdot \overline{X}_4$

d) $g_4 = \overline{X}_1 + \overline{X}_2 + \overline{X}_3 + \overline{X}_4$

e) $g_5 = 0.1 \cdot \overline{X}_1 + 0.2 \cdot \overline{X}_2 + 0.3 \cdot \overline{X}_3 + 0.4 \cdot \overline{X}_4$

f) $g_6 = \overline{X}_4$

g) Keine der obigen Antworten ist richtig.

3-ii) Welche der folgenden Funktionen sind erwartungstreue Schätzer für den Erwartungswert der Milchproduktion pro Kuh E(X) <u>und</u> sind entweder am wirksamsten oder haben eine Varianz von $\frac{1}{64}\sigma^2$ (mehrere richtige Antworten möglich)?

a) $g_1 = 0.7 \cdot \overline{X}_2 + 0.3 \cdot \overline{X}_4$

b) $g_2 = 0.3 \cdot \overline{X}_1 + 0.9 \cdot \overline{X}_3$

c) $g_3 = 0.25 \cdot \overline{X}_1 + 0.25 \cdot \overline{X}_2 + 0.5 \cdot \overline{X}_4$

d) $g_4 = \overline{X}_1 + \overline{X}_2 + \overline{X}_3 + \overline{X}_4$

e) $g_5 = 0.1 \cdot \overline{X}_1 + 0.2 \cdot \overline{X}_2 + 0.3 \cdot \overline{X}_3 + 0.4 \cdot \overline{X}_4$

f) $g_6 = \overline{X}_4$

g) Keine der obigen Antworten ist richtig.

4.

Ein Automobilkonzern ist in zwei Werke aufgeteilt. In Werk I arbeiten 6000 Beschäftigte (= Grundgesamtheit G_1) und in Werk II arbeiten 4000 Beschäftigte (= Grundgesamtheit G_2). Die Vielzahl von neuen Auftragseingängen aufgrund der Abwrackprämie will die Firmenleitung mit einer neuen Arbeitszeitregelung bewältigen. Der Betriebsrat will die Anteile p_1, p_2 bzw. p der Befürworter der vorgeschlagenen neuen Arbeitszeitregelung in G_1, G_2 bzw. in der Gesamtbelegschaft $G = G_1 \bigcup G_2$ schätzen.
Dazu wird in G_1 eine einfache Stichprobe vom Umfang n_1 und in G_2 eine einfache Stichprobe vom Umfang n_2 gezogen, in der jeder Befragte die neue Regelung befürworten kann (= Ergebnis 1) oder nicht (= Ergebnis 0). Es sei $\overline{X}_1$ bzw. $\overline{X}_2$ der zufallsabhängige Anteil der Befürworter in der Stichprobe aus G_1 bzw. G_2.

4-i) Welche der folgenden Funktionen sind für beliebige n_1 und n_2 erwartungstreue Schätzer für p, wobei gilt: $p = \frac{6000 \cdot p_1 + 4000 \cdot p_2}{10000}$?

a) $\frac{1}{n_1+n_2} \cdot (n_1 \cdot \overline{X}_1 + n_2 \cdot \overline{X}_2)$

b) $\frac{1}{2} \cdot (\overline{X}_1 + \overline{X}_2)$

c) $\frac{1}{10000} \cdot (6000 \cdot \overline{X}_1 + 4000 \cdot \overline{X}_2)$

d) Keine der obigen Antworten ist richtig.

4-ii) Von $n_1 = 100$ Befragten aus G_1 waren 40 und von $n_2 = 50$ Befragten aus G_2 waren 30 für die neue Regelung. Effiziente Schätzer für p_1, p_2 und p sind:

a) $p_1 = 0.4, p_2 = 0.6$ und $p = 0.50$

b) $p_1 = 0.4, p_2 = 0.6$ und $p = 0.48$

c) $p_1 = 0.6, p_2 = 0.4$ und $p = 0.50$

d) Keine der obigen Antworten ist richtig.

Wettbewerb Aufgabenblatt 3
Abgabe bis Dienstag, 8. November, 12 Uhr

1.

Gegeben sei ein gezinkter Würfel. Es gilt:

$$P(X=2) = P(X=3) = P(X=4) = P(X=5) = \frac{1}{6}$$

$$P(X=1) = \frac{1}{6} - \theta \qquad P(X=6) = \frac{1}{6} + \theta \qquad \theta \in \left[0; \frac{1}{6}\right]$$

1-i) Welche der folgenden Aussagen ist zutreffend:

a) Für $\theta = 0$ handelt es sich um einen fairen Würfel.

b) Für $\theta = \frac{1}{6}$ handelt es sich um einen fairen Würfel.

c) Keine der obigen Antworten ist richtig.

1-ii) Bei einer einfachen Stichprobe wird folgendes Resultat erzielt:

$$x = (6, 4, 2, 1, 2, 3, 6, 3, 4, 5)$$

Welcher Wert für den unbekannten Parameter θ ist für die gegeben Beobachtungen richtig? Benutzen Sie zur Bestimmung des Parameters die ML-Methode.

a) $\hat{\theta}$=$\frac{1}{18}$

b) $\hat{\theta}$=$\frac{1}{19}$

c) $\hat{\theta}$=$\frac{1}{17}$

d) $\hat{\theta}$=$\frac{1}{20}$

e) Keine der obigen Antworten ist richtig.

2.

Gegeben sei eine poissonverteilte Zufallsvariable $X \sim Po(\lambda)$ und eine Stichprobe $(x_1, ..., x_n)$.

2-i) Schätzen Sie mit Hilfe der ML-Methode den Parameter λ und wählen Sie die richtige Lösung aus.

a) $\widehat{\lambda} = \ln \overline{x}$

b) $\widehat{\lambda} = \overline{x}$

c) $\widehat{\lambda} = \frac{1}{\overline{x}}$

d) $\widehat{\lambda} = \frac{1}{\ln \overline{x}}$

e) Keine der obigen Antworten ist richtig.

2-ii) Welche der folgenden Aussagen sind korrekt (mehrere richtige Antworten möglich)?

a) $L(x_1, ..., x_n|\theta)$ ist ein Maß für die Wahrscheinlichkeit, $(x_1, ..., x_n)$ zu beobachten, wenn θ vorliegt.

b) Das Maximum von L bei gegebenen $x_1, ..., x_n$ in Abhängigkeit von θ liefert die ML-Schätzung von θ.

c) Um die Informationen zu verdichten und eine genauere Schätzung zu erhalten wendet man die Logarithmus-Funktion an.

d) Das Ziel ist θ so zu bestimmen, dass der Stichprobenrealisation $x_1, ..., x_n$ die größte Likelihood zukommt.

e) Keine der obigen Antworten ist richtig.

3.

Gegeben sei die Stichprobe einer Poisson-verteilten Zufallsvariable:

x	0	1	2	3
Häufigkeit	64	27	9	2

Bestimmen Sie den ML-Schätzer für λ!

4.

Eine Zufallsvariable X sei exponentialverteilt mit $X \sim Ex(\lambda)$. Es liegen folgende Beobachtungen vor: $20, 11, 19, 31, 26, 23, 11, 9, 32, 28$.

4-i) Bestimmen Sie mit Hilfe der Beobachtungen den Momentenschätzer θ für λ auf Basis des ersten Moments. Runden Sie dabei auf drei Stellen nach dem Komma.

4-ii) Der Momentschätzer $\hat{\lambda}$ soll nun mit Hilfe des zweiten Moments geschätzt werden. Mit welchem Befehl in R kann man dies tun, wenn die Beobachtungen bereits in der Variablen x gespeichert sind (mehrere richtige Antworten möglich)?

a) `1/sd(x)`

b) `1/sqrt(var(x))`

c) `var(x)`

d) `sd(x)`

e) `sqrt(1/var(x))`

f) Keine der obigen Antworten ist richtig.

Wettbewerb Aufgabenblatt 4
Abgabe bis Dienstag, 15. November, 12 Uhr

1.

In einer Mühle wird Getreide gemahlen und das Mehl in Tüten verpackt. Das Gewicht einer Tüte Mehl kann dabei als normalverteilt angenommen werden. Die Varianz des Gewichts ist aus langjähriger Erfahrung bekannt und beträgt 2500 Gramm^2. Eine einfache Stichprobe vom Umfang $n = 25$ ergibt ein Gesamtgewicht von 26.000 Gramm.

1-i) Wie groß ist das 95%-Konfidenzintervall für den Mittelwert des Gewichts der Mehltüten?

a) $[950; 1130]$

b) $[942; 1138]$

c) $[927; 1063]$

d) $[899; 1067]$

e) $[998; 1214]$

f) Keine der obigen Antworten ist richtig.

1-ii) Die Länge des Konfidenzintervalls soll nun durch eine größere Stichprobe verkürzt werden. Wie ist n zu wählen, damit das Konfidenzintervall maximal eine Länge von 100 hat?

a) $n \leq 250$

b) $n \geq 103$

c) $n \geq 97$

d) $n \geq 56$

e) $n \leq 93$

f) Keine der obigen Antworten ist richtig.

2.

Eine Maschine wickelt Klopapier auf Rollen, auf denen es verkauft wird. Die Länge des Papiers, das auf eine Rolle gewickelt wird ist mit Mittlewert μ und Varianz 0,36 m^2 normalverteilt. Sie stellen fest, dass die Maschine auf die letzten 25 Rollen insgesamt 995 m Klopapier aufgewickelt hat.

2-i) Wie lautet die beste erwartungstreue Schätzung für den Erwartungswert μ der Normalverteilung?

a) $0.36 \cdot \sqrt{995}$

b) 995/25

c) $\frac{995}{25 \cdot 0.36}$

d) $\frac{995}{25 \cdot \sqrt{0.36}}$

e) $\frac{995}{\sqrt{25 \cdot 0.36}}$

f) 25/995

2-ii) Nehmen Sie an, die Variable `muhut` enthält den oben geschätzten Erwartungswert von μ. Wir bestimmen Unter- und Obergrenze des 99%-Konfidenzintervall für μ mit dem Kommando
`c (muhut + qnorm(A) * B , muhut - qnorm(A) * B )`
Welche Werte setzen wir für `A` ein (mehrere richtige Antworten möglich)?

a) `A=.005`

b) `A=.01`

c) `A=.99`

d) `A=.995`

2-iii) Welche Werte setzen wir für `B` ein (mehrere richtige Antworten möglich)?

a) `B=.36`

b) `B=sqrt(36)`

c) `B=.36/25`

d) `B=.36/sqrt(25)`

e) `B=sqrt(.36/25)`

f) `B=sqrt(.36)/25`

2-iv) Wie viele Rollen müssten wir abwickeln und exakt nachmessen, um ein 99%-Konfidenzintervall für μ aufstellen zu können, das nicht breiter als 0,1 m ist (mehrere richtige Antworten möglich)

a) `2*10*qnorm(0.995)*0.36`

b) `2*10*qnorm(0.995)*0.6`

c) `(2*10*qnorm(0.995)*0.6)^2`

d) `(2*10*qnorm(0.995)*0.36)^2`

e) `(10*qnorm(0.995)*0.6)^2`

f) `(2*10*qnorm(0.99)*0.6)^2`

g) `(2*10*qnorm(0.99)*0.36)^2`

3.

3-i) Die Lebensdauer von Batterien ist normalverteilt. Ein physikalisches Forschungsinstitut hat im Auftrag eines großen Batterieherstellers herausgefunden, dass die Standardabweichung der Lebensdauer einer Batterie 30 (Stunden) beträgt. Die Messung der Lebensdauer von 100 der Produktion zufällig entnommenen Batterien ergab eine Gesamtleistungsdauer von 6935,75 Stunden. Berechnen Sie das 90% Konfidenzintervall für den Mittelwert der Lebensdauer. *(Bitte benutzen Sie die Werte aus der Verteilungstabelle im Anhang und runden Sie erst am Ende.)*

a) $[64.4; 74.3]$

b) $[65.5; 73.2]$

c) $[68,5; 70,3]$

d) $[0; 217.9]$

e) Keine der obigen Antworten ist richtig.

