Technische Hochschule Köln
Fakultät für Wirtschafts- und Rechtswissenschaften
Prof. Dr. Arrenberg
Raum 221, Tel. 39 14
jutta.arrenberg@th-koeln.de

Wiederholung für die W-Statistik-Klausur am 03.07.2018

Aufgabe 1 (27.09.2010)
Ein international tätiges Unternehmen mit mehreren Niederlassungen in Deutschland und dem übrigen Europa hat seine überfälligen Forderungen in Millionen Euro (Mio. EUR) untersucht, nach Fälligkeit in Tagen eingeteilt und einander gegenüber gestellt.

Klasse Nr.	Fälligkeit von über ... bis maximal ... (in Tagen)	Höhe überfälliger Forderungen in Deutschland (in Mio. EUR)	Höhe überfälliger Forderungen im übrigen Europa (in Mio. EUR)
1	0 bis 30	24	12
2	30 bis 90	48	36
3	90 bis 180	72	66
4	180 bis 360	16	6
Summe		160	120

a) Wie lange sind 30 Prozent der Forderungen in Deutschland höchstens überfällig?

b) Wie viel Prozent der Forderungen im übrigen Europa sind länger als 225 Tage überfällig?

c) Ermitteln Sie unter Zugrundelegung des arithmetischen Mittels, ob die Forderungen in Deutschland oder die Forderungen im übrigen Europa durchschnittlich länger überfällig sind.

d) In welcher der beiden Regionen schwankt die Fälligkeit der Forderungen stärker? Begründen Sie Ihre Antwort durch Berechnung einer geeigneten statistischen Maßzahl.

e) Wie ist die Spannweite definiert? Kann diese bei den vorliegenden klassierten Daten als Parameter verwendet werden? Begründen Sie bitte kurz Ihre Antwort.

Aufgabe 5 (16.07.2002)
Die Bundestagswahl im September 2002 wirft ihre Schatten voraus. Im Auftrag des Fernsehsenders „n-tv“ führt das Meinungsforschungsinstitut „Emnid“ aus Bielefeld in regelmäßigen Abständen repräsentative Befragungen von rund 1 000 zufällig ausgewählten Wahlberechtigten in Deutschland durch und lässt sich dabei immer auch die berühmte „Sonntagsfrage“ beantworten: „Welcher Partei würden Sie Ihre Stimme geben, wenn am nächsten Sonntag Bundestagswahlen wären?“
Normalerweise liegen dabei die großen Volksparteien SPD und CDU/CSU ungefähr in einem Bereich von 30 bis 40 % derjenigen Wahlberechtigten, die sich bei der Umfrage für eine bestimmte Partei entscheiden.

a) Wie groß müsste eine Zufallsstichprobe (ohne Berücksichtigung der Antwortverweigerer, die auf die Sonntagsfrage nicht antworten bzw. angeben, nicht zur Wahl zu gehen bzw. sich noch nicht entschieden zu haben) sein, damit man mit einer Sicherheit von rund 95% davon ausgehen könnte, dass eine Partei bei allen Wahlberechtigten einen Stimmenanteil zwischen 34 und 36 % hat, wenn sich in der Zufallsstichprobe genau 35 % für diese Partei aussprechen?

b) Normalerweise sind rund 25 % der zufällig für dei Zufallsstichprobe ausgewählten Wahlberechtigten Antwortverweigerer [im Sinne der Darstellung unter a)], wobei es in der Natur der Sache liegt, dass dieser Anteil keine Konstante ist; gehen Sie indessen bei der Beantwortung von Frage b) der Einfachheit halber von einem Erfahrungssatz von 25 % Antwortverweigerern aus.
Wie müsste man die Größe der Zufallsstichprobe, die Sie unter a) berechnet haben, verändern, um bei sonst gleichen Voraussetzungen (Sicherheit: 95 %; Fehlerspielraum: 34 bis 36 %) die Antwortverweigerer zu berücksichtigen?

c) Nehmen Sie an, dass sich bei der Zufallsstichprobe [auf Basis der Teilaufgabe a); das heißt: verwenden Sie den dort berechneten Stichprobenumfang] 18% derjenigen, die eine Partei ihrer Wahl angeben, für die FDP votieren. Wie groß ist dann - wiederum bei einer Sicherheit von rund 95% - der Fehlerspielraum für den Stimmenanteil der FDP? Begründen Sie bitte Ihre Antwort durch Berechnung und anschließende Interpretation des Ergebnisses!

