Übungsaufgaben 2

Abgabetermin: 29.10.02

1. Beispiel aus der Gerichtsmedizin (Reißland, p.4): Bei der Ermittlung des Kindesvaters benutzt man Daten des neugeborenen Kindes, aus denen man auf die Dauer der Schwangerschaft schließen kann. Hier ist eine entsprechende Tabelle von Meßwerten

x_i[cm]	48	49	50	51	52
y_i[Tage]	277,1	279,3	281,4	283,2	284,8

Bestimme die Regressionsgerade $y = a + bx$, dabei ist

$$b = \frac{\sum_{i=1}^{n} x_i y_i - n \cdot \overline{x} \cdot \overline{y}}{\sum_{i=1}^{n} x_i^2 - n \cdot \overline{x}^2},$$
$$a = \overline{y} - b\overline{x}$$

mit den Mittelwerten $\overline{x} = \frac{1}{n}\sum x_i$, und $\overline{y} = \frac{1}{n}\sum y_i$.

Hinweis zum Aufschreiben: Man verwende folgende Anordnung:

x_i	48	49	50	51	52	$\sum x_i$	$=$	...	$\overline{x}$	$=$	...
y_i	277,1	279,3	281,4	283,2	284,8	$\sum y_i$	$=$	...	$\overline{y}$	$=$	...
$x_i y_i$	...	...	...	...	...	$\sum x_i y_i$	$=$	...			
x_i^2	...	...	...	...	...	$\sum x_i^2$	$=$	...			

2. Seien folgende Datenpaare (x_i, y_i) gegeben

x_i	1	3	6
y_i	2	5	6

Bestimme die Regressionsgerade $y = a + bx$ (für die Abhängigkeit der y-Werte von den x-Werten).

3. Seien die gleichen Datenpaare wie in Aufgabe 2 gegeben, allerdings untersuche man nun, wie x von y abhängt, man bilde also die Regressionsgerade $x = a' + b'y$ für die Tabelle

y_i	2	5	6
x_i	1	3	6

Man würde erwarten, daß $b' = \frac{1}{b}$ sein sollte — zeige, daß dies **nicht** zutrifft!

Man begründe den Unterschied der beiden Ergebnisse: welche Summenwerte werden in den beiden verschiedenen Aufgabenstellungen minimiert?

4. Betrachte die gleichen Datenpaare (x_i, y_i) wie in Aufgabe 1. Diesmal bestimmen wir aber die Regressionsgerade $x = a' + b'y$, dabei ist

$$\begin{aligned} b' &= \frac{\sum x_i y_i - n \cdot \overline{x} \cdot \overline{y}}{\sum y_i^2 - n \cdot \overline{y}^2}, \\ a' &= \overline{x} - b'\overline{y}. \end{aligned}$$

Zusätzlich soll der Korrelations-Koeffizient

$$r_{xy} = \frac{\sum x_i y_i - n \cdot \overline{x} \cdot \overline{y}}{\sqrt{\sum x_i^2 - n \cdot \overline{x}^2} \cdot \sqrt{\sum y_i^2 - n \cdot \overline{y}^2}}.$$

bestimmt werden.

Diese Aufgaben sollen **ohne** Verwendung spezieller statistischer Programme gelöst werden. Zum Rechnen darf natürlich ein Taschenrechner verwendet werden, aber keine spezielle statistische Funktionstaste. Denn die Aufgaben dienen dazu, ein Gefühl dafür zu entwickeln, welche Rechenschritte notwendig sind, um eine Regressionsgerade (und den Korrelations-Koeffizienten) zu berechnen.

Hier eine Internet-Adresse, die Materialien und interaktive Übungen zur linearen Regression und zur Bestimmung des Korrelations-Koeffizienten bereitstellt:

`http://medweb.uni-muenster.de/institute/imib/`
lehre/skripte/biomathe/bio/bio.html

(JUMBO - **J**ava-**u**nterstützte **M**ünsteraner **B**iometrie-**O**berfläche, vom Institut fr Medizinische Informatik und Biomathematik der Universität Münster).

Von der Homepage der Vorlesung gibt es über “Links” einen direkten Link dorthin, und zwar auch auf spezielle Übungen.