3-ii) Die Geschäftsleitung des Batterieherstellers möchte, dass die Länge des Konfidenzintervalls höchstens 5 (Stunden) beträgt. Wie groß muss die Stichprobe gewählt werden?

a) mindestens 99

b) mindestens 594

c) mindestens 390

d) Keine der obigen Antworten ist richtig.

3-iii) Da der Batteriehersteller sich nicht nur auf eine Meinung verlassen möchte, wurde ebenfalls ein anderes physikalisches Forschungsinstitut damit beauftragt, ein 90%-Konfidenzintervall für den Erwartungswert der Lebensdauer von Batterien anzugeben. Auch sie gehen davon aus, dass die Lebensdauer von Batterien normalverteilt ist. Da dieses Forschungsinstitut technisch nicht so modern ausgerüstet ist, können Sie die Varianz nicht ermitteln. Das Institut hat eine Stichprobe vom Umfang 20 gezogen und die Lebensdauer der Batterien (in Stunden) gemessen: (60,5; 80; 71; 73,7; 65; 68; 64,4; 62,9; 74; 78; 72,9; 74; 67,5; 72,8; 61,9; 71; 58; 61; 72,8; 73). Bestimmen Sie das 90% Konfidenzintervall.

a) $[54,3233; 83,9167]$

b) $[66,7338; 71,5062]$

c) $[66,8377; 71,4023]$

d) Keine der obigen Antworten ist richtig.

Wettbewerb Aufgabenblatt 5
Abgabe bis Dienstag, 22. November, 12 Uhr

1.

Um die Unfallzahlen zu senken, wurde ein Gesetz erlassen, dass Fahren mit Abblendlicht am Tag vorschreibt. Ein Automobilclub möchte die Wirksamkeit der Maßnahme testen und vermutet, dass die Unfallzahlen im Durchschnitt gesunken sind.
Die Variable X enthält die Zahl der Unfälle im Monat und die Variable $\mu_{x,0}$ enthält den langjährigen Mittelwert aus der Vergangenheit (in der gleichen Einheit).
Stellen Sie die Null- und Alternativhypothese auf.

a) $H_0 : \mathrm{E}(X) = \mu_{x,0}, H_1 : \mathrm{E}(X) \neq \mu_{x,0}$

b) $H_0 : \mathrm{E}(X) \neq \mu_{x,0}, H_1 : \mathrm{E}(X) = \mu_{x,0}$

c) $H_0 : \mathrm{E}(X) = \mu_{x,0}, H_1 : \mathrm{E}(X) > \mu_{x,0}$

d) $H_0 : \mathrm{E}(X) = \mu_{x,0}, H_1 : \mathrm{E}(X) \geq \mu_{x,0}$

e) $H_0 : \mathrm{E}(X) = \mu_{x,0}, H_1 : \mathrm{E}(X) < \mu_{x,0}$

f) $H_0 : \mathrm{E}(X) = \mu_{x,0}, H_1 : \mathrm{E}(X) \leq \mu_{x,0}$

g) Keine der obigen Antworten ist richtig.

2.

Bei einem Signifikanztest wurde die Nullhypothese auf dem 1% Signifikanzniveau verworfen. Welche der folgenden Aussagen sind richtig (mehrere richtige Antworten möglich)?

a) Die Nullhypothese ist nachweislich eindeutig falsch.

b) Die Alternativhypothese ist nachweislich eindeutig wahr.

c) Mit einer Wahrscheinlichkeit von 99% gilt H_1.

d) Die Nullhypothese kann man mit einer Wahrscheinlichkeit von 1% verwerfen.

e) Keine der obigen Antworten ist richtig.

3.

Ein Bauunternehmer rühmt sich mit der Behauptung den bisherigen Energiebedarf seiner Einfamilienhäuser von bislang durchschnittlich 15 kWh pro Monat durch neue Dämmtechniken signifikant gesenkt zu haben. Dies untermauert er durch eine Stichprobe von $n = 16$ Häusers, die im letzten Monat einen durchschnittlichen Verbrauch von $13,5$ kWh ergeben hat. Der Energieverbrauch pro Haus und pro Monat sei unabhängig und normalverteilt mit einer Varianz von $6,25$ kWh^2.

3-i) Wie muss der Bauunternehmer die Nullhypothese und die Alternativhypothese wählen um seine Behauptung zu bestätigen ?

a) $H_0 : \theta = 15, H_1 : \theta \neq 15$

b) $H_0 : \theta < 15, H_1 : \theta \geq 15$

c) $H_0 : \theta \geq 15, H_1 : \theta < 15$

d) $H_0 : \theta \neq 15, H_1 : \theta = 15$

e) Keine der obigen Antworten ist richtig.

3-ii) Welche der Antworten sind zu einem Signifikanzniveau von $\alpha = 5\%$ richtig (mehrere richtige Antworten möglich)?

a) H_0 muss abgelehnt werden.

b) Hausbesitzer die den Eindruck haben, dass sie sogar mehr Energie verbrauchen müssten folgende Hypothesen wählen: $H_0 : \theta = 15, H_1 : \theta \neq 15$

c) Bevor die Stichprobe gezogen wird, überlegt sich der Bauunternehmer, dass H_0 umso eher angenommen wird, je größer die Standardabweichung des Energieverbrauches ist.

d) H_0 kann angenommen werden.

e) α ist die Wahrscheinlichkeit für die Ablehnung einer richtigen Nullhypothese.

f) Keine der obigen Antworten ist richtig.

4.

Sie absolvieren ein Praktikum in einem Kosmetikunternehmen und überprüfen heute die Abfüllmaschine für Duschgel. Aus einer Stichprobe von 15 Duschgelflaschen bestimmen Sie eine mittlere Füllmenge von 250 ml bei einer Varianz von 24. Sie nehmen an, dass die Füllmenge normalverteilt ist.

4-i) Sie erzählen Ihrem Chef von ihrer Überprüfung und er ist beunruhigt, weil der Hersteller der Maschine hat Ihm versprochen, die Varianz der Füllmenge sei langfristig 14. Ihr Chef beschwert sich beim Hersteller der Abfüllmaschine. Dessen Servicehotline versichert Ihm, dass die von Ihnen gemessene Abweichung zufällig sei und langsfristig auch Ihre Maschine eine Varianz von nicht größer als 14 haben würde. Welche Hypothesen stellen Sie richtigerweise auf, um die Aussage der Servicehotline zu überprüfen?

a) $H_0 : \sigma_X^2 = 14, H_1 : \sigma_X^2 \neq 14$

b) $H_0 : \sigma_X^2 \neq 14, H_1 : \sigma_X^2 = 14$

c) $H_0 : \sigma_X^2 = 14, H_1 : \sigma_X^2 > 14$

d) $H_0 : \sigma_X^2 = 24, H_1 : \sigma_X^2 \geq 14$

e) $H_0 : \sigma_X^2 = 14, H_1 : \sigma_X^2 < 14$

f) $H_0 : \sigma_X^2 = 24, H_1 : \sigma_X^2 \leq 14$

g) Keine der obigen Antworten ist richtig.

4-ii) Testen Sie die Hypothese mit einem einseitigen Test bei einem Signifikanzniveau von 5%. Beachten Sie, dass wir hier eine Varianz testen. In der Vorlesung hatten wir zum Test von Mittelwerten den Zusammenhang $(\bar{x} - \mu_0)/\sigma_{\bar{x}} \sim N(0,1)$ benutzt. Zum Testen von Varianzen verwenden Sie hier die Testfunktion: $(n-1) \cdot \hat{\sigma}^2/\sigma_0^2 \sim \chi^2_{n-1}$. Die Verteilung und die Quantile der χ^2-Verteilung bestimmen Sie in R mit `pchisq` und `qchisq`. Die χ^2-Verteilung hängt, genauso wie die t-Verteilung, auch von Freiheitsgraden ab.

Was ist richtig (mehrere richtige Antworten möglich)?

a) Die Testfunktion ergibt für diese Stichprobe einen Wert von 24.

b) Die Testfunktion ergibt für diese Stichprobe einen Wert von 14.

c) Die Testfunktion ergibt für diese Stichprobe einen Wert von 2.

d) Der Ablehnungsbereich für die Testfunktion reicht von $-\infty$ bis $23,69$.

e) Der Ablehnungsbereich für die Testfunktion reicht von $23,69$ bis ∞.

f) Die Nullhypothese kann abgelehnt werden.

g) Keine der obigen Antworten ist richtig.

4-iii) Der Hersteller Ihrer Maschine überprüft die gesamte Angelegenheit und geht nun von einer langfristigen Varianz von 25 aus. Wie groß muss Ihre Stichprobe sein, damit das 95%-Konfidenzintervall für die mittlere Füllmenge eine Breite von genau 1.96 ml hat? (verwenden Sie für diese Aufgabe die oben angegebenen Quantile)

a) $n = 10$

b) $n = 25$

c) $n = 100$

d) $n = 250$

e) $n = 1000$

f) $n = 2500$

g) Keine der obigen Antworten ist richtig.

5.

In einer Stadt werden regelmäßig Niederschlagsmessungen durchgeführt. Das Jahr 2008 war recht kalt und es hat oft geregnet. Meteorologen vermuten daher, dass im Durchschnitt höhere Niederschlagsmengen in 2008 als in den vorhergegangenen Jahren gefallen sind.

Die Variable X enthält den jährlichen Niederschlag und die Variable $\mu_{x,0}$ enthält den langjährigen Mittelwert aus der Vergangenheit (in der gleichen Einheit).

Stellen Sie die Null- und Alternativhypothese auf.

a) $H_0 : \mathrm{E}(X) = \mu_{x,0}, H_1 : \mathrm{E}(X) \neq \mu_{x,0}$

b) $H_0 : \mathrm{E}(X) \neq \mu_{x,0}, H_1 : \mathrm{E}(X) = \mu_{x,0}$

c) $H_0 : \mathrm{E}(X) = \mu_{x,0}, H_1 : \mathrm{E}(X) > \mu_{x,0}$

d) $H_0 : \mathrm{E}(X) = \mu_{x,0}, H_1 : \mathrm{E}(X) \geq \mu_{x,0}$

e) $H_0 : \mathrm{E}(X) = \mu_{x,0}, H_1 : \mathrm{E}(X) < \mu_{x,0}$

f) $H_0 : \mathrm{E}(X) = \mu_{x,0}, H_1 : \mathrm{E}(X) \leq \mu_{x,0}$

g) Keine der obigen Antworten ist richtig.

6.

Zur Herstellung von Kaffeepulver werden die Bohnen in einer Mühle gemahlen und anschließend in eine Tüte gefüllt. Das Gewicht X der Tüte ist normalverteilt und hat den Erwartungswert θ. Aus der hundertjährigen Tradition des Unternehmens sei bekannt, dass die Varianz 256 Gramm^2 beträgt. Eine einfache Stichprobe vom Umfang $n = 25$ hat ein Gesamtgewicht von 12400 Gramm ergeben.
Können Sie zum Niveau $\alpha = 0,05$ nachweisen, dass θ signifikant von 501 Gramm abweicht?

a) Mit den gegebenen Daten kann die Aufgabe nicht gelöst werden.

b) Ja, θ weicht signifikant von 501 Gramm ab.

c) Nein, θ weicht nicht signifikant von 501 Gramm ab.

d) Keine der obigen Antworten ist richtig.

7.

