

Einführung in die Informatik	3. Übung
Thema: haben Nasenlänge und Schuhgröße etwas miteinander zu tun?	WS 02/03
	Ausgabetermin: 25.Nov.02 letzter Fertigstellungstermin: 16.12.02

Programmierbasis: Visual Basic

Problemstellung: In einer statistischen Untersuchung soll der Frage nachgegangen werden, welcher Zusammenhang zwischen Nasenlänge und Schuhgröße in einer Gruppe von Personen besteht. Bei N befragten Personen sind dabei folgende statistischen Größen von Interesse, wobei die Bezeichnungen gelten

- n_i = Nasenlänge der i-ten Person in cm

- s_i = Schuhgröße der i-ten Person in cm

mittlere Nasenlänge:
$$\bar{n} = \frac{1}{N} \sum_{i=1}^N n_i \quad (1)$$

mittlere Schuhgröße:
$$\bar{s} = \frac{1}{N} \sum_{i=1}^N s_i \quad (2)$$

Varianz der Nasenlänge:
$$v_n = \frac{1}{N} \sum_{i=1}^N (n_i - \bar{n})^2 \quad (3)$$

Varianz der Schuhgröße:
$$v_s = \frac{1}{N} \sum_{i=1}^N (s_i - \bar{s})^2 \quad (4)$$

Korrelationskoeffizient:
$$k = \frac{\frac{1}{N} \sum_{i=1}^N (s_i - \bar{s}) \cdot (n_i - \bar{n})}{\sqrt{v_n \cdot v_s}} \quad (5)$$

Zur Vorbereitung der Programmerstellung müssen zuerst einmal alle Ausdrücke von Hand ausmultipliziert und die resultierenden Einzelsummen gebildet werden. Im Ergebnis dieser Rechnung kommen danach nur die folgenden 5 Summen vor:

$$\sum n_i \quad \sum s_i \quad \sum n_i^2 \quad \sum s_i^2 \quad \text{und} \quad \sum s_i \cdot n_i \quad (6)$$

Da es mühsam und umständlich wäre, alle Eingabewerte n_i , s_i von Hand einzulesen, befinden sie sich auf einer ASCII-Datei mit dem Namen `nase.dat`, die Ihnen im Verzeichnis "`handout\luedecke\...`" zur Verfügung steht. Sie können diese Datei