

# Statistik-Klausur vom 16. April 2004

Bearbeitungszeit: 90 Minuten

## Aufgabe 1

Getränkekonsum in Deutschland von 1997 bis 2002

Jahr	Getränkekonsum (in Litern pro Person und Jahr)		
	alkoholhaltig	alkoholfrei	
1997	160	240	
1998	156	241	
1999	156	248	
2000	154	253	
2001	152	262	
2002	152	271	

Quelle: Institut der deutschen Wirtschaft Köln: iwd, Nr. 34, 21. August 2003

- Wie hoch dürfte der Konsum alkoholhaltiger Getränke (in Litern pro Person und Jahr) in Deutschland im Jahre 2005 sein?
- Beschreiben Sie verbal (d.h. ohne Berechnungen), welcher Zusammenhang zwischen der Höhe des Konsums alkoholhaltiger Getränke einerseits und der Höhe des Konsums alkoholfreier Getränke andererseits besteht.
- Welche statistische Maßzahl ist geeignet, um die Stärke des linearen Zusammenhangs zwischen der Höhe des Konsums alkoholhaltiger Getränke einerseits und der Höhe des Konsums alkoholfreier Getränke andererseits zu kennzeichnen? (Hinweis: Die verbale Benennung dieser Maßzahl mit Begründung genügt; Berechnungen werden nicht verlangt.)

## Aufgabe 2

Das Statistische Bundesamt in Wiesbaden veröffentlicht in regelmäßigen Abständen den Verbraucherpreisindex. Für die Jahre 1998 bis 2002 ergeben sich folgende Werte:

Jahr	Verbraucherpreisindex (2000 = 100)
1998	98,0
1999	98,6
2000	100,0
2001	102,0
2002	103,4

- Um wie viel Prozent sind die Verbraucherpreise zwischen 1998 und 2002 gestiegen?

- b) Um wie viel Prozent sind die Verbraucherpreise zwischen 1998 und 2002 durchschnittlich pro Jahr gestiegen?
- c) Wie hoch ist der Kaufkraftverlust im Zeitraum 1998 bis 2002?
- d) Welche Werte nimmt der Verbraucherpreisindex für die Jahre 1999 bis 2002 an, wenn man als Basisjahr das Jahr 1998 wählt?

### Aufgabe 3

Autofarbe der 2003 in Deutschland zugelassenen Pkw

Farbe	silber/grau	schwarz	blau	grün	rot	weiß	sonstige
Anteil	44,9%	22,6%	18,6%	5,0%	4,6%	2,4%	1,9%

Quelle: ADACmotorwelt 3/2004

- a) Wie viel Prozent der 2003 zugelassenen Pkw waren entweder silber/grau oder schwarz oder blau?
- b) Ein Pkw wurde 2003 zugelassen und hat einen Unfall verursacht. Wie groß muss für diesen Pkw die Wahrscheinlichkeit sein, dass er schwarz lackiert ist, damit für die 2003 zugelassenen Pkw die Ereignisse:

$A$  = Pkw-Farbe ist schwarz

$B$  = Pkw verursacht einen Unfall

stochastisch unabhängig sind?

### Aufgabe 4

Bei einer Flugreise wird Reisegepäck bis zu 20 kg kostenlos befördert. Reisegepäck über 20 kg gilt als Übergepäck und kann nur mit Zustimmung der Fluggesellschaft und gegen Bezahlung mitgenommen werden.

- Die Fluggäste  $A, B, C, D, E$  hatten folgendes Reisegepäck:  
18 kg, 15 kg, 21 kg, 17 kg, 19 kg.
  - Wie hoch ist das Durchschnittsgewicht dieser fünf Reisegepäck?
  - Wie hoch ist die Standardabweichung dieses Datensatzes?
- Nehmen Sie an, das Reisegepäck (in kg) eines Fluggastes ist normalverteilt mit dem Erwartungswert von 18 kg und einer Standardabweichung von 2 kg.
  - Ein Fluggast wird zufällig ausgewählt. Wie hoch ist die Wahrscheinlichkeit, dass dies ein Fluggast mit Übergepäck ist?
  - Wie groß ist die Wahrscheinlichkeit, unter 20 Fluggästen genau drei Fluggäste mit Übergepäck zu haben?
  - Ein Fluggast wird zufällig ausgewählt. Berechnen Sie die Wahrscheinlichkeiten dafür, dass dieser Fluggast ein Reisegepäck hat zwischen
    - 21 und 22 kg

- 22 und 23 kg
- 23 und 24 kg
- 24 und 25 kg

Für jedes volle Kilogramm über 20 kg Reisegepäck sind jeweils 5 € zu bezahlen. Mit welchen zusätzlichen Einnahmen pro Fluggast kann die Fluggesellschaft in etwa rechnen?