Sie arbeiten für einen Getränkehersteller, bei dem verschiedene Getränke abgefüllt werden. Nachdem Sie das Ganze einige Zeit beobachtet haben, finden Sie, dass die Flaschen einer Sorte leerer aussehen als andere und vermuten daher, dass die Soll-Füllmenge von $1000ml$ nicht eingehalten wird. Dies möchten Sie nun anhand eines Tests überprüfen.

7-i) Wie lautet die Null- und die Alternativhypothese, wenn Sie beweisen wollen, dass die Füllmenge zu klein ist?

a) $H_0 : E(X) = 1000, H_1 : E(X) > 1000$

b) $H_0 : E(X) < 1000, H_1 : E(X) > 1000$

c) $H_0 : E(X) = 1000, H_1 : E(X) < 1000$

d) Keine der obigen Antworten ist richtig.

7-ii) Sie entnehmen nun eine Stichprobe vom Umfang $n = 25$. Dabei ergab sich eine mittlere Füllmenge von $997ml$. Außerdem ist bekannt, dass die Varianz beim Abfüllen (σ_x^2) $400ml^2$ beträgt. Welche der folgenden Aussagen sind bei einem Signifikanzniveau von 5% richtig (mehrere richtige Antworten möglich)?

a) H_0 wird abgelehnt.

b) H_0 wird angenommen.

c) Wenn H_0 zutrifft, ist die Wahrscheinlichkeit, einen Fehler 1. Art zu machen 5%.

d) Die Wahrscheinlichkeit, einen Fehler 2. Art zu machen, beträgt 5%

e) Keine der obigen Antworten ist richtig.

Wettbewerb Aufgabenblatt 6
Abgabe bis Dienstag, 29. November, 12 Uhr

1.

Kürzlich haben Sie in einer Zeitung gelesen, dass Studenten in Deutschland durchschnittlich 105 Euro im Monat für Freizeitaktivitäten ausgeben. Sie interessieren sich nun dafür, ob das auch auf die Studenten in Jena zutrifft und führen eine Umfrage mit 100 Studenten durch. Sie nehmen an, dass die Ausgaben für Freizeitaktivitäten normalverteilt sind. Das Signifikanzniveau beträgt 5%.

1-i) Wie lautet der richtige R-Befehl, für den Test den Sie ausführen, wenn Sie testen wollen, ob Studenten in Jena weniger für Freizeitaktivitäten ausgeben, als der Bundesdurchschnitt? Der Vektor x enthält dabei Ihre Stichprobe.

a) `t.test(x, alternative = "two.sided", mu = 105)`

b) `t.test(x, alternative = "less", mu = 105)`

c) `t.test(x, alternative = "less", mu = 100)`

d) Keine der obigen Antworten ist richtig.

1-ii) Sie führen den Test nun durch und erhalten das folgende Ergebnis:

```
        One Sample t-test

data:  x
t = -2.1082, df = 99, p-value = 0.01877
alternative hypothesis: true mean is less than 105
95 percent confidence interval:
     -Inf 102.7933
sample estimates:
mean of x
 94.61176
```

Welche der folgenden Aussagen sind korrekt (mehrere richtige Antworten möglich)?

a) Zu einem Signifikanzniveau von 5% wird die Nullhypothese, dass die Studenten in Jena mehr ausgeben, abgelehnt.

b) Zu einem Signifikanzniveau von 5% wird die Nullhypothese, dass die Studenten in Jena mehr ausgeben, angenommen.

c) Der Wert der Teststatistik beträgt 0,01877.

d) Der Wert der Teststatistik beträgt -2,1082.

e) Der Test umfasst nur 99 Beobachtungen aus der Stichprobe.

f) Keine der obigen Antworten ist richtig.

2.

Kennzeichnen Sie die wahren Aussagen (mehrere richtige Antworten möglich)!

a) Mit einem t-Test kann man testen ob, ob der Mittelwert einer Grundgesamtheit gleich einem bestimmten Wert ist.

b) t Die Teststatistik $g = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ gilt sowohl für einseitige als auch für zweiseitige Tests.

c) Je kleiner das Signifikanzniveau α, um so eher wird ceteris paribus H_0 angenommen.

d) Der Fehler, der entsteht wenn H_0 angenommen wird obwohl H_0 falsch ist, wird durch ein vorgegebenes Signifikanzniveau α beschränkt.

e) Bei einem zweiseitigen Test muss eine zweidimensionale Stichprobe vorliegen.

f) Je größer die Varianz der Grundgesamtheit, desto unwahrscheinlicher ist ceteris paribus die Ablehnung einer Nullhypothese.

g) Keine der obigen Antworten ist richtig.

3.

Eine Verbraucherorganisation möchte prüfen, ob die Gewichtsangaben von 1 kg auf Lebensmittelverpackungen von Zucker der Wahrheit entsprechen. Dazu kaufen sie in einem willkürlich ausgewählten Supermarkt 20 Päckchen Zucker und führen anschließend in einem Labor die Kontrollen durch. Die Gewichtsmessungen (in g) ergaben: 987; 1002; 993; 999; 981; 1009; 1013; 995; 1002; 1001; 998; 997; 994; 1005; 1007; 985; 995; 998; 1014; 1003.

Es wird angenommen, dass die Füllmengen normalverteilt sind. Die Verbraucherorganisation möchte nun prüfen, ob der Mittelwert des Gewichts der Zuckerpackungen 1 kg entspricht. Das Signifikanzniveau beträgt 5%.

3-i) Wie lauten die Null- und die Alternativhypothese, die die Verbraucherorganisation aufstellen muss?

a) $H_0 : E(X) \leq 1000 \quad ; \quad H_1 : E(X) > 1000$

b) $H_0 : E(X) \geq 1000 \quad ; \quad H_1 : E(X) < 1000$

c) $H_0 : E(X) = 1000 \quad ; \quad H_1 : E(X) > 1000$

d) $H_0 : E(X) = 1000 \quad ; \quad H_1 : E(X) \neq 1000$

e) $H_0 : E(X) \neq 1000 \quad ; \quad H_1 : E(X) = 1000$

f) Keine der obigen Antworten ist richtig.

3-ii) Wie lautet näherungsweise der Wert der Teststatistik?

a) $g(X_1, ..., X_{20}) = -0,5687$

b) $g(X_1, ..., X_{20}) = -0,0835$

c) $g\left(X_1, ..., X_{20}\right) = 0,5687$

d) $g\left(X_1, ..., X_{20}\right) = 515,9$

e) Keine der obigen Antworten ist richtig.

3-iii) Wie lautet der Annahmebereich etwa?

a) $[-2,093; 2,093]$

b) $[-1,96; 1,96]$

c) $[-\infty; 1,729]$

d) $[-1,729; \infty]$

e) Keine der obigen Antworten ist richtig.

3-iv) Wird die Nullhypothese abgelehnt?

a) ja

b) nein

c) Keine der obigen Antworten ist richtig.

Wettbewerb Aufgabenblatt 7
Abgabe bis Dienstag, 6. Dezember, 12 Uhr

1.

Im Rahmen einer Umfrage zum Wahlverhalten bei der nächsten Komunalwahl wurden 1000 Passanten befragt.
Die Umfrage ergab, dass 183 von 480 befragten Männern CDU und 166 SPD wählen würden. Hingegen würden nur 142 Frauen CDU, jedoch 184 SPD wählen. Für die Grünen würden sich 147 Frauen von 225 Befragten entscheiden und die FDP könnte sich über immerhin 53 Stimmen von den männlichen Befragten freuen.

Überprüfen Sie die Erhebung dahingehend, ob sich eine Abhängigkeit zwischen Wahlpräferenz und Geschlecht zu einem Signifikanzniveau von $\alpha = 5\%$ nachweisen lässt.

1-i) Wie müssen in diesem Fall die Hypothesen lauten (mehrere richtige Antworten möglich)?

a) H_0 : Das Wahlverhalten ist nicht gleichverteilt. H_1 : Das Wahlverhalten ist gleichverteilt.

b) H_0 : Geschlecht und Wahlverhalten sind unabhängig voneinander.H_1 : Geschlecht und Wahlverhalten sind abhängig voneinander.

c) H_0 : Das Wahlverhalten ist gleichverteilt. H_1 : Das Wahlverhalten ist nicht gleichverteilt.

d) H_0 : Geschlecht und Wahlverhalten sind abhängig voneinander., H_1 : Geschlecht und Wahlverhalten sind unabhängig voneinander.

e) Keine der obigen Antworten ist richtig.

1-ii) Wählen Sie den korrekten Ablehnungsbereich aus (mehrere richtige Antworten möglich)!

a) $g \in [7,8147;\infty]$

b) $g \in [-\infty;7,8147]$

c) $g < Q_3^{\chi^2}(0,95)$

d) $g > Q_3^{\chi^2}(0,95)$

e) Keine der obigen Antworten ist richtig.

1-iii) Lässt sich die Nullhypothese ablehnen?

a) Ja, mit einem Teststatistikwert von g = 26,0597

b) Nein, mit einem Teststatistikwert von g = 2,3335

c) Ja, mit einem Teststatistikwert von g = 13,2981

d) Nein, mit einem Teststatistikwert von g = 8,4611

e) Keine der obigen Antworten ist richtig.

2.

Eine Umfrage zu den Essgewohnheiten von 1000 zufällig Befragten ergab folgendes Ergebnis:

Gewicht	Regelmäßigkeit der Mahlzeiten regelmäßig	leicht unregelmäßig	stark unregelmäßig	$\sum$
Normalgewicht	350	150	100	600
Über- und Untergewicht	100	50	250	400
$\sum$	450	200	350	1000

Es soll zum Signifikanzniveau $\alpha = 10\%$ getestet werden, ob eine Abhängigkeit zwischen dem Gewicht der Testpersonen und der Regelmäßigkeit der Mahlzeiten vorliegt.

2-i) Welche Hypothesen müssen Sie aufstellen?

a) H_0: Die zwei Merkmale sind unabhängig voneinander,
H_1: Die zwei Merkmale sind nicht unabhängig voneinander

b) H_0: Die zwei Merkmale sind nicht unabhängig voneinander,
H_1: Die zwei Merkmale sind unabhängig voneinander

c) H_0: Die zwei Merkmale gehören zum selben Verteilungstyp,
H_1: Die zwei Merkmale gehören nicht zum selben Verteilungstyp

d) H_0: Die zwei Merkmale gehören nicht zum selben Verteilungstyp,
H_1: Die zwei Merkmale gehören zum selben Verteilungstyp

e) Keine der obigen Antworten ist richtig.

2-ii) Wie lautet ungefähr der Wert der Teststatistik (auf 4 Stellen gerundet)

a) $g \approx 2,6753$

b) $g \approx 222,0569$

c) $g \approx 275,6857$

d) keine der obigen Antworten ist richtig

2-iii) Kann die Nullhypothese angenommen werden?

a) Ja, denn H_0 kann abgelehnt werden

b) Ja,denn H_1 kann abgelehnt werden

c) Nein, denn H_0 kann abgelehnt werden

d) Nein, denn H_1 kann abgelehnt werden

e) Keine der obigen Antworten ist richtig.

3.