Wiederholung für die W-Statistik-Klausur 10/2018

Aufgabe 1 (02.02.2016)
Zwischen dem Absatz X und dem Verkaufspreis Y eines Produkts wird ein linearer Zusammenhang vermutet. Bei den letzten vier Preisveränderungen (in GE) ergaben sich folgende Absätze (in ME):

Absatz	Preis
25	4
21	5
20	7
16	8

a) Welcher Preis ist anzusetzen, um mit einen Absatz von 18 ME rechnen zu können?

b) Mit welchem Absatz ist zu rechnen, wenn der Preis auf 9 GE erhöht wird? Interpretieren Sie die Steigung der Regressionsgeraden, die zu dieser Prognose herangezogen wird.

c) Beurteilen Sie die Stärke des linearen Zusammenhangs anhand einer geeigneten Maßzahl.

d) Sind die Berechnungen unter den Teilaufgaben a) und b) aus statistischer Sicht zuverlässig? (Begründung!)

Aufgabe 1 (20.07.2017)

a) Es ist bekannt, dass die Tagesrendite einer Aktie (Angabe in %) normalverteilt ist mit Erwartungswert 0,1% und Standardabweichung 0,7%.

1. Bestimmen Sie die Wahrscheinlichkeit, dass die Tagesrendite dieser Aktie höher als 1,5% ist!
2. Welcher konkrete Wert der Tagesrendite wird mit einer Wahrscheinlichkeit von 99% nicht unterschritten?
3. Wie groß ist die Wahrscheinlichkeit, dass die Tagesrendite an drei Tagen genau zweimal höher als 1,5% ist, wenn die Tagesrenditen stochastisch unabhängig voneinander sind?

c) Die Tagesrendite einer weiteren Aktie sei ebenfalls normalverteilt und es sei ferner das zweifache Schwankungsintervall bekannt. Die untere Grenze dieses Intervalls ist gegeben durch $-1{,}9886$ und die obere Grenze durch 2,3886.
Bestimmen Sie aus diesen Angaben den Erwartungswert und die Varianz der Tagesrendite dieser zweiten Aktie!

Lösung zu Aufgabe 1: (27.09.2010)
X = Fälligkeit (in Tagen) einer Forderung in Höhe von 1 Mio. EUR in Deutschland
Y = Fälligkeit (in Tagen) einer Forderung in Höhe von 1. Mio. EUR im übrigen Europa
n_j^x = Anzahl der Forderungen über eine Mio. EUR in der j-en Klasse in Deutschland
n_j^y = Anzahl der Forderungen über eine Mio. EUR in der j-en Klasse im übrigen Europa

j	$x_{j-1}^* < x \leq x_j^*$	x_j'	b_j	n_j^x/n	$F(x_j^*)$	n_j^y/n	$F(y_j^*)$
1	$0-30$	15	30	0,15	0,15	0,10	0,10
2	$30-90$	60	60	0,30	0,45	0,30	0,40
3	$90-180$	135	90	0,45	0,90	0,55	0,95
4	$180-360$	270	180	0,10	1,00	0,05	1,00
$\sum$				1,00		1,00	

a) $x_{0{,}30} \approx 30 + \dfrac{0{,}30-0{,}15}{0{,}30} \cdot 60 = 60$
d.h. etwa 30% aller Forderungen in Deutschland sind höchstens 60 Tage überfällig.

b) $F(225) \approx 0{,}95 + \dfrac{0{,}05}{180}(225-180) = 0{,}9625$
$100\% - 96{,}25\% = 3{,}75\%$
d.h. etwa 4% aller Forderungen im übrigen Europa sind länger als 225 Tage überfällig.

c) $\overline{y} \approx \dfrac{1}{120}\left[15 \cdot 12 + 60 \cdot 36 + 135 \cdot 66 + 270 \cdot 6\right] = 107{,}25$
d.h. im übrigen Europa ist eine Forderung in Höhe von einer Mio. EUR im Durchschnitt etwa 107 Tage überfällig.
$\overline{x} \approx \dfrac{1}{160}\left[15 \cdot 24 + 60 \cdot 48 + 135 \cdot 72 + 270 \cdot 16\right] = 108$
d.h. in Deutschland ist eine Forderung in Höhe von einer Mio. EUR im Durchschnitt 108 Tage überfällig und damit etwas länger überfällig als im übrigen Europa.