### Aufgabe 5

In einer Stadt mit 100 000 Einwohnern soll der Anteil der Raucher an der Gesamtbevölkerung geschätzt werden. Das Gesundheitsamt der Stadt verlangt, dass die durchzuführende Studie zu Ergebnissen kommt, die mit einer Wahrscheinlichkeit von 95% zutreffen. Bei einer Umfrage unter 897 Einwohnern befanden sich genau 269 Raucher.

Schätzen Sie anhand der Umfrage den Anteil der Raucher in der Stadt, wenn die oben genannte Anforderung erfüllt werden soll. Und interpretieren Sie Ihr Ergebnis.

### Lösungen:

*Lösung von Aufgabe 1:*

a) Lineare Regression

$x$  = Zeit = unabhängige Variable

$y$  = Alkoholkonsum (in l pro Kopf) = abhängige Variable

Regressionsgerade  $f(x) = a_1 + b_1 \cdot x$

$x_i$	$y_i$	$x_i \cdot y_i$	$x_i^2$
1	160	160	1
2	156		4
3	156		9
4	154		16
5	152		25
6	152		36
21	930	3 228	91

Jahr 1997  $\hat{=}$  Zeitpunkt 1

Jahr 2005  $\hat{=}$  Zeitpunkt 9

Gesucht:  $f(9) = a_1 + b_1 \cdot 9 = ?$

$$b_1 = \frac{6 \cdot 3\,228 - 21 \cdot 930}{6 \cdot 91 - 21^2} = -1,542857$$

$$a_1 = \frac{930 - (-1,542857) \cdot 21}{6} = 160,4$$

$$f(9) = 160,4 - 1,542857 \cdot 9 = 146,5143$$

d.h. im Jahr 2005 ist gemäß der Methode der kleinsten Quadrate mit einem Alkoholkonsum von etwa 147 Litern pro Kopf zu rechnen.

- b)  $x =$  Alkoholkonsum (in l pro Kopf)  
 $y =$  Konsum alkoholfreier Getränke (in l pro Kopf)  
 Große  $x$ -Werte gehen mit kleinen  $y$ -Werten einher. Und kleine  $x$ -Werte gehen mit großen  $y$ -Werten einher. Es besteht ein linearer negativer Zusammenhang (negative Korrelation) zwischen den  $x$ - und  $y$ -Werten; d.h. je höher der Alkoholkonsum desto geringer ist der Verbrauch an alkoholfreien Getränken.
- c) Es kann das Bestimmtheitsmaß  $B$  berechnet werden.  $B$  liegt zwischen 0 und 1. Je näher der Wert von  $B$  bei 1 liegt desto stärker ist der lineare Zusammenhang. Es kann der Korrelationskoeffizient  $r$  berechnet werden,  $r$  liegt zwischen  $-1$  und  $+1$ . Je näher  $r$  an  $-1$  liegt, desto höher ist hier der negative lineare Zusammenhang.

*Lösung von Aufgabe 2:*

- a)  $\frac{103,4}{98,0} = 1,0551$   
 d.h. im Zeitraum von 1998 bis 2002 sind die Verbraucherpreise um insgesamt 5,5% gestiegen.
- b)  $\sqrt[2002-1998]{1,0551} = \sqrt[4]{1,0551} = 1,0135$   
 d.h. im Zeitraum von 1998 bis 2002 sind die Verbraucherpreise um durchschnittlich 1,4% pro Jahr gestiegen.
- c) Kaufkraft =  $\frac{1}{\text{Preisindex}} = \frac{1}{1,0551} = 0,9478$   
 d.h. im Zeitraum von 1998 bis 2002 ist die Kaufkraft um insgesamt 5,2% gesunken.

d)

Jahr	Index
1998	$\frac{98,0}{98,0} \cdot 100 = 100,0$
1999	$\frac{98,6}{98,0} \cdot 100 = 100,6$
2000	$\frac{100,0}{98,0} \cdot 100 = 102,0$
2001	$\frac{102,0}{98,0} \cdot 100 = 104,1$
2002	$\frac{103,4}{98,0} \cdot 100 = 105,5$

*Lösung zu Aufgabe 3:*

- a)  $44,9 + 22,6 + 18,6 = 86,1$   
 d.h. der Anteil beträgt 86,1%.

