

# **Lösungsvorschläge**

**Klausur**

**Statistische Methodenlehre I**

**Daniel Lambert**

**Aufgabe 1:**

Man möchte eine Trendgerade der Form  $y = a + b \cdot t$  kalkulieren. Hierfür benötigt man die Parameter  $a$  und  $b$ . Für die Methode der Kleinsten Quadrate berechnet man zunächst eine Hilfstabelle, um dann die Zahlen  $a$  und  $b$  kalkulieren zu können.

<b>t</b>	<b>y</b>	<b>t·y</b>	<b>t<sup>2</sup></b>
1	5	5	1
2	7	14	4
3	6	18	9
4	8	32	16
5	9	45	25
6	11	66	36
$\Sigma = 21$	$\Sigma = 46$	$\Sigma = 180$	$\Sigma = 91$

Das arithmetische Mittel der Zeitwerte ist

$$\bar{t} = \frac{1}{n} \cdot \sum t = \frac{1}{6} \cdot 21 = 3,5.$$

Für die  $y$ -Daten erhält man analog  $\bar{y} = \frac{1}{n} \cdot \sum_{i=1}^n y_i = \frac{1}{6} \cdot 46 = 7,67$ .

Man rechnet schließlich das Steigungsmaß  $b$  aus als

$$b = \frac{\frac{1}{n} \cdot \sum t \cdot y - \bar{t} \cdot \bar{y}}{\frac{1}{n} \cdot \sum t^2 - \bar{t}^2} = \frac{\frac{1}{6} \cdot \sum 180 - 3,5 \cdot 7,67}{\frac{1}{6} \cdot \sum 91 - 3,5^2} = 1,0857.$$

Hiermit geht man in die Formel für den Ordinatenabschnitt  $a$ :

$$a = \bar{y} - b \cdot \bar{t} = 7,67 - 1,0857 \cdot 3,5 = 3,867.$$

Damit lautet der Trend der Zeitreihe mit der Methode der Kleinsten Quadrate schließlich

$$y(t) = 3,867 + 1,0857 \cdot t.$$

Den Trendwert für das Jahr 2003 berechnet man durch Einsetzen der Zahl  $t = 3$  in die obige Trendfunktion:

$$y(3) = 3,867 + 1,0857 \cdot 3 = 7,124.$$

### Aufgabe 2:

Ein Fall für die Tschebyscheffsche Ungleichung.

#### LAMBERT-KOCHREZEPT TSCHEBYSCHEFFSCHE UNGLEICHUNG:

Voraussetzungen

- die konkrete Verteilung ist unbekannt
- Erwartungswert  $\mu$  und Varianz  $\sigma^2$  hingegen sind bekannt.
- es ist gefragt nach einer **Mindest-** oder nach einer **Höchstwahrscheinlichkeit**, nicht danach, eine Wahrscheinlichkeit exakt zu berechnen.
- der fragliche Bereich, über den eine Wahrscheinlichkeitsaussage gemacht werden soll, liegt *symmetrisch* um den Erwartungswert  $\mu$  herum. Rechts und links um den Erwartungswert herum beträgt die Länge des fraglichen Bereichs genau  $c$ .

#### Anwendung der Ungleichung

##### 1. Fall

- verbal: dann ist die Wahrscheinlichkeit, dass die Zufallsvariable  $X$  vom eigenen Erwartungswert  $\mu$  um höchstens  $c$  abweicht, mindestens  $1 - \frac{\sigma^2}{c^2}$
- in Formeln:  $W(|X - \mu| < c) \geq 1 - \frac{\sigma^2}{c^2}$

##### 2. Fall

- verbal: die Wahrscheinlichkeit, dass die Zufallsvariable  $X$  vom eigenen Erwartungswert um mindestens  $c$  abweicht, beträgt höchstens  $\frac{\sigma^2}{c^2}$
- in Formeln:  $W(|X - \mu| > c) \leq \frac{\sigma^2}{c^2}$ .

