

Hausaufgaben für LL-08, Statistik zum 9.4.2009

Aufgabe 1

Lese Abschnitt 2.6 *Zweidimensionale Häufigkeitsverteilungen*. Den Begriff der Kovarianz werden wir in der kommenden Woche kennen lernen. Es ist eine gute Übung schon vorab zu testen, wie viel man von dem noch unbekanntem Material versteht.

Anmerkung zu Formel (2.57): Setzt man $a_i = x_i - \bar{x}$ und $b_i = y_i - \bar{y}$, so erhält man (nach Kürzen von $1/(n-1)$)

$$\left| \sum_{i=1}^n a_i \cdot b_i \right| \leq \sqrt{\sum_{i=1}^n a_i^2} \sqrt{\sum_{i=1}^n b_i^2}$$

Teilt man durch die rechte Seite ergibt sich die aus der Vektorrechnung „bekannte“ Beziehung

$$\frac{\left| \sum_{i=1}^n a_i \cdot b_i \right|}{\sqrt{\sum_{i=1}^n a_i^2} \sqrt{\sum_{i=1}^n b_i^2}} = |\cos \alpha| \leq 1$$

Die Kovarianz ist formelmäßig mit dem Skalarprodukt vergleichbar, während die Standardabweichungen zur Länge von Vektoren korrespondieren.

Bearbeite Aufgabe 2.16.

Aufgabe 2

Lese Abschnitt 2.8 *Regressionsrechnung*.

Beachte, dass hier die Steigung mit b und der Achsenabschnitt mit a bezeichnet wird. Gerade bei b besteht Verwirrungspotenzial mit der Notation in der Vorlesung.

Ich bezeichne die Variablen x und y (s.S. 59) oft als unabhängige bzw. abhängige Variable (s. auch Duller, Kapitel 9, Regression).

Auf Seite 61 findet Ihr die mathematische Herleitung, die man noch nicht verstehen muss, aber durchaus verstehen kann. In einigen Wochen werden wir in Mathe über partielle Ableitungen und Extrema von Funktionen in mehreren Variablen sprechen.

Zu Seite 63: Formel (2.77) ist die Punkt-Steigungs-Form der Geraden durch den Punkt der Mittelwerte. Eigenschaft d) beschreibt den Ausgleichs-Charakter - die Abweichungen nach oben (positive Residuen) sind gleich den Abweichungen nach unten (negative Residuen).

Bearbeite Aufgabe 2.19 mit dem angegebenen Modellansatz

$$\Delta l = l \alpha \Delta T$$

mit der Steigung $l\alpha$ und ohne Achsenabschnitt. Die angegebene Aufgabenstellung, beide Parameter zu berechnen ist physikalisch unsinnig (vgl. auch erster Messwert).

Aufgabe 2.20 wurde zu großen Teilen in der Vorlesung behandelt. Wiederholt die Ausführungen dazu.

Aufgabe 3

Geben Sie Beispieldaten für eine lineare Regression in EXCEL ein. Probieren Sie die Analyse-Funktion „Regression“ aus.

Probieren Sie die folgende Alternative aus: Rufen Sie mit den Daten den Diagrammtyp XY auf. Right-Klick auf die Datenreihe liefert die Option Trendlinie hinzufügen. (Vgl. Buch von Duller, Abschnitt 9.2).

Aufgabe 4

Die folgenden vier Beispiele wurden von Frank Anscombe im Jahre 1973 veröffentlicht. Zeichnen Sie die Streudiagramme, berechnen Sie die Mittelwerte, Standardabweichungen, Korrelationskoeffizienten und Regressionsgeraden. Zeichnen Sie die Regressionsgeraden in das jeweilige Streudiagramm. Kommentieren Sie die Ergebnisse.

Beispiel 1:

x	10	8	13	9	11	14	6	4	12	7	5
y	8,04	6,95	7,58	8,81	8,33	9,96	7,24	4,26	10,84	4,82	5,68

Beispiel 2:

x	10	8	13	9	11	14	6	4	12	7	5
y	9,14	8,14	8,74	8,77	9,26	8,1	6,13	3,10	9,13	7,26	4,74

Beispiel 3:

x	10	8	13	9	11	14	6	4	12	7	5
y	7,46	6,77	12,74	7,11	7,81	8,84	6,08	5,39	8,15	6,42	5,73

Beispiel 4: (Der x -Wert 19 ist kein Tippfehler.)

x	8	8	8	8	8	8	8	19	8	8	8
y	6,58	5,76	7,71	8,84	8,47	7,04	5,25	12,50	5,56	7,91	6,89

Aufgabe 5

Überfliege Kapitel 9 im Buch von Duller. Überspringe den letzten Abschnitt zu SPSS.

Bearbeite Aufgabe 9.1. Man muss dazu die Bedeutung des Bestimmtheitsmaßes nicht verstanden haben; es genügt, in die Formel einzusetzen.