Ein Speditionsunternehmen hat 100 Fahrzeuge. Im vergangenen Jahr waren alle in der Werkstatt und bei allen mussten Reparaturen vorgenommen werden. Ein Controller der Firma hat die Reparaturfälle und dazu die Hersteller der Fahrzeuge zusammengetragen wie in der folgenden Tabelle. Er will jetzt testen, ob die aufgetretenen Reparaturfälle vom Hersteller abhängen.

n_{ij} / e_{ij} / $(n_{ij}-e_{ij})^2$ / $\frac{(n_{ij}-e_{ij})^2}{e_{ij}}$	Bremsen	Motor	gesamt $n_{.j}$
Volvo	**21**/12/81/6,75	**9**/18/E/4,5	**30**
Mercedes	**A**/12/9/G	**21**/C/F/H	**30**
MAN	**10**/D/36/2,25	**B**/24/36/1,5	**40**
gesamt $n_{i.}$	**40**	**60**	**100**

3-i) Ergänzen Sie die fehlenden Zahlen in der Tabelle.

a) $A =$

A. 9

B. 10

C. 21

D. 30

E. andere

c) $C =$

A. 12

B. 16

C. 18

D. 24

E. andere

e) $E =$

A. 4

B. 9

C. 36

D. 81

E. andere

g) $G =$

A. 0.5

B. 0.75

C. 2.25

D. 6.75

E. andere

b) $B =$

A. 9

B. 10

C. 21

D. 30

E. andere

d) $D =$

A. 12

B. 16

C. 18

D. 24

E. andere

f) $F =$

A. 4

B. 9

C. 36

D. 81

E. andere

h) $H =$

A. 0.5

B. 0.75

C. 2.25

D. 6.75

E. andere

3-ii) Wie lautet der Wert der Testfunktion?

a) $g((X_1, Y_1), ..., (X_{100}, Y_{100})) = 2,52$

b) $g((X_1, Y_1), ..., (X_{100}, Y_{100})) = 16,25$

c) $g((X_1, Y_1), ..., (X_{100}, Y_{100})) = 100$

d) $g((X_1, Y_1), ..., (X_{100}, Y_{100})) = 252$

e) Keine der obigen Antworten ist richtig.

3-iii) Nehmen Sie an, der Wert der Testfunktion wäre $g = 6$. Könnte dann die Nullhypothese der Unabhängigkeit zwischen der Reparaturanfälligkeit einzelner Bestandteile und dem Hersteller verworfen werden (mehrere richtige Antworten möglich)?

a) Ja, bei einem Signifikanzniveau von 10%.

b) Ja, bei einem Signifikanzniveau von 5%.

c) Ja, bei einem Signifikanzniveau von 2.5%.

d) Ja, bei einem Signifikanzniveau von 1%.

e) Keine der obigen Antworten ist richtig.

Wettbewerb Aufgabenblatt 8
Abgabe bis Dienstag, 13. Dezember, 12 Uhr

1.

Kennzeichnen Sie die wahren Aussagen (mehrere richtige Antworten möglich)!

a) Die endogene Variable ist unabhängig.

b) Bei der KQ-Methode wird der senkrechte Abstand der einzelnen Datenpunkte zur Regressionsgeraden minimiert.

c) Das Merkmal Alter in ganzen Jahren erklärt das Einkommen. Obwohl nur 70 jährige Menschen untersucht werden kann eine Regressionsgerade erstellt werden.

d) Im Allgemeinen können erklärende und erklärte Variable beliebig getauscht werden, ohne dass sich das Ergebnis verändert.

e) Wenn das Bestimmtheitsmaß 1 ist liegen alle Datenpunkte auf der Regressionsgeraden.

f) Die Wurzel des Bestimmtheitsmaßes entspricht dem Korrelationskoeffizient nach Pearson.

g) Keine der obigen Antworten ist richtig.

2.

Es soll untersucht werden, ob die Ausgaben für Werbung einen Einfluss auf die Absatzmenge eines Produktes haben. Dazu wurde folgende Stichprobe erhoben:

Ausgaben für Werbung	900	1300	1200	400	700	800	1000
Absatz des Produktes	400	700	550	100	250	300	500

Die Ausgaben für Werbung werden in der Variable X definiert, die Absatzmenge des Produktes in der Variable Y.
Sie führen eine lineare Regression zur Überprüfung des Zusammenhangs durch.

2-i) Welche der folgenden Aussagen sind wahr, bzw. welche Annahmen werden gemacht (mehrere richtige Antworten möglich)?

a) X ist die unabhängige Variable, Y ist die abhängige Variable

b) X ist die abhängige Variable, Y ist die unabhängige Variable

c) Sie wissen noch nicht, was die abhängige und was die unabhängige Variable ist. Deshalb führen Sie eine lineare Regression durch.

d) Sie vermuten zwischen X und Y einen linearen Zusammenhang.

e) Die Residuen sind normalverteilt.

f) Die Residuen korrelieren mit der abhängigen Variable.

g) Die Residuen korrelieren mit sich selbst (= Autokorrelation der Residuen).

h) Keine der obigen Antworten ist richtig.

2-ii) Wie lautet die Regressionsgerade?

a) $Y_i = 228\frac{1}{6} \cdot X_i - 204912\frac{1}{2}$

b) $Y_i = 1\frac{39}{73} \cdot X_i - 980\frac{60}{73}$

c) $Y_i = \frac{73}{112} \cdot X_i - 186\frac{17}{28}$

d) Keine der obigen Antworten ist richtig.

2-iii) Welchen Wert hat das Bestimmtheitsmaß R^2 ungefähr?

a) $R^2 = 0,9710$

b) $R^2 = 0,9854$

c) $R^2 = 1,0298$

d) Keine der obigen Antworten ist richtig.

3.

Gegeben seien folgende Beobachtungen:

i	1	2	3	4
x_i	-1	-1	1	1
y_i	0	2	2	0

Wie muss in diesem Fall die beste Regressionsgerade lauten (Antwort ohne zu Rechnen möglich)?

a) $Y_i = 0$

b) $Y_i = X_i$

c) $Y_i = 1$

d) $Y_i = 1 - X_i$

e) $Y_i = 2$

f) Keine der obigen Antworten ist richtig.

4.

Ein Erlebnisbad will untersuchen, inwiefern sich die durchschnittliche Temperatur (X) auf die Besucherzahlen (Y) auswirkt. Dazu liegen folgende Daten vor:

	durchschnittliche Temperatur	Besucherzahlen
Januar/Februar	-0,3°	6327
März/April	6,8°	6703
Mai/Juni	14,9°	5688
Juli/August	20,1°	2492
September/Oktober	12,2°	5182
November/Dezember	2,3°	6159

4-i) Welches ist hierbei die erklärende Variable und welches die erklärte?

a) Y ist die erklärende Variable, X die erklärte.

b) X ist die erklärende Variable, Y die erklärte.

c) Keine der obigen Antworten ist richtig.

4-ii) Wie können Sie die Regressionsgerade mit Hilfe von R bestimmen (mehrere richtige Antworten möglich)? Die Befehle

```
> x <- c(-0.3,6.8,14.9,20.1,12.2,2.3)
> y <- c(6327,6703,5688,2492,5182,6159)
```

wurden bereits ausgeführt.

a) `lm(x ~ y)`

b) `lm(y ~ x)`

c) `summary(lm(x ~ y))`

d) `summary(lm(y ~ x))`

e) Keine der obigen Antworten ist richtig.

4-iii) Wie lautet die Regressionsgerade?

a) $Y_i = 6896,7 - 157,7 \cdot X_i$

b) $Y_i = -157,6 + 6896,7 \cdot X_i$

c) $Y_i = 6896,7 + 157,6 \cdot X_i$

d) $X_i = 6896,7 - 157,6 \cdot Y_i$

e) $X_i = 6896,7 + 157,6 \cdot Y_i$

f) Keine der obigen Antworten ist richtig.

4-iv) Ist β_0 signifikant von 0 verschieden?

a) Ja, es ist signifikant von 0 verschieden.

b) Nein, es ist nicht signifikant von 0 verschieden.

c) Keine der obigen Antworten ist richtig.

Wettbewerb Aufgabenblatt 9
Abgabe bis Dienstag, 20. Dezember, 12 Uhr

1.

In den folgenden Situationen sei der wahre Zusammenhang jeweils durch die Gleichung

$$Y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + u$$

gegeben. Sie schätzen jedoch

$$Y = \beta_0 + \beta_1 \cdot X_1 + u.$$

Betrachten Sie die die folgenden Situationen. Welches Vorzeichen sollte man für die Differenz $E(\hat{\beta}_1) - \beta_1$ erwarten?

1-i) Y = Einkommen von Arbeitern
X_1 = Arbeitserfahrung
X_2 = Alter des Arbeiters
Datenbasis = Alle männlichen Arbeiter einer Autofabrik in Michigan, 1997

a) $E(\hat{\beta}_1) - \beta_1 > 0$

b) $E(\hat{\beta}_1) - \beta_1 < 0$

c) $E(\hat{\beta}_1) - \beta_1 = 0$

d) Keine der obigen Antworten ist richtig.

1-ii) Y = Anzahl der Besucher einer Open-Air Veranstaltung der Kulturarena
X_1 = Wochenende-Dummy (= 1, falls Wochenende)
X_2 = Temperatur in °C
Datenbasis = alle Open-Air Veranstaltungen der Kulturarena in Jena im Sommer 2007.

a) $E(\hat{\beta}_1) - \beta_1 > 0$

b) $E(\hat{\beta}_1) - \beta_1 < 0$

c) $E(\hat{\beta}_1) - \beta_1 = 0$

d) Keine der obigen Antworten ist richtig.

1-iii) Y = Nachfrage nach Schokolade
X_1 = Preis für Schokolade
X_2 = verfügbares Einkommen
Datenbasis = jährliche Nachfrage in den USA von 1972–2002. Preise und Einkommen sind nicht inflationsbereinigt. (Gehen Sie davon aus, dass Schokolade kein Giffen-Gut ist.)

a) $E(\hat{\beta}_1) - \beta_1 > 0$

b) $E(\hat{\beta}_1) - \beta_1 < 0$

c) $E(\hat{\beta}_1) - \beta_1 = 0$

d) Keine der obigen Antworten ist richtig.

2.

Welche der folgenden Aussagen sind wahr (mehrere richtige Antworten möglich)?

a) Schätzfehler entstehen im Regressionsmodell aufgrund unberücksichtigter Einflussgrößen und aufgrund von Fehlern in den Daten (Mess- und Auswahlfehler)

b) Um die Varianz der Schätzer zu minimieren, ist es ratsam, alle verfügbaren Variablen in das Modell zu integrieren.

c) Haben die Residuen eine konstante Varianz, liegt Heteroskedastizität vor.

d) Autokorrelation liegt vor, wenn die Residuen der Grundgesamtheit unkorreliert sind.

e) Multikollinearität kann durch die Betrachtung der Korrelationsmatrix aufgedeckt werden. Betragsmäßig niedrige Korrelationskoeffizienten zwischen den unabhängigen Variablen bedeuten Multikollinearität.

f) Der Wert eines Regressionskoeffizienten dient als Maß der Wichtigkeit der betreffenden Variablen.

g) Das Bestimmtheitsmaß R^2 misst die Güte der Anpassung der Regressionsfunktion an die empirischen Daten.

h) Keine der obigen Antworten ist richtig.

3.

Ein Delikatessenhersteller in Deutschland besitzt 5 Filialen in 5 verschiedenen Städten. Das Geschäft läuft gut; es soll eine weitere Filiale eröffnet werden. Um eine richtige Standortwahl treffen zu können, sollen die externen Erfolgsfaktoren des Gewinns identifiziert werden. Dazu stehen ihnen folgende Daten zur Verfügung:

```
> earn <- c(20000,32500,42000,12000,28000) # Gewinn
> hab <- c(12000,72000,164000,60000,16000) # Einwohner
> park <- c(600,3600,8200,3000,800) # Parkplätze
> mil <- c(15,19,24,11,17) # Anzahl Millionäre
```

3-i) Bestimmen Sie zu einem Signifikanzniveau von $\alpha = 5\%$ die Erfolgsfaktoren des Delikatessenhändlers auf Basis der Ihnen zur Verfügung stehenden Daten. Nutzen Sie dazu die einfache lineare Regression (d.h. nur eine erklärende Variable). Welche der folgenden Aussagen sind richtig (mehrere richtige Antworten möglich)?

a) Die Anzahl der Einwohner pro Stadt ist ein Erfolgsfaktor.

b) Die Anzahl der Parkplätze pro Stadt ist ein Erfolgsfaktor.

c) Die Anzahl der Millionäre pro Stadt ist ein Erfolgsfaktor.

d) Keine der obigen Antworten ist richtig.

