Fachbereich Mathematik und Informatik
der Universität Marburg
Dr. Helga Lohöfer

Sommersemester 2012

Mathematik für Natur- und Sozialwissenschaftler

- Blatt 10 -
Abgabe: Mittwoch, den 27.06.2012, 12:10 Uhr vor der Vorlesung

Skript und Aufgabenblätter: Im Internet unter www.mathematik.uni-marburg.de/~lohoefer/
Lektürehinweis: Skript Kap 3.1
Wichtige Begriffe: Lineare Regression, Korrelationskoeffizient r (von Pearson).

1. Bei der praktischen Berechnung der Kovarianz mittels der Formel $s_{xy} = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n}$ macht der Zähler ziemlich viel Arbeit. Der Rechenaufwand halbiert sich fast, wenn man folgende Gleichung ausnutzt:

$$\sum (x_i - \bar{x}) \cdot (y_i - \bar{y}) = \left(\sum x_i \cdot y_i\right) - n \cdot \bar{x} \cdot \bar{y}$$

Rechnen Sie nach, dass diese Gleichung stimmt. (3)
Tipp: Es ist $\sum x_i = n \cdot \bar{x}$, analog für $\sum y_i$.

2. Es besteht ein Zusammenhang zwischen der Wasseroberflächentemperatur X (in ^{0}C) des Meeres und der Verdunstung Y (Class-A-Kessel, mm pro Tag). Verbundene Messungen in den sieben Monaten April bis Oktober eines Jahres ergaben folgende Daten:

Temperatur X [^{0}C]	9,1	15,9	18,7	22,1	19,2	15,8	11,2
Verdunstung Y [$\text{mm} \cdot \text{d}^{-1}$]	2,0	3,2	3,6	3,8	2,5	1,6	0,8

a) Visualisieren Sie die Punkte (x_i | y_i) in einem x-y-Koordinatensystem (Maßstab X-Achse: 1 Kästchen = 1^0C, Y-Achse: 5 Kästchen = 1 $\text{mm} \cdot \text{d}^{-1}$). (1)
b) Aufgrund welcher Kriterien darf aus der Punktwolke auf lineare Korreliertheit von X und Y geschlossen werden? (1)
c) Berechnen Sie die Ausgleichsgerade $\tilde{Y} = A + B \cdot X$ (Drei-Schritt-Verfahren, A und B mit 2 Nachkommastellen) und zeichnen Sie sie in die Skizze aus a) ein. (4)
d) Berechnen Sie den Korrelationskoeffizienten r (4 Nachkommastellen). Liegt eine starke oder schwache lineare Korrelation vor? (3)

3. Gegeben sei eine verbundene Messreihe zu den Variablen X, Y:

X	x_1	...	x_n
Y	y_1	...	x_n

a) Bildet man die Vektoren $\vec{a} = (x_1 - \bar{x}, x_2 - \bar{x}, \dots, x_n - \bar{x})$ und $\vec{b} = (y_1 - \bar{y}, y_2 - \bar{y}, \dots, y_n - \bar{y})$, so gilt für den Winkel α zwischen diesen beiden Vektoren stets:

$\cos \alpha = r$ (Korrelationskoeffizient von Pearson).

Rechnen Sie nach, dass diese Gleichung stimmt. (1)
b) Die stärkstmögliche lineare Korrelation besteht, wenn alle Punkte (x_i | y_i) genau auf der Ausgleichsgeraden $\tilde{Y} = A + B \cdot X$ liegen, d.h. wenn nicht nur, wie stets, $\bar{y} = A + B \cdot \bar{x}$ gilt, sondern auch $y_i = A + B \cdot x_i$ für alle i = 1, ..., n. Rechnen Sie nach, dass dann sogar gilt:
r = 1, falls die Steigung B positiv ist, und r = –1, falls die Steigung B negativ ist. (3)

Bedingungen für die erfolgreiche Teilnahme:
- Regelmäßige Teilnahme an der Vorlesung (maximal 2 mal Fehlen).
- Auf mindestens 9 Übungsblätter jeweils mindestens 5 Punkte erhalten, insgesamt jedoch mindestens 75 Punkte. Es sind 11 Übungsblätter mit jeweils mindestens 15 Punkten geplant.
- Dazu das Bestehen einer Klausur.