d)
$$\begin{aligned} s_x^2 &\approx \frac{1}{160}\Big[(15-108)^2 \cdot 24 + (60-108)^2 \cdot 48 + (135-108)^2 \cdot 72 \\ &\quad +(270-108)^2 \cdot 16\Big] \\ &= 4\,941 \end{aligned}$$

$$\begin{aligned} s_y^2 &\approx \frac{1}{120}\big[(15-107{,}25)^2\cdot 12+(60-107{,}25)^2\cdot 36+(135-107{,}25)^2\cdot 66 \\ &\quad +(270-107{,}25)^2\cdot 6\big] \\ &= 3\,268{,}688 \end{aligned}$$

d.h. gemessen mit der Varianz sind die Unterschiede der Fälligkeitsdauern der Forderungen in Deutschland stärker.

e) Die Spannweite ist die Differenz aus dem größten beobachteten Wert und dem kleinsten beobachteten Wert. Da bei klassierten Daten (wie hier im Beispiel) nicht mehr die Einzelwerte vorliegen, sondern nur noch die Klassen, in die die Einzelwerte fallen, können hier weder ein größter noch ein kleinster beobachteter Wert bestimmt werden und somit kann die Spannweite nicht berechnet werden.

Lösung: (16.07.2002)

a) Auf Grund einer Stichprobe beträgt das 0,95-Konfidenzintervall für den Anteilswert laut Aufgabentext [0,34; 0,36]. Also beträgt die Breite des Konfidenzintervalls 0,02 und die halbe Breite $e = 0{,}01$. Ferner beträgt der Anteilswert in der Stichprobe laut Aufgabentext 0,35. Gesucht ist der Mindeststichprobenumfang n des 0,95-Konfidenzintervalls für den Anteilswert:

$$n \geq \frac{(1{,}96)^2\cdot 0{,}35\cdot 0{,}65}{(0{,}01)^2} = 8739{,}64$$

d.h. der Mindeststichprobenumfang beträgt etwa 8740.
Das in der Aufgabe gegebene Konfidenzintervall [0,34; 0,36] ist ein Konfidenzintervall für den Anteilswert. Also müssen wir einerseits prüfen, ob wir die Binomialverteilung durch die Normalverteilung annähern dürfen. Die Faustregel $n \geq 100$ ist erfüllt. Und andererseits müssen wir überprüfen, ob überhaupt eine Binomialverteilung vorliegt, weil es sich bei einer Umfrage um eine Stichprobe ohne Zurücklegen handelt. Die Faustregel dazu $\frac{n}{N} = \frac{8740}{40\,000\,000} \leq 0{,}05$ stimmt. (Die Anzahl aller Wahlberechtigten in Deutschland beträgt etwa 40 Millionen.)

b) Dreisatz:

$$\begin{aligned} 75\% &\;\widehat{=}\; 8\,740 \text{ Befragte} \\ 100\% &\;\widehat{=}\; \frac{8\,740 \text{ Befragte}}{75\%}\cdot 100\% = 11\,653{,}3 \text{ Befragte} \end{aligned}$$

d.h. der Stichprobenumfang müsste etwa 11 650 betragen.

c) Gesucht ist ein 0,95-Konfidenzintervall für den Stimmenanteil der FDP-Wähler bei der Wahl. Der Stimmenanteil der FDP-Wähler in der Stichprobe beträgt 0,18. Somit ergibt sich folgendes Konfidenzintervall:

$$\left[0{,}18-1{,}96\cdot\sqrt{\frac{0{,}18\cdot 0{,}82}{8740}};\,0{,}18+1{,}96\cdot\sqrt{\frac{0{,}18\cdot 0{,}82}{8740}}\right] = [0{,}1719; 0{,}1881]$$

d.h. das Konfidenzintervall beträgt etwa [17,2%; 18,8%].
Interpretation: Das berechnete Konfidenzintervall ist ein geschätzter Bereich für das Intervall, in dem der FDP-Stimmenanteil in der Bevölkerung mit der Wahrscheinlichkeit 0,95 liegt.