Dass die konkrete Verteilung hier sogar bekannt ist, bedeutet, dass man die Wahrscheinlichkeit exakt berechnen könnte. Da allerdings nach einer Abschätzung gefragt ist, nehmen wir Tschebyscheff. Also konkret:

$W(|X - \mu| < c) \geq 1 - \frac{\sigma^2}{c^2}$ , d.h.  $c = 2$ . Die Varianz und der Erwartungswert einer F-verteilten Zufallsvariablen sind bekannt:

$X \sim F(m,n)$ , also

$$E(X) = \frac{n}{n-2} = \frac{100}{98} = 1,0204 \text{ und}$$

$$\begin{aligned} \text{Var}(X) &= \frac{2 \cdot n^2 \cdot (n+m-2)}{m \cdot (n-4) \cdot (n-2)^2} = \frac{2 \cdot 100^2 \cdot (100+100-2)}{100 \cdot (100-4) \cdot (100-2)^2} \\ &= \frac{396}{921.984} = 0,0004295. \end{aligned}$$

Damit rechnet man also

$$W(|X - \mu| < c) \geq 1 - \frac{\sigma^2}{c^2}$$

$\Leftrightarrow W(|X - \mu| < 2) \geq 1 - \frac{0,0004295}{2^2} = 0,999893$ . Man erhält also eine untere Abschätzung für die gesuchte Wahrscheinlichkeit von 0,999893.

### Aufgabe 3:

Die Zufallsvariable  $X$  ist normalverteilt zu einem Erwartungswert von  $\mu = 100$  und einer Standardabweichung von  $\sigma = 4^{0,5} = 2$ .

a) Daher rechnet man

$$\begin{aligned} W(96 \leq X \leq 104) &= \Phi\left(\frac{104-100}{4^{0,5}}\right) - \Phi\left(\frac{96-100}{4^{0,5}}\right) \\ &= \Phi(2) - \Phi(-2) \\ &= \Phi(2) - [1 - \Phi(-2)] \end{aligned}$$

$$\begin{aligned}
 &= 2 \cdot \Phi(2) - 1 \\
 &= 2 \cdot 0,977250 - 1 \\
 &= 0,9545.
 \end{aligned}$$

b) Gefragt ist nach der Wahrscheinlichkeit, dass die Zufallsvariable echt kleiner ist als 94:

$$\begin{aligned}
 W(X < 94) &= W(X \leq 94) \\
 &= \Phi\left(\frac{94 - 100}{4^{0,5}}\right) \\
 &= \Phi(-3) \\
 &= 1 - \Phi(3) \\
 &= 1 - 0,99865 \\
 &= 0,00135.
 \end{aligned}$$

#### **LAMBERT-METHODE:**

Bei stetigen Zufallsvariablen sind die Wahrscheinlichkeiten  $W(X < a)$  und  $W(X \leq a)$  beide identisch. Deswegen hatten wir oben auch  $W(X < 94) = W(X \leq 94)$  geschrieben.

c) Es gilt folgendes.

#### **LAMBERT-REGEL:**

Die Wahrscheinlichkeit der Gleichheit ist bei stetigen Zufallsvariablen stets null, d.h. es gilt  $W(X = a) = 0$ , wenn  $X$  stetig verteilt ist.