Klausurtermine:
1. Klausur: Do, 19.07.12, 11:00 – 12:45, Hörsaalgebäude Chemie Lahnberge, Hans-Meerwein-Str., HS A + B
2. Klausur: Do, 11.10.12, 11:00 – 12:45, Hörsaalgebäude Chemie Lahnberge, Hans-Meerwein-Str., HS B

Bezeichnungen:

Varianz von X $\quad s_x{}^2 = \frac{\sum (x_i - \overline{x})^2}{n}$

Varianz von Y $\quad s_y{}^2 = \frac{\sum (y_i - \overline{y})^2}{n}$

Kovarianz von X und Y $\quad s_{xy} = \frac{\sum (x_i - \overline{x}) \cdot (y_i - \overline{y})}{n}$

Alternative Berechnungsformeln (siehe Aufgabenblatt 3, Nr. 2, und Aufgabe Nr. 1 oben) :

Varianz von X $\quad s_x{}^2 = \frac{\sum x_i^2 - n \cdot \overline{x}^2}{n}$

Varianz von Y $\quad s_y{}^2 = \frac{\sum y_i^2 - n \cdot \overline{y}^2}{n}$

Kovarianz von X und Y $\quad s_{xy} = \frac{\sum x_i \cdot y_i - n \cdot \overline{x} \cdot \overline{y}}{n}$

Lineare Regression

Gegeben n verbundene Messungen (x_i,y_i) zweier Variabler X und Y. Visualisierung der Daten zeige, dass die Punkte $(x_i|y_i)$ (i = 1,2,..., n), ungefähr auf einer steigenden oder fallenden Geraden liegen oder in einer Punktwolke mit bandartiger Optik, welche ungefähr geradlinig steigt oder fällt. .

Die **Ausgleichsgerade** oder **Regressionsgerade** $\tilde{Y} = A + B \cdot X$ zur Messreihe ist dadurch eindeutig charakterisiert, dass für sie der Abstand zwischen der empirischen Datenzeile $(y_1, \ldots, y_n)$ und der theoretischen Datenzeile $(A+B \cdot x_1, \ldots, A+B \cdot x_n)$ minimal wird.

Berechnung der Konstanten A, B in der Formel $\tilde{Y} = A + B \cdot X$ **(Drei-Schritt-Verfahren):**

1. Schritt: Berechne die Mittelwerte $\overline{x} = \frac{\sum x_i}{n}$ und $\overline{y} = \frac{\sum y_i}{n}$,

die Quadratsumme $\sum x_i^2$ und

die Produktsumme $\sum x_i \cdot y_i$

2. Schritt: Berechne die **Steigung der Geraden** $B = \frac{\sum x_i \cdot y_i - n \cdot \overline{x} \cdot \overline{y}}{\sum x_i^2 - n \cdot \overline{x}^2}$. Es gilt hiermit $B = \frac{s_{xy}}{s_x{}^2}$.

3. Schritt: Nach dem Runden von B berechne den **Schnittpunkt mit der senkrechten Achse** $A = \overline{y} - B \cdot \overline{x}$.

Bemerkung: Die Ausgleichsgerade verläuft durch die zwei Punkte (0 |A) und $(\overline{x}|\overline{y})$.

Soll die Ausgleichsgerade eine Proportionalität modellieren, also durch den Ursprung verlaufen, so **muss** man für den Schnittpunkt mit der senkrechten Achse **a priori** A = 0, d.h. also $\overline{y} - B \cdot \overline{x} = 0$ fordern. Das bedeutet eine **Zusatzbedingung** an die Steigung der Geraden. Daraus ergibt sich eine geänderte

Berechnung der Ursprungsgeraden $\tilde{Y} = C \cdot X$ **(Ein-Schritt-Verfahren):** $\quad C = \frac{\sum x_i \cdot y_i}{\sum x_i^2}$

Korrelationskoeffizient r (aussagekräftig nur bei vorliegender <u>linearer</u> Korrelation von X und Y)

$$r = \frac{\sum (x_i - \overline{x}) \cdot (y_i - \overline{y})}{\sqrt{\sum (x_i - \overline{x})^2} \cdot \sqrt{\sum (y_i - \overline{y})^2}} \quad \text{oder auch} \quad r = \frac{\sum x_i \cdot y_i - n \cdot \overline{x} \cdot \overline{y}}{\sqrt{\sum x_i^2 - n \cdot \overline{x}^2} \cdot \sqrt{\sum y_i^2 - n \cdot \overline{y}^2}}$$

Winkelrechnung im $\mathbb{R}^n$

Sind $\vec{a} = (a_1, \ldots, a_n)$ und $\vec{b} = (b_1, \ldots, b_n)$ zwei beliebige Vektoren der Länge $\|\vec{a}\| = \sqrt{\sum a_i^2}$ und $\|\vec{b}\| = \sqrt{\sum b_i^2}$ mit dem Skalarprodukt $\vec{a} \cdot \vec{b} = \sum a_i \cdot b_i$, so gilt für den Winkel α zwischen den beiden Vektoren allgemein:

$$\cos \measuredangle(\vec{a}, \vec{b}) = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \cdot \|\vec{b}\|}.$$

Dieser verallgemeinerte Kosinus hat immer einen Wert zwischen −1 und 1. Er ist = 1 bzw. = −1, wenn die Vektoren sich nur durch eine positiven bzw. negativen skalaren Faktor unterscheiden.