Lösung zu Aufgabe 1 (02.02.2016)

i	x_i	y_i	$x_i \cdot y_i$	x_i^2	y_i^2
1	25	4			
2	21	5			
3	20	7			
4	16	8			
$\sum$	82	24	473	1 722	154

a) Gesucht $a_1 + b_1 \cdot 18 =?$

$$b_1 = \frac{4 \cdot 473 - 82 \cdot 24}{4 \cdot 1\,722 - 82^2} = \frac{-76}{164} = -0{,}4634146 \approx -0{,}46$$

$$a_1 = \frac{24 - (-0{,}46) \cdot 82}{4} = 15{,}43$$

$15{,}43 - 0{,}46 \cdot 18 = 7{,}15$

d.h. es ist ein Preis von etwa 7 GE anzusetzen.

b) Gesucht $a_2 + b_2 \cdot 9 =?$

$$b_2 = \frac{-76}{4 \cdot 154 - 24^2} = \frac{-76}{40} = -1{,}9$$

d.h. steigt der Preis um eine GE, so sinkt der Absatz um etwa 1,9 ME.

$$a_2 = \frac{82 - (-1{,}9) \cdot 24}{4} = 31{,}9$$

$31{,}9 - 1{,}9 \cdot 9 = 14{,}8 \approx 15$

d.h. es ist mit einem Absatz von etwa 15 ME zu rechnen.

c) $r = -\sqrt{(-0{,}46) \cdot (-1{,}9)} = -\sqrt{0{,}874} = -0{,}9348797 \approx -0{,}93$

d.h es liegt eine starke (negative) Korrelation vor.

d) Auf den Prognosewert aus Teilaufgabe a) ist Verlass, da es sich um einen interpolierten Wert ($18 \in [16; 25]$) bei gleichzeitig starker Korrelation handelt. Auf den Prognosewert aus Teilaufgabe b) ist kein Verlass, da es sich um einen extrapolierten Wert ($9 \notin [4; 8]$) handelt.

Lösung zu Aufgabe 1: (20.07.2017)

a) X=Tagesrendite (in %)

$X \sim \mathsf{N}(\mu = 0{,}1; \sigma = 0{,}7)$

1. $P(X > 1{,}5) = 1 - P(X \leq 1{,}5) = F_U\left(\frac{1{,}5 - 0{,}1}{0{,}7}\right) = 1 - F_U(2) = 1 - 0{,}977 = 0{,}023 = 2{,}3\%$
2. $0{,}01 = P(X \leq x) \Leftrightarrow -2{,}3263 = \frac{x - 0{,}1}{0{,}7} \Leftrightarrow x = 0{,}1 - 2{,}3263 \cdot 0{,}7 = -1{,}52841$

   d.h. die gesuchte Tagesrendite beträgt $-1{,}52841\%$.
3. Z = Anzahl der Tage, an denen die Rendite über 1,5% liegt

   $Z \sim \mathsf{B}(n = 3; p = 0{,}023)$

   $P(Z = 2) = \binom{3}{2} \cdot 0{,}023^2 \cdot 0{,}977 = 0{,}0016$

   d.h. die Wahrscheinlichkeit beträgt 0,0016.

b) Y=Tagesrendite (in %)

$Y \sim \mathsf{N}(\mu;\sigma)$

$[\mu - 2\cdot\sigma; \mu + 2\cdot\sigma] = [-1{,}9886; +2{,}3886]$

1. Lösungsweg:

$$\begin{array}{lrcl}
\text{I} & -1{,}9886 & = & \mu - 2\cdot\sigma \\
\text{II} & 2{,}3886 & = & \mu + 2\cdot\sigma \\
\hline
\text{II} - \text{I} & 4{,}3772 & = & 4\cdot\sigma \Leftrightarrow \sigma = 1{,}0943 \Leftrightarrow \sigma^2 = 1{,}0943^2 = 1{,}1975 \\
\text{I} & -1{,}9886 & = & \mu - 2\cdot 1{,}0943 \Leftrightarrow \mu = 2\cdot 1{,}0943 - 1{,}9886 = 0{,}2
\end{array}$$

2. Lösungsweg:

μ=Intervallmitte = (Untergrenze + Obergrenze) $\div$ 2 = $(-1{,}9886 + 2{,}3886) \div 2 = 0{,}2$

Intervallänge = Obergrenze minus Untergrenze $= 4\cdot\sigma = 2{,}3886 - (-1{,}9886) = 4{,}3772 \Leftrightarrow \sigma = \frac{4{,}3772}{4} = 1{,}0943 \Leftrightarrow \sigma^2 = 1{,}0943^2 = 1{,}1975$