3-ii) Ihr Kollege reicht Ihnen die Auswertung der Daten mittels einer multiplen Regression.

```
> est1=lm(earn~hab+park)
> est2=lm(earn~hab+mil)
> est3=lm(earn~mil+park)
> est4=lm(earn~hab+park+mil)
> mtable(est1, est2, est3, est4, coef.style="all", summary.stats=c("N"))
```

```
Calls:
est1: lm(formula = earn ~ hab + park)
est2: lm(formula = earn ~ hab + mil)
est3: lm(formula = earn ~ mil + park)
est4: lm(formula = earn ~ hab + park + mil)

============================================================
                 est1        est2        est3        est4
------------------------------------------------------------
(Intercept)  18650.509  -14873.813  -14873.813  -14873.813
             (6749.754)  (4066.118)  (4066.118)  (4066.118)
                (2.763)    (-3.658)    (-3.658)    (-3.658)
                (0.070)     (0.067)     (0.067)     (0.067)
hab              0.127      -0.011                  -0.011
                (0.079)     (0.022)                 (0.022)
                (1.602)    (-0.482)                (-0.482)
                (0.207)     (0.677)                 (0.677)
mil                       2469.044*   2469.044*   2469.044*
                          (283.160)   (283.160)   (283.160)
                            (8.720)     (8.720)     (8.720)
                            (0.013)     (0.013)     (0.013)
park                                    -0.214
                                        (0.444)
                                       (-0.482)
                                        (0.677)
------------------------------------------------------------
N                  5           5           5           5
============================================================
```

Welche der folgenden Aussagen sind zu einem Signifikanzniveau von $\alpha = 5\%$ richtig (mehrere richtige Antworten möglich)?

a) Die Anzahl der Einwohner pro Stadt ist ein Erfolgsfaktor.

b) Die Anzahl der Parkplätze pro Stadt ist ein Erfolgsfaktor.

c) Die Anzahl der Millionäre pro Stadt ist ein Erfolgsfaktor.

d) Die Anzahl der Einwohner und Parkplätze pro Stadt sind Erfolgsfaktoren.

e) Die Anzahl der Einwohner und Millionäre pro Stadt sind Erfolgsfaktoren.

f) Die Anzahl der Parkplätze und Millionäre pro Stadt sind Erfolgsfaktoren.

g) Es existieren kollineare Regressoren.

h) Die Anzahl der Parkplätze ist die Omitted Variable.

i) Keine der obigen Antworten ist richtig.

Wettbewerb Aufgabenblatt 10
Abgabe bis Dienstag, 3. Januar, 12 Uhr

1.

Ein Pharmaunternehmen forscht an einem neuem Medikament gegen Schlaflosigkeit. Hierzu führte es vor Anwendung des Medikaments eine Kurzstudie unter 8 Probanden aus 8 verschiedenen Städten durch, um herauszufinden, welche Größe am meisten Einfluss auf die Schlafdauer ausübt. Folgende Daten kamen dabei heraus:

```
> Schlafdauer <- c(6,7,8,10,9,11,8,5) # in Stunden
> Alter <- c(49,47,20,18,52,19,65,72) # in Jahren
> Gewicht <- c(70,90,85,100,69,52,62,90) # in kg
> Groesse <- c(172,196,182,181,176,169,175,190) # in cm
```

1-i) Welche Einflussgröße hat zu einem Signifikanzniveau von $\alpha = 5\%$ Einfluss auf die Schlafdauer? Führen Sie dazu eine (multiple) lineare Regression durch. Welche Aussagen sind richtig? (mehrere richtige Antworten möglich)

a) Das Alter hat signifikanten Einfluss.

b) Das Gewicht hat signifikanten Einfluss.

c) Die Größe hat signifikanten Einfluss.

d) Keine der obigen Antworten ist richtig.

1-ii) Führen Sie nun eine multiple Regression durch. Betrachten Sie hierbei folgende Zusammenhänge:

```
> est1 = lm(Schlafdauer ~ Alter + Gewicht)
> est2 = lm(Schlafdauer ~ Alter + Groesse)
> est3 = lm(Schlafdauer ~ Groesse +  Gewicht)
> est4 = lm(Schlafdauer ~ Alter + Groesse + Gewicht)
> library(memisc)
> mtable(est1, est2, est3, est4, coef.style = "all", summary.stats = c("N"))
```

```
Calls:
est1: lm(formula = Schlafdauer ~ Alter + Gewicht)
est2: lm(formula = Schlafdauer ~ Alter + Groesse)
est3: lm(formula = Schlafdauer ~ Groesse + Gewicht)
est4: lm(formula = Schlafdauer ~ Alter + Groesse + Gewicht)

====================================================
                est1      est2      est3      est4
----------------------------------------------------
(Intercept)   14.561**  24.045    31.404    16.648
              (2.758)  (10.814)  (19.690)  (16.311)
              (5.280)   (2.223)   (1.595)   (1.021)
              (0.003)   (0.077)   (0.172)   (0.365)
Alter         -0.070*   -0.059              -0.068
```

```
               (0.024)   (0.026)              (0.031)
              (-2.891)  (-2.261)             (-2.188)
               (0.034)   (0.073)              (0.094)
Gewicht        -0.046               0.020     -0.039
               (0.032)             (0.073)    (0.062)
              (-1.460)             (0.273)   (-0.640)
               (0.204)             (0.796)    (0.557)
Groesse                  -0.075    -0.139     -0.015
                         (0.061)   (0.132)    (0.114)
                        (-1.223)  (-1.052)   (-0.130)
                         (0.276)   (0.341)    (0.903)
----------------------------------------------------
N                 8         8         8          8
====================================================
```

Welche Aussagen sind zu einem Signifikanzniveau von $\alpha = 5\%$ korrekt (mehrere richtige Antworten möglich)?

a) Es existieren kollineare Regressoren.

b) Bei Betrachtung aller 3 Größen in est4 hat das Alter den größten Einfluss.

c) Der t-Wert des Alters in est4 beträgt -2.188.

d) Ein AIC-Wert vom Einfluss der Größe und des Gewichts auf die Schlafdauer beträgt 13.52641.

e) Keine der obigen Antworten ist richtig.

2.

Bitte beantworten Sie folgende Fragen bzw. entscheiden Sie, ob folgende Aussagen zum Thema „t-Statistik für individuelle Koeffizienten“ wahr oder falsch sind.

2-i) Sie haben eine lineare Regression geschätzt und eine Regressionsfunktion erhalten. Ein geeignetes Prüfkriterium um zu testen, ob die unabhängigen Variablen einen signifikanten Einfluss auf die abhängige Variable haben, ist die t-Statistik.

a) wahr

b) falsch

c) Keine der obigen Antworten ist richtig.

2-ii) Die t-Statistik berechnet sich

a) aus der Differenz zwischen geschätztem Regressionskoeffizienten und dem nullhypothetischen Regeressionskoeffizientens dividiert durch den Standardfehler des Regressionskoeffizienten.

b) aus der Differenz zwischen geschätztem Regressionskoeffizienten und dem wahren Regeressionskoeffizienten dividiert durch den Standardfehler des Regressionskoeffizienten.

c) Keine der obigen Antworten ist richtig.

2-iii) Üblicherweise stellen Sie folgende Nullhypothese auf, um zu überprüfen, ob eine Variable in ein Modell gehört:

a) $H_0 : \beta_i = 0$

b) $H_0 : \beta_i \neq 0$

c) Keine der obigen Antworten ist richtig.

2-iv) Wenn Sie die Nullhypothese nicht verwerfen,...

a) können Sie sich recht sicher sein, dass ein Einfluss der unabhängigen Variablen X_i auf die abhängige Variable Y existiert.

b) können Sie sich recht sicher sein, dass kein Einfluss der unabhängigen Variablen X_i auf die abhängige Variable Y existiert.

c) Keine der obigen Antworten ist richtig.

2-v) Lehnen Sie die Nullhypothese ab,

a) können Sie folgern, dass ein Einfluss der unabhängigen Variablen X_i auf die abhängige Variable Y existiert.

b) können Sie folgern, dass kein Einfluss der unabhängigen Variablen X_i auf die abhängige Variable Y existiert.

c) Keine der obigen Antworten ist richtig.

2-vi) Sie führen in R folgende Regressionsanalyse durch:

```
> Absatzmenge <- c(300, 250, 100, 400, 600, 800)
> Preis <- c(250, 225, 210, 300, 325, 250)
> AusgabenWerbung <- c(600, 550, 450, 750, 900, 1100)
> est <- lm (Absatzmenge ~ Preis + AusgabenWerbung)
> t_values <- coef(est)/sqrt(diag(vcov(est)))
> t_values
```

```
   (Intercept)          Preis AusgabenWerbung
    -4.0646354     -0.3102413      15.7876720
```

Ihr Signifikanzniveau beträgt 10%. Welche Variablen haben signifikanten Einfluss?

a) Preis und Werbung

b) nur Preis

c) nur Werbung

d) weder Preis noch Werbung

e) Keine der obigen Antworten ist richtig.

Wettbewerb Aufgabenblatt 11
Abgabe bis Dienstag, 10. Januar, 12 Uhr

1.

Sie möchten untersuchen, ob das Geschlecht eines Arbeitnehmers Einfluss auf das Bruttoeinkommen hat. Dazu führen Sie eine lineare Regression durch und führen für das Geschlecht eine Dummy-Variable ein. Die Variable `Maennlich` nimmt den Wert 1 an, wenn der Arbeitnehmer männlich ist. Ist der Arbeitnehmer weiblich, nimmt die Dummy-Variable den Wert 0 an.
Sie haben Ihre Analyse in R durchgeführt und folgendes Ergebnis erhalten:

```
> Bruttoeinkommen<-c(2500,4000,3000,7000,5000,2900,1500)
> Maennlich<-c(1,0,0,0,0,1,1)
> summary(lm(Bruttoeinkommen~Maennlich))
```

```
Call:
lm(formula = Bruttoeinkommen ~ Maennlich)

Residuals:
    1     2     3     4     5     6     7
  200  -750 -1750  2250   250   600  -800

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4750.0      699.6   6.789  0.00105 **
Maennlich    -2450.0     1068.7  -2.292  0.07043 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1399 on 5 degrees of freedom
Multiple R-squared: 0.5125,       Adjusted R-squared: 0.4149
F-statistic: 5.255 on 1 and 5 DF,  p-value: 0.07043
```

1-i) Welche ist die abhängige und welche die unabhängige Variable?

a) abhängige Variable: `Bruttoeinkommen`, unabhängige Variable: `Maennlich`

b) abhängige Variable: `Maennlich`, unabhängige Variable: `Bruttoeinkommen`

c) Keine der obigen Antworten ist richtig.

1-ii) Wie lautet die Regressionsgleichung?

a) $\texttt{Bruttoeinkommen} = 4750 - 2450 \cdot \texttt{Maennlich} + u_i$

b) $\texttt{Bruttoeinkommen} = -2450 + 4750 \cdot \texttt{Maennlich} + u_i$

c) $\texttt{Bruttoeinkommen} = 6,789 - 2,292 \cdot \texttt{Maennlich} + u_i$

d) $\texttt{Bruttoeinkommen} = -2,292 + 6,789 \cdot \texttt{Maennlich} + u_i$

e) Keine der obigen Antworten ist richtig.

1-iii) Welchen Wert nimmt R^2 an?

a) $R^2 = 0,5125$

b) $R^2 = 0,4149$

c) Keine der obigen Antworten ist richtig.

1-iv) Das durchschnittliche Bruttoeinkommen der Männer ist:

a) 7200

b) 4750

c) 2300

d) Kann man nicht sagen

e) Keine der obigen Antworten ist richtig.

