Fachbereich Mathematik
M. Kohler / A. Fromkorth

TECHNISCHE UNIVERSITÄT DARMSTADT

WS 2007/08
19. November 2007

Lösungsvorschläge zum 4. Übungsblatt zur „Mathematik und Statistik für Biologen“

Lösung zur Aufgabe 11 (3 Punkte)

(a) i. $1 + 2 + 3 + \cdots + 10 = \sum_{i=1}^{10} i$

ii. $1 + x + x^2 + x^3 + \ldots x^n = \sum_{i=0}^{n} x^i$

(b) $\sum_{k=1}^{5} x_k = \sum_{k=1}^{5} (2^k + 5k) = 7 + 14 + 23 + 36 + 57 = 137$

Lösung zur Aufgabe 12 (3 Punkte)

(a) i. $\mathcal{P}(A) = \{\emptyset, \{2\}, \{5\}, \{9\}, \{2,5\}, \{2,9\}, \{5,9\}, \{2,5,9\}\}$

ii. $A \cup B = \{0,2,5,9\}, A \cap B = \{5\}, A \setminus B = \{2,9\}$

iii. $B \cup \mathbb{N} = \mathbb{N}_0$, $\mathbb{N} \setminus \mathbb{Z} = \emptyset$

(b) Nein. Gegenbeispiel sind z.B. die Mengen A und B aus dem ersten Teil der Aufgabe.

Lösung zur Aufgabe 13 (3 Punkte)

(a) Wir wollen die Punkte durch eine Gerade approximieren. Demnach versuchen wir eine allgemeine Geradengleichung der Form

$$g(x) = ax + b$$

an die gegebenen Punkte anzupassen. (Also a und b so zu wählen, dass die berechnete Gerade möglichst dicht bei den Punkten liegt.) Werte für a und b findet man durch Minimieren der Funktion

$$f(a,b) := \sum_{i=1}^{3} (y_i - g(x_i))^2.$$

Analog zum Minimieren von Funktionen einer Veränderlicher kann man hier ansetzen mit:

$$\begin{aligned} \frac{d}{da} f(a,b) &= 0 \\ \frac{d}{db} f(a,b) &= 0 \end{aligned}$$

(D.h. man berechnet die Ableitung nach der entsprechenden Variable). Dies liefert ein Gleichungssystem bestehend aus 2 Gleichungen für 2 Unbekannte.

$$\begin{aligned} f(a,b) &= \sum_{i=1}^{3}(y_i - g(x_i))^2 \\ &= (2-g(1))^2 + (3-g(2))^2 + (5-g(3))^2 \\ &= (2-(a+b))^2 + (3-(2a+b))^2 + (5-(3a+b))^2 \end{aligned}$$

Ableiten nach a ergibt:

$$\begin{aligned} \frac{d}{da} f(a,b) &= 2(2-(a+b))(-1) + 2(3-(2a+b))(-2) + 2(3-3a+b))(-3) \\ &= -46 + 28a + 12b \end{aligned}$$

Ableiten nach b ergibt:

$$\begin{aligned} \frac{d}{db} f(a,b) &= 2(2-(a+b))(-1) + 2(3-(2a+b))(-1) + 2(3-3a+b))(-1) \\ &= -20 + 12a + 6b. \end{aligned}$$

Zu lösen ist also das lineare Gleichungssystem

$$\begin{aligned} 28a + 12b &= 46 \\ 12a + 6b &= 20 \end{aligned}$$

Man erhält: $a = \frac{3}{2}$, $b = \frac{1}{3}$. Die Regressionsgerade ist also

$$g(x) = \frac{3}{2}x + \frac{1}{3}.$$

(b) Man erhält:

$$\begin{aligned} \bar{x} &= \frac{1}{3}(1+2+3) = 2 \\ \bar{y} &= \frac{1}{3}(2+3+5) = \frac{10}{3} \\ s_{x,y} &= \frac{1}{2}\left((1-2)(2-\frac{10}{3}) + (2-2)(3-\frac{10}{3}) + (3-2)(5-\frac{10}{3})\right) \\ &= \frac{1}{2}\left(\frac{4}{3} + \frac{5}{3}\right) = \frac{3}{2} \\ s_x^2 &= \frac{1}{2}\left((1-2)^2 + (2-2)^2 + (3-2)^2\right) = 1 \\ \hat{a} &= \frac{s_{x,y}}{s_x^2} = \frac{3}{2} \end{aligned}$$

Damit erhält man die Regressionsgerade

$$y = \hat{a}(x-\bar{x}) + \bar{y} = \frac{3}{2}(x-2) + \frac{10}{3} = \frac{3}{2}x + \frac{1}{3}.$$

Diese Gerade stimmt mit der Geraden aus Teil a) überein.

Lösung zur Aufgabe 14 (3 Punkte)

Abb 1-3 Aufgrund der Lage der Datenpunkte wird die Steigung der Regressionsgerade negativ sein. Da die Korrelation das gleiche Vorzeichen hat wie die Steigung und immer im Intervall $[-1, 1]$ liegt, folgt $-1 < r_{x,y} < 0$. (Da die Punkte nicht alle auf einer Geraden liegen, ist $r_{x,y} \neq -1$. Da die Regressionsgerade nicht waagerecht verläuft, ist auch $r_{x,y} \neq 0$.)

Abb 4 Da die Steigung der Regressionsgerade positiv ist, ist auch die empirische Korrelation positiv. wie zuvor können 1 und 0 nicht vorkommen. Also gilt $r_{x,y} \in (0, 1)$.