- b) Gemäß Tabelle beträgt  $P(A) = 0,226$ . Also muss  $P(A | B)$  ebenfalls 0,226 betragen, damit  $A, B$  stochastisch unabhängig sind.

*Lösung zu Aufgabe 4:*

1. a)  $\bar{x} = \frac{1}{5}[18 + 15 + 21 + 17 + 19] = 18$   
d.h. im Durchschnitt wog das Reisegepäck 18 kg.
- b)  $s^2 = \frac{1}{5}[(18 - 18)^2 + (15 - 18)^2 + (21 - 18)^2 + (17 - 18)^2 + (19 - 18)^2] =$   
 $\frac{1}{5}[0 + 9 + 9 + 1 + 1] = 4$   
 $s = \sqrt{4} = 2$   
d.h. die empirische Standardabweichung beträgt 2 kg.

2. a)  $X =$  Reisegepäck (in kg) eines Fluggastes

$$X \sim \mathbf{N}(\mu = 18; \sigma = 2)$$

$$P(X > 20) = 1 - P(X \leq 20) = 1 - F_U\left(\frac{20 - 18}{2}\right) = 1 - F_U(1) =$$

$$1 - 0,841 = 0,159$$

d.h. die Wahrscheinlichkeit beträgt etwa 0,16.

- b)  $Y =$  Anzahl der Fluggäste mit Übergepäck

$$Y \sim \mathbf{B}(n = 20; p = 0,159)$$

$$P(Y = 3) = \binom{20}{3} \cdot 0,159^3 \cdot 0,841^{17} = 0,2413$$

d.h. die Wahrscheinlichkeit beträgt etwa 0,24.

- c)  $P(X \leq 22) - P(X \leq 21) = F_U\left(\frac{22-18}{2}\right) - F_U\left(\frac{21-18}{2}\right) = F_U(2) - F_U(1,5)$

$$= 0,977 - 0,933 = 0,044$$

$$P(X \leq 23) - P(X \leq 22) = F_U\left(\frac{23-18}{2}\right) - 0,977 = F_U(2,5) - 0,977$$

$$= 0,994 - 0,977 = 0,017$$

$$P(X \leq 24) - P(X \leq 23) = F_U\left(\frac{24-18}{2}\right) - 0,994 = F_U(3) - 0,994$$

$$= 0,999 - 0,994 = 0,005$$

$$P(X \leq 25) - P(X \leq 24) = F_U\left(\frac{25-18}{2}\right) - 0,999 = F_U(3,5) - 0,999$$

$$\approx 1 - 0,999 = 0,001$$

$Z =$  Anzahl der vollen kg über 20 kg Reisegepäck eines Fluggastes

$z$	0	1	2	3	4	5 oder mehr
$P(Z = z)$	0,933	0,044	0,017	0,005	0,001	$\approx 0$

$$0 \cdot 0,933 + 5 \cdot 0,044 + 10 \cdot 0,017 + 15 \cdot 0,005 + 20 \cdot 0,001 = 0,485$$

d.h. pro Fluggast ist mit einer Abgabe von 48,5 Cent für Übergepäck zu rechnen.

*Lösung von Aufgabe 5*

$p =$  Anteil der Raucher in der Stadt

$$\hat{p} = \frac{269}{897} \approx 0,30 = \text{Anteil der Raucher in der Stichprobe}$$

Stichprobenumfang:  $n = 897$

Faustregel:  $n \geq 100$  ist erfüllt

0,95-KI für  $p$ :

$$\left[ 0,3 \pm 1,96 \cdot \sqrt{\frac{0,3 \cdot 0,7}{897}} \right] = [0,3 \pm 0,03] = [0,27; 0,33]$$

d.h.  $[0,27; 0,33]$  ist ein geschätztes Intervall für den Bereich, in dem der Anteil der Raucher in der Stadt mit einer Wahrscheinlichkeit von 0,95 liegt.