2.

Welcher der folgenden Aussagen sind wahr (mehrere richtige Antworten möglich)? (Aus technischen Gründen sind die Antworten auf zwei Teilaufgaben geteilt, bitte verstehen Sie die beiden folgenden Teilaufgaben als eine Aufgabe.)

2-i) a) Die Variable Studienabschluss (Diplom, Bachlor, Master) ist diskret.

b) Die Variable Studienabschluss (Diplom, Bachlor, Master) ist stetig.

c) Die Variable Studienabschluss (Diplom, Bachlor, Master) ist binär.

d) Die Variable Größe in cm (mit unendlich vielen Nachkommastellen) ist diskret.

e) Die Variable Größe in cm (mit unendlich vielen Nachkommastellen) ist stetig.

f) Die Variable Größe in cm (mit unendlich vielen Nachkommastellen) ist binär.

2-ii) (mehrere richtige Antworten möglich)

a) Die Variable Größe in cm (ohne Nachkommastellen) ist diskret.

b) Die Variable Größe in cm (ohne Nachkommastellen) ist stetig.

c) Die Variable Größe in cm (ohne Nachkommastellen) ist binär.

d) Die Variable Ehestatus (verheiratet ja/nein) ist diskret.

e) Die Variable Ehestatus (verheiratet ja/nein) ist stetig.

f) Die Variable Ehestatus (verheiratet ja/nein) ist binär.

g) Keine der obigen Antworten ist richtig.

3.

Gegeben seien die folgenden Beobachtungen.

Person	1	2	3	4	5	6	7	8
Einkommen in EUR	2000	1000	680	450	400	1050	690	1100
Konsum	1850	980	600	450	395	1000	660	980
Körpergröße	1,82	1,75	1,72	1,74	1,73	1,76	1,72	1,80

3-i) Führen Sie eine lineare Regression durch, wobei das Einkommen den Konsum erklären soll und kennzeichnen Sie die wahren Aussagen (mehrere richtige Antworten möglich)! β_0 sei hier der Achsenabschnitt und β_1 sei die Steigung der exogenen Variablen im Regressionsmodell.

a) Eine Einkommenserhöhung um 1 Euro erhöht den Konsum um ca. 91 Cent.

b) Sowohl β_0 als auch β_1 sind signifikant von Null verschieden.

c) Das Modell weist einen hohen Erklärungsgehalt auf.

d) Keine der obigen Antworten ist richtig.

3-ii) Wandeln Sie nun die exogene Variable Einkommen in eine Dummy Variable um. Unterscheiden Sie Ausprägungen größer als 1000 und kleiner bzw. gleich 1000. Prüfen Sie die folgenden Aussagen (mehrere richtige Antworten möglich)!

a) Das durschnittliche Einkommen der Personen, die über 1000 Euro verdienen beträgt 1276,67 Euro.

b) Der Unterschied der Mittelwerte beider Gruppen beträgt 617.

c) Für $\alpha = 0.05$ ist β_1 signifikant von Null verschieden.

d) Keine der obigen Antworten ist richtig.

3-iii) Berücksichtigen Sie nun (zusätzlich zu der Situation aus 3-ii) auch die Beobachtungen zum Merkmal Körpergröße. Sowohl Einkommen als auch Körpergröße werden als binäre Variable modelliert. Unterscheiden Sie bei Körpergröße Beobachtungen mit einer Größe von mindestens 1.75m von solchen mit einer geringeren Körpergröße. Prüfen Sie, ob man die Interaktion zwischen den beiden erklärenden Variablen in das Modell aufnehmen kann. (mehrere richtige Antworten möglich):

a) Es liegen zwei binäre Regressoren vor.

b) Der Koeffizient für Größe ist signifikant von Null verschieden, wenn man eventuelle Interaktionen zwischen den Regressoren unberücksichtigt lässt.

c) Werden die Interaktionsbeziehungen zwischen den beiden Regressoren im Regressionsmodell zugelassen, so erhält man ein Bestimmtheitsmaß von 0.645.

d) Der Interaktionsterm kann wegen Multikolliniarität nicht bestimmt werden.

e) Personen die mindestens 1.75m groß sind und ein Einkommen von höchstens 1000 Euro haben, haben im Mittel einen Konsum von 960 Euro.

f) Menschen die kleiner als 1.75m groß sind und höchstens 1000 Euro verdienen, haben ein Durchschnittseinkommen von 526,2.

g) Keine der obigen Antworten ist richtig.

Wettbewerb Aufgabenblatt 12
Abgabe bis Dienstag, 17. Januar, 12 Uhr

1.

Gegeben seien die folgenden Beobachtungen.

Person	1	2	3	4	5	6	7	8	9
Einkommen in EUR	6900	1000	680	450	400	1050	20000	1100	4000
Marginale Konsumquote	0.6	0.85	0.98	1	0.97	0.82	0.53	0.9	0.7

1-i) Führen Sie eine lineare Regression durch, wobei das Einkommen die marginale Konsumquote erklären soll (mehrere richtige Antworten möglich). Es gilt:

$$c = \beta_0 + \beta_1 \cdot \text{Einkommen}$$

a) Eine Einkommenserhöhung führt zu einer signifikanten ($\alpha = 5\%$) Reduktion der Konsumquote.

b) Die Anwendung des linearen Modells ist empfehlenwert.

c) Die signifikanten Schätzergebnisse zeigen eindeutig, dass das Modell gut ist.

d) Keine der obigen Antworten ist richtig.

1-ii) Verwenden Sie nun ein Polynom zweiten Grades. Anschließend schätzen Sie erneut mit einem Polynom dritten Grades (mehrere richtige Antworten möglich).

a) Auf Basis der Signifikanz der Parameter sollten wir das Modell mit einem Polynom zweiten Grades einem Modell mit einem Polynom dritten Grades vorziehen.

b) Das quadratische Modell ist dem linearen Modell nicht vorzuziehen.

c) Das quadratische Modell hat einen hohen Erklärungsgehalt.

d) Keine der obigen Antworten ist richtig.

2.

Ein Züchter von Bonsaibäumen nutzt Düngermittel, um das Wachstum zu beschleunigen. Um die Wirkung zu analysieren, hat er im letzten Jahr eine statistische Erhebung über die Wirksamkeit vorgenommen. Diese ergab folgende Daten. Y beschreibt das Gewicht (in Gramm) eines Bonsais nach einem Jahr und X die täglich eingesetzte Menge an Dünger (in Gramm).

Eingesetztes Düngermittel (in Gramm)	Gewicht des Bonsais (in Gramm)
1	4300
5	5400
8	5800
10	6600
14	7500
17	7200
20	8000
22	7200
25	8300
28	8100
30	8400
33	8300

2-i) Führen sie eine einfache lineare Regression durch. Wie lautet die Regressionsgrade?

a) $Y_i = 3820.09 + 1266.45 \cdot \ln(x_i) + u_i$

b) $\ln(Y_i) = 3820.09 + 1266.45 \cdot x_i + u_i$

c) $Y_i = 4571.47 + 138.32 \cdot x_i + u_i$

d) $Y_i = 4991.47 + 118.32 \cdot x_i + u_i$

e) Keine der obigen Antworten ist richtig.

2-ii) Führen sie eine lineare Regression mit dem linear-log Modell durch. Wie lautet die Regressionsgrade?

a) $Y_i = 3820.1 + 1266.4 \cdot \ln(x_i) + u_i$

b) $\ln(Y_i) = 3820.1 + 1266.4 \cdot x_i + u_i$

c) $Y_i = 8.524 + 0.018 \cdot \ln(x_i) + u_i$

d) $\ln(Y_i) = 8.524 + 0.018 \cdot x_i + u_i$

e) Keine der obigen Antworten ist richtig.

2-iii) Welchen der folgenden Aussagen stimmen sie zu (mehrere richtige Antworten möglich)?

a) Der Grenznutzen des Düngers wird in (2-i) als konstant angenommen.

b) Der Grenznutzen des Düngers wird in (2-ii) als konstant angenommen.

c) Der Grenznutzen des Düngers wird in (2-ii) als abnehmend angenommen.

d) Der Grenznutzen des Düngers wird in (2-ii) als zunehmend angenommen.

e) Auf Basis des R^2 ist das Modell aus (2-i) vorzuziehen.

f) Auf Basis des R^2 ist das Modell aus (2-ii) vorzuziehen.

g) Keine der obigen Antworten ist richtig.

3.

Sie schätzen die folgende Gleichung:

$$Y_i = 20 + 5 \cdot \ln X_i + u_i$$

Welche der folgenden Aussagen sind wahr (mehrere richtige Antworten möglich)?

a) Es liegt ein log-linares Modell vor.

b) Es liegt ein linear-log Modell vor.

c) Der marginale Effekt einer Änderung von X_i ist $\frac{\partial Y_i}{\partial X_i} = 5$.

d) Der marginale Effekt einer Änderung von X_i ist $\frac{\partial Y_i}{\partial X_i} = 5 \cdot \frac{1}{X_i}$.

e) Wenn sich X_i um 1% ändert, dann ändert sich Y_i um 0,05 Eineiten.

f) Keine der obigen Antworten ist richtig.

Wettbewerb Aufgabenblatt 13
Abgabe bis Dienstag, 24. Januar, 12 Uhr

1.

Ein Marktforschungsunternehmen untersucht den Zusammenhang zwischen der Nachfrage nach Schokolade, dem verfügbaren Einkommen und der Nachfrage nach Gummitieren. Dabei soll folgender Zusammenhang gelten:

$$\begin{aligned}\text{Nachfrage nach Schokolade} =& \beta_1 * \text{verfügbares Einkommen} \\ &+ \beta_2 * \text{Nachfrage nach Gummitieren} + \beta_0 + u\end{aligned}$$

Dazu wurden die Daten der Nachfrage nach Schokolade (in Tsd.), das durchschnittliche verfügbare Einkommen (in Euro) und der Nachfrage nach Gummitieren (in Tsd.) der letzten 10 Jahre analysiert. Dazu gab es den folgenden R-Output:

```
Call:
lm(formula = Schokolade ~ Einkommen + Gummitiere)

Residuals:
    Min      1Q  Median      3Q     Max
-2072.2  -560.2   131.8   453.5  2038.7

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.501e+04  1.602e+03  -9.374 3.27e-05 ***
Einkommen    5.994e+00  6.650e-01   9.014 4.22e-05 ***
Gummitiere   5.789e-01  3.629e-01   1.595    0.155
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1218 on 7 degrees of freedom
Multiple R-squared: 0.923,       Adjusted R-squared: 0.901
F-statistic: 41.96 on 2 and 7 DF,  p-value: 0.0001266
```

1-i) Wie groß sind die Werte von β_0, β_1 und β_2?

a) $\beta_0 = 15010, \beta_1 = -5.994, \beta_2 = -0.5789$

b) $\beta_0 = 5.994, \beta_1 = -15010, \beta_2 = 0.5789$

c) $\beta_0 = 0.5789, \beta_1 = 5.994, \beta_2 = 15010$

d) $\beta_0 = -15010, \beta_1 = 5.994, \beta_2 = 0.5789$

e) $\beta_0 = 15010, \beta_1 = 5.994, \beta_2 = 0.5789$

f) Keine der obigen Antworten ist richtig.

1-ii) Welche dieser Aussagen sind wahr (mehrere richtige Antworten möglich)?

a) Das Bestimmtheitsmaß beträgt 0.901.

b) Das Bestimmtheitsmaß beträgt 0.923.

c) Das verfügbare Einkommen hat eigentlich keinen Einfluss auf die Nachfrage nach Schokolade.

d) Die Nachfrage nach Gummitieren hat keinen signifikanten Einfluss auf die Nachfrage nach Schokolade.

e) Wenn sich das durchschnittlich verfügbare Einkommen um einen Euro erhöht, dann erhöht sich die Nachfrage nach Schokolade um 5.994.

f) Erhöht sich die Nachfrage nach Gummitieren um 1000, dann sinkt die Nachfrage nach Schokolade um 578.9.

g) Keine der obigen Antworten ist richtig.