So auch hier, denn die Normalverteilung ist eine stetige Verteilung:

$$W(X = 100) = 0.$$

d) Man interessiert sich für die Frage, welches Geldmengenwachstum  $a$  in 97,5 % der Fälle unterschritten wird. Dafür rechnet man also

$$W(X \leq a) = 0,975 \Leftrightarrow \Phi\left(\frac{a - 100}{4^{0,5}}\right) = 0,975.$$

Alsdann schlägt man das 0,975-Fraktile in der Tabelle der Standardnormalverteilung nach und sieht  $x_{0,975;N(0,1)} = 1,96$ . Dies ist also jene Zahl,

die in der Klammer stehen muss, damit die Wahrscheinlichkeit bei 0,975

liegt:  $\frac{a-100}{4^{0,5}} = 1,96$ . Aufgelöst erhält man folglich  $a = 1,96 \cdot 2 + 100 =$

103,92, d.h. das 0,975-Fraktile der o.e. Verteilung ist 103,92, in Zeichen:  $X_{0,975;N(100,2)} = 103,92$ .

e) Man rechnet

$$W(X < 105) = 0,975 \quad \Leftrightarrow \Phi\left(\frac{105 - \mu}{\sigma}\right) = 0,975$$

$$\Leftrightarrow \frac{105 - \mu}{\sigma} = 1,96$$

$$\begin{aligned} \Leftrightarrow \mu &= 105 - 1,96 \cdot \sigma \\ &= 105 - 1,96 \cdot 2 \\ &= 101,08. \end{aligned}$$

#### Aufgabe 4:

Von der zweidimensionalen Verteilung ist zunächst lediglich folgendes bekannt:

X/Y	Y = 1	Y = 2	Σ
X = 2	0,2		0,4
X = 3			0,4
X = 4			
Σ			1

Die beiden bedingten Wahrscheinlichkeiten liefern uns die nächsten beiden Felder:

$$W(Y=1|X=3) = \frac{W(Y=1, X=3)}{W(X=3)}, \text{ mithin}$$

$$\begin{aligned} W(Y=1, X=3) &= W(Y=1|X=3) \cdot W(X=3) \\ &= 0,2 \cdot 0,4 \\ &= 0,08. \end{aligned}$$

Genauso kalkuliert man

$$\begin{aligned}
 W(Y=2, X=4) &= W(Y=2|X=4) \cdot W(X=4) \\
 &= 0,3 \cdot [1 - W(X=2) - W(X=3)] \\
 &= 0,3 \cdot 0,2 \\
 &= 0,06.
 \end{aligned}$$

Dadurch sind mehr Felder bekannt:

X/Y	Y = 1	Y = 2	Σ
X = 2	0,2		0,4
X = 3	0,08		0,4
X = 4		0,06	0,2
Σ			1

Schließlich lassen sich noch manche Felder als Differenz bestimmen. So ist in der ersten Zeile die Summe 0,4, d.h.  $W(X=2, Y=2) = 0,2$ , genauso  $W(X=3, Y=2) = 0,4 - W(X=3, Y=1) = 0,4 - 0,08 = 0,32$ .

Damit lässt sich die zweidimensionale Verteilung komplettieren:

X/Y	Y = 1	Y = 2	Σ
X = 2	0,2	0,2	0,4
X = 3	0,08	0,32	0,4
X = 4	0,14	0,06	0,2
Σ	0,42	0,58	1

b) Folgendes gilt:

### LAMBERT-REGEL UNABHÄNGIGKEIT VON ZUFALLSVARIABLEN:

Zwei Zufallsvariablen sind **unabhängig**, wenn in einer zweidimensionalen Verteilungstabelle das **Produkt der Randwahrscheinlichkeiten** gleich dem **jeweiligen Eintrag in der Zelle** ist.

Mathematisch korrekt ausgedrückt: X und Y unabhängig, wenn

$$W(X = x) \cdot W(Y = y) = W(X = x, Y = y)$$

für alle möglichen Werte x und y ist.

Dies ist aber hier nicht der Fall, so z.B.  $W(X = 2) \cdot W(Y = 1) = 0,4 \cdot 0,2 = 0,08$ , aber  $W(X = 2, Y = 1) = 0,2 \neq 0,08$ .

Wegen  $W(X = 2) \cdot W(Y = 1) \neq W(X = 2, Y = 1)$  sind insgesamt die beiden Zufallsvariablen X und Y damit abhängig.