2.

2-i) Gegeben sei ein Konfidenzintervall zum Niveau $1 - \alpha$. Kennzeichnen Sie die richtigen Aussagen (mehrere richtige Antworten möglich):

a) Die Intervallgrenzen eines Konfidenzintervalls sind eine Zufallsvariable.

b) Der zu schätzende wahre Parameter liegt mit einer Wahrscheinlichkeit von $1 - \alpha$ im Konfindenzintervall.

c) In $(1 - \alpha) * 100\%$ aller Fälle erhält man ein Intervall, das den unbekannten wahren Parameter enthält.

d) Das Konfidenzintervall ist um so kleiner, je kleiner α ist.

e) Keine der obigen Antworten ist richtig.

2-ii) Zur Herstellung von Kaffeepulver werden die Bohnen in einer Mühle gemahlen und anschließend in eine Tüte gefüllt. Das Gewicht X der Tüte ist normalverteilt und hat den Erwartungswert θ. Aus der hundertjährigen Tradition des Unternehmens sei bekannt, dass die Varianz 256 Gramm^2 beträgt. Eine einfache Stichprobe vom Umfang $n = 25$ hat ein Gesamtgewicht von 12400 Gramm ergeben.
Bestimmen Sie ein 0.95-Konfidenzintervall für den Parameter θ. Runden Sie das Ergebnis auf drei Nachkommastellen.

a) [491,452; 499,678]

b) [487,767; 504,243]

c) [490,746; 501,264]

d) [489,728; 502, 272]

e) Keine der obigen Antworten ist richtig.

3.

Einer Pressererklärung einer kleinen amerikanischen Universität zufolge erhalten 50% der Studierenden finanzielle Unterstützung durch die Universität. Ihre Nullhypothese ist, dass diese Zahl zutrifft. Sie befragen 70 Studierende, und stellen fest, dass davon 30 finanzielle Unterstützung durch ihre Universität erhalten. Wenn Sie Ihre Teststatistik als approximativ Normalverteilt annehmen, wie bestimmen Sie einen p-Wert um Ihre Nullhypothese zu prüfen?

a) `pt((30-35)/(50*.5*.5),df=49)`

b) `pnorm((30-35)/sqrt(50*.5))`

c) `pt((30-35)/sqrt(50*.5),df=29)`

d) `2*pnorm((30-35)/sqrt(50*.5*.5))`

e) `2*pnorm((30-35)/sqrt(50*.7*.3))`

f) `pnorm((30-35)/sqrt(50*.5*.5))`

g) Keine der obigen Antworten ist richtig.

4.

4-i) X ist eine rechteckverteilte Zufallsvariable im Intervall $[a, b]$. Welche dieser Schätzfunktionen ist erwartungstreu zum Schätzen der unteren Intervallgrenze a, wenn eine Stichprobe vom Umfang n=10 vorliegt und die X_i unabhängig und identisch verteilt sind?

a) $f_1 (X_1, ..., X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} X_i$

b) $f_2 (X_1, ..., X_n) = \frac{1}{2n} \sum_{i=1}^{n} X_i - \frac{2}{n-2} \cdot b$

c) $f_3 (X_1, ..., X_n) = \sum_{i=1}^{n} X_i - \frac{8}{n} \cdot \sum_{i=1}^{n} X_i - b$

d) Keine der obigen Antworten ist richtig.

4-ii) Berechnen Sie den mittleren quadratischen Fehler der 2. Schätzfunktion aus Aufgabe 4-i.

a) $\text{MSE}(f_2) = \frac{1}{15}b^2 - \frac{1}{10}ab + \frac{1}{15}a^2$

b) $\text{MSE}(f_2) = \frac{1}{12}\left(\frac{1}{4}b^2 - \frac{1}{2}ab + 7a^2\right)$

c) $\text{MSE}(f_2) = \frac{5}{13}(a-b)^2$

d) $\text{MSE}(f_2) = \frac{1}{480} \cdot (b-a)^2 + \frac{9}{16}a^2$

e) Keine der obigen Antworten ist richtig.

4-iii) Zum Schätzen der unteren Intervallgrenze a ist zusätzlich zu den Schätzfunktionen aus 4-i eine weitere Schätzfunktion $f_4 = \frac{2}{n} \sum_{i=1}^{n} X_i - b$ gegeben. Welche der beiden Schätzfunktionen f_3, f_4 ist wirksamer?

a) f_3 ist wirksamer

b) f_4 ist wirksamer

c) beide sind gleich effizient

d) Keine der obigen Antworten ist richtig.

Wettbewerb Aufgabenblatt 14
Abgabe bis Dienstag, 31. Januar, 12 Uhr

1.

Die Füllmenge einer Kaffeetasse sei normalverteilt. Ein Kaffeemaschinenhersteller möchte eine seiner neuen Maschinen testen und herausfinden, wie viel Kaffee sie bei einem normalen Durchlauf pro Tasse erzeugt. Dazu nimmt er sein Testmodell und lässt dieses ingesamt 80 Mal durchlaufen. Am Ende des Versuchstages kennt er die gesamte Kaffeemenge der 80 Versuche, welche 15.860ml beträgt. Allgemein sei außerdem bekannt, dass die Standardabweichung einer Kaffeetassenfüllung 20ml beträgt.

1-i) Bestimmen Sie das 95% Konfidenzintervall für den Mittelwert der Füllmenge!

a) [159,051; 237,449]

b) [0; 202,633]

c) [193,867; 202,633]

d) [189,485; 207,015]

e) Keine der obigen Antworten ist richtig.

1-ii) Der Kaffeemaschinenhersteller möchte, dass die Länge des Konfidenzintervalls höchstens 6ml beträgt. Wie groß muss die neue Stichprobe mindestens gewählt werden?

2.

Susi jobbt neben ihrem Studium als Kellnerin. Sie meint, dass ihr Trinkgeld von der Arbeitszeit abhängt. In den letzten Tagen hat sie gearbeitet und Trinkgeld erhalten wie in folgender Tabelle.

Datum	Arbeitszeit	Trinkgeld
03.01.2008	5h	50 EUR
04.01.2008	4h	25 EUR
06.01.2008	2h	10 EUR
07.01.2008	7h	60 EUR
10.01.2008	3h	30 EUR

Dieser Sachverhalt soll in einem einfachen linearen Regressionsmodell abgebildet werden.

2-i) Welches ist die erklärende Variable x_i?

a) Datum

b) Arbeitszeit

c) Studium

d) Trinkgeld

e) Keine der obigen Antworten ist richtig.

2-ii) Wie berechnet sich die Spalte 8 der Tabelle 1 ($\hat{y}_i$)?

a) $(y_i - \bar{y})(x_i - \bar{x})$

b) $(y_i - \bar{y})^2$

c) $\hat{b}x_i + \hat{a}$

d) $(\bar{y} - \hat{b}x_i)$

e) Keine der obigen Antworten ist richtig.

2-iii) Was steht in Spalte 11 ($\hat{u}_i$)?

a) $\hat{u}_i = y_i - \bar{y}$, also Werte zur Berechnung der Varianz der y_i

b) $\hat{u}_i = \bar{y} - \hat{y}_i$, also die Abweichung der theoretischen Werte von Durchschnitt der Beobachtungen

c) $\hat{u}_i = y_i - \hat{y}_i$, also die Differenz zwischen beobachteten und theoretischen Werten

d) $\hat{u}_i = \hat{x}_i - \hat{y}_i$, also die Differenz zwischen unabhängigen und abhängigen Variablen

e) Keine der obigen Antworten ist richtig.

2-iv) Wie kann der Wert aus der Spalte 14, Zeile 7 interpretiert werden?

a) Erwartungswert der Residuen

b) Varianz der beobachteten Werte der erklärten Variable

c) Varianz der Residuen

d) Varianz der theoretischen Werte der erklärenden Variable

e) Keine der obigen Antworten ist richtig.

2-v) Berechnen Sie das Bestimmtheitsmaß!

a) $R^2 = \frac{35{,}878}{320} = 0,1121$

b) $R^2 = \frac{284{,}122}{320} = 0,8879$

c) $R^2 = \frac{35{,}878}{284{,}122} = 0,1263$

d) $R^2 = 1 - \frac{35{,}878}{284{,}122} = 0,8737$

e) Keine der obigen Antworten ist richtig.

2-vi) Was sagt das Bestimmtheitsmaß aus? (mehrere richtige Antworten möglich)

a) Je größer R^2, desto so besser ist die Anpassung durch die Regressionsgerade.

b) Je größer R^2, desto schlechter ist die Anpassung durch die Regressionsgerade.

c) R^2 gibt an, die groß die Varianz der erklärten Variable ist.

d) $R^2 \cdot 100\%$ der Varianz der beobachteten Werte der erklärten Variable wird durch die Varianz der theoretischen Werte erklärt.

e) R^2 ist ein Maß für die Steigung der Regressionsgeraden.

Tabelle 1 (Die Werte sind auf drei Nachkommastellen gerundet.)

Spalte	1	2	3	4	5	6	7	8
Zeile	x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$	$\hat{y}_i$
1	5	50	0.8	15	0.64	225	12	42.838
2	4	25	-0.2	-10	0.04	100	2	33.041
3	2	10	-2.2	-25	4.84	625	55	13.446
4	7	60	2.8	25	7.84	625	70	62.432
5	3	30	-1.2	-5	1.44	25	6	23.243
6 $\sum_{i=1}^{n}$	21	175	0	0	14.8	1600	145	175.000
7 $\frac{1}{5}\sum_{i=1}^{n}$	4.2	35	0	0	2.96	320	29	35.000

Fortsetzung

Spalte	9	10	11	12	13	14	15
Zeile	$(\hat{y}_i - \bar{\hat{y}}_i)$	$(\hat{y}_i - \bar{\hat{y}}_i)^2$	$\hat{u}_i$	$\hat{u}_i^2$	$(\hat{u}_i - \bar{\hat{u}}_i)$	$(\hat{u}_i - \bar{\hat{u}}_i)^2$	$((x_i - \bar{x}) \cdot \hat{u}_i)^2$
1	7.838	61.432	7.162	51.297	7.162	51.297	32.830
2	-1.959	3.839	-8.041	64.650	-8.041	64.650	2.586
3	-21.554	464.577	-3.446	11.875	-3.446	11.875	57.473
4	27.432	752.538	-2.432	5.917	-2.432	5.917	46.387
5	-11.757	138.221	6.757	45.654	6.757	45.654	65.741
6 $\sum_{i=1}^{n}$	0.000	1420.608	0.000	179.392	0.000	179.392	205.017
7 $\frac{1}{5}\sum_{i=1}^{n}$	0.000	284.122	0.000	35.878	0.000	35.878	41.003

3.

Kennzeichnen Sie die wahren Aussagen zu nichtlinearen Regressionsmodellen (mehrere richtige Antworten möglich)!

a) Für nichtlineare Regressionsmodelle werden immer Polynome verwendet. Der Grad des Polynoms wird je nach Datenlage angepasst.

b) Im log-linear Modell entspricht der Regressionskoeffizient dem marginalen Effekt.

c) Bei allen nichtlinearen Modellen entsprechen die Regressionskoeffizienten der Elastizität

d) Das Bestimmtheitsmaß ist zum Vergleich der nichtlinearen Modelle am besten geeignet.

e) Das OLS Verfahren kann nur angewandt werden, wenn ein Regressionsmodell linear in den Parametern dargestellt werden kann.

f) Keine der obigen Antworten ist richtig.

4.

Es gibt Menschen, die unter Schlafstörungen, d.h. Abweichungen vom gesunden Schlafverhalten, leiden. Eine Form der Schlafstörung ist das vorzeitige Erwachen, ohne danach wieder einschlafen zu können. Sie arbeiten in einem Schlaflabor und erheben für 16 Patienten die Schlafdauer vor (x) und nach (y) der Einnahme eines Medikaments. Ihr Assistent hat zwei Tests in R durchgeführt und präsentiert ihnen folgende Ergebnisse:

Test 1

```
> t.test(x, y, paired = TRUE)
```

```
        Paired t-test

data:  x and y
t = -4.3188, df = 15, p-value = 0.0006084
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.351107 -1.136393
sample estimates:
mean of the differences
               -2.24375
```

Test 2

```
> t.test(x,y)
```

```
        Welch Two Sample t-test

data:  x and y
t = -4.3764, df = 24.221, p-value = 0.0001993
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.301381 -1.186119
sample estimates:
mean of x mean of y
  3.53125   5.77500
```

4-i) Welche Annahmen werden von beiden obigen Tests vorausgesetzt? ((mehrere richtige Antworten möglich))

a) x und y sind binomialverteilt

b) x und y sind normalverteilt oder die Stichproben sind sehr groß.

c) Die Stichproben sind sehr klein.

d) Keine der obigen Antworten ist richtig.

4-ii) Nehmen Sie an, diese Annahmen sind erfüllt. Welcher der beiden Test ist in diesem Fall angemessen?

a) Test 1 ist angemessen.

b) Test 2 ist angemessen.

c) Keiner der beiden Tests ist angemessen.

4-iii) Ihre Nullhpothese ist, dass das Medikament keinen Einfluss auf die Schlafdauer Ihrer Patienten hat. Ihr Signifikanzniveau ist 5 %. Welche Antwort ist korrekt?

a) Die Nullhypothese kann abgelehnt werden.

b) Die Nullhypothese kann nicht abgelehnt werden.

c) Kann mithilfe des Tests nicht beurteilt werden.

Verteilungen in R:

Typ	Verteilung	Quantil
Normalverteilung	`pnorm`	`qnorm`
t-Verteilung	`pt`	`qt`
χ^2-Verteilung	`pchisq`	`qchisq`
F-Verteilung	`pf`	`qf`

Poisson Verteilung: $P_\lambda(X=k) = \lambda^k \cdot e^{-\lambda}/k!$; $E[X] = \lambda$; $\mathrm{var}(X) = \lambda$

Exponentialverteilung: $f_\lambda(X) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{sonst} \end{cases}$; $F_\lambda(X) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & \text{sonst} \end{cases}$; $E[X] = 1/\lambda$; $\mathrm{var}(X) = 1/\lambda^2$

Einige Stammfunktionen: $\int x\,dx = \frac{1}{2}x^2 + C$; $\int x^n\,dx = x^{n+1}/(n+1) + C$; $\int \frac{1}{x}\,dx = \ln x + C$; $\int a^x\,dx = a^x / \ln a + C$

Ableitung der Log-Likelihood Funktion:
$\frac{d}{d\theta} \ln L(x_1, \ldots, x_n|\theta) = \frac{f'(x_1|\theta)}{f(x_1|\theta)} + \cdots + \frac{f'(x_n|\theta)}{f(x_n|\theta)}$

Erwartungswert: $E(c \cdot X) = c \cdot E(X)$; $E(X+Y) = E(X) + E(Y)$

Varianz: $\mathrm{var}(c \cdot X) = c^2 \cdot \mathrm{var}(X)$; $\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y) + 2 \cdot \mathrm{cov}(X,Y)$

Varianz von $\bar{x}$: $\mathrm{var}(\bar{x}) = \sigma_x^2/n$

Standardabweichung von $\bar{x}$: $\sigma_{\bar{x}} = \sigma_x/\sqrt{n}$

Schätzer für Erwartungswert: $\hat{\mu}_X = \bar{x} = \frac{1}{n}\sum_i x_i$

Schätzer für Varianz: $\hat{\sigma}_X^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2$

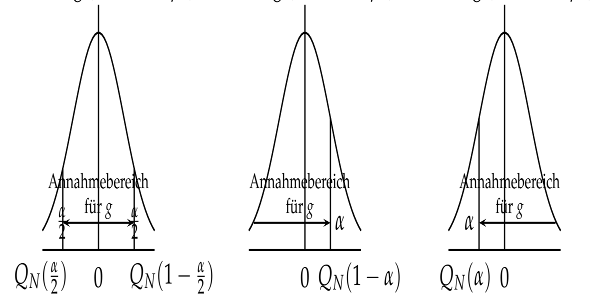
Schätzer für Standardabweichung von X:
$\hat{\sigma}_X = \sqrt{\frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2}$

Schätzer für $\sigma_{\bar{x}}$: $\hat{\sigma}_{\bar{x}} = \hat{\sigma}_x/\sqrt{n}$

Signifikanztest:

Teststatistik $g = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$

zweiseitig $(H_0 : \bar{X} = \mu_0)$ einseitig $(H_0 : \bar{X} \leq \mu_0)$ einseitig $(H_0 : \bar{X} \geq \mu_0)$



H_0 wird abgelehnt, falls g nicht im Annahmebereich liegt.

Bias: $\mathrm{B}ias(\hat{\theta}, \theta) = E(\hat{\theta}) - \theta$

Konfidenzintervall für den Mittelwert:
$\left[\bar{x} + \sigma_{\bar{x}} \cdot Q\left(\frac{\alpha}{2}\right); \bar{x} - \sigma_{\bar{x}} \cdot Q\left(\frac{\alpha}{2}\right)\right]$

Fehler 1. und 2. Art:

Testergebnis		tatsächliche Situation: H_0 falsch	H_0 wahr
	H_0 wird abgelehnt (positiv)	$1-\beta$, Power Sensitivität	α, Signifikanzniveau Fehler 1. Art
	H_0 wird angenommen (negativ)	β Fehler 2. Art	$1-\alpha$ Spezifität

Vergleich von Mittelwerten (unverbundene Stichproben)

$$\frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{\hat{\sigma}_A^2}{n_A} + \frac{\hat{\sigma}_B^2}{n_B}}} \sim t_{n_A + n_B - 2}$$

Vergleich von Mittelwerten (verbundene Stichproben)

$g = \frac{\bar{\Delta}}{\hat{\sigma}_{\bar{\Delta}}} \sim t_{n-1}$ mit $\Delta_i = X_i - Y_i$ und
$\hat{\sigma}_{\bar{\Delta}} = \frac{1}{\sqrt{n}}\sqrt{\frac{\sum_{i=1}^n (\Delta_i - \bar{\Delta})^2}{n-1}}$

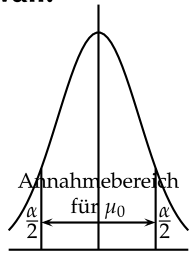
χ^2-Kontingenztest $e_{ij} = \frac{\sum_{j=1}^k x_{ij} \cdot \sum_{i=1}^n x_{ij}}{\sum_{i=1}^n \sum_{j=1}^k x_{ij}}$
$g = \sum_{i=1}^n \sum_{j=1}^k \frac{\left(x_{ij} - e_{ij}\right)^2}{e_{ij}} \sim \chi^2_{(n-1)\cdot(k-1)}$

χ^2-Anpassungstest: $g = \sum_{i=1}^k \frac{(X(a_i) - n \cdot P(a_i))^2}{n \cdot P(a_i)} \sim \chi^2_{k-1}$

Test von Mittelwerten: $g = \frac{\bar{x} - \mu_{X,0}}{\hat{\sigma}_{\bar{x}}}$ wobei $g \sim t_{n-1}$ falls X normalverteilt, und $g \sim N(0,1)$ falls $n \to \infty$

AIC $= -2 \cdot L + 2 \cdot k$ (dabei ist L die Likelihood des Modells und k die Anzahl der Parameter).

Konfidenzintervall:

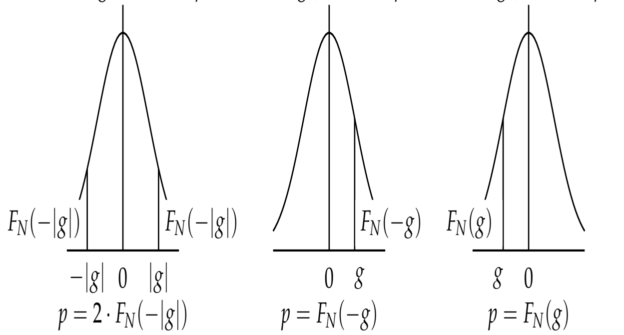


$\bar{x} + \frac{\sigma}{\sqrt{n}} \cdot Q_N(\frac{\alpha}{2})$ $\bar{x}$ $\bar{x} + \frac{\sigma}{\sqrt{n}} \cdot Q_N(1 - \frac{\alpha}{2})$

$H_0 : \bar{X} = \mu_0$ wird abgelehnt, falls μ_0 nicht im Annahmebereich liegt.

p**-Wert:** Teststatistik $g = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$

zweiseitig ($H_0 : \bar{X} = \mu_0$) einseitig ($H_0 : \bar{X} \leq \mu_0$) einseitig ($H_0 : \bar{X} \geq \mu_0$)



H_0 wird abgelehnt falls $p < \alpha$

	0.9	0.95	0.975	0.99	0.995	0.9975	0.999
qnorm(x)	1.28	1.64	1.96	2.33	2.58	2.81	3.09
qt(x,1)	3.08	6.31	12.71	31.82	63.66	127.32	318.31
qt(x,2)	1.89	2.92	4.30	6.96	9.92	14.09	22.33
qt(x,3)	1.64	2.35	3.18	4.54	5.84	7.45	10.21
qt(x,4)	1.53	2.13	2.78	3.75	4.60	5.60	7.17
qt(x,5)	1.48	2.02	2.57	3.36	4.03	4.77	5.89
qt(x,6)	1.44	1.94	2.45	3.14	3.71	4.32	5.21
qt(x,7)	1.41	1.89	2.36	3.00	3.50	4.03	4.79
qt(x,8)	1.40	1.86	2.31	2.90	3.36	3.83	4.50
qt(x,9)	1.38	1.83	2.26	2.82	3.25	3.69	4.30
qt(x,10)	1.37	1.81	2.23	2.76	3.17	3.58	4.14

	0.001	0.0025	0.005	0.01	0.025	0.05	0.1	0.9	0.95	0.975	0.99	0.995	0.9975	0.999
qchisq(x,1)	0.00	0.00	0.00	0.00	0.00	0.00	0.02	2.71	3.84	5.02	6.63	7.88	9.14	10.83
qchisq(x,2)	0.00	0.01	0.01	0.02	0.05	0.10	0.21	4.61	5.99	7.38	9.21	10.60	11.98	13.82
qchisq(x,3)	0.02	0.04	0.07	0.11	0.22	0.35	0.58	6.25	7.81	9.35	11.34	12.84	14.32	16.27
qchisq(x,4)	0.09	0.14	0.21	0.30	0.48	0.71	1.06	7.78	9.49	11.14	13.28	14.86	16.42	18.47
qchisq(x,5)	0.21	0.31	0.41	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09	16.75	18.39	20.52
qchisq(x,6)	0.38	0.53	0.68	0.87	1.24	1.64	2.20	10.64	12.59	14.45	16.81	18.55	20.25	22.46
qchisq(x,7)	0.60	0.79	0.99	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28	22.04	24.32
qchisq(x,8)	0.86	1.10	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.95	23.77	26.12
qchisq(x,9)	1.15	1.45	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59	25.46	27.88
qchisq(x,10)	1.48	1.83	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19	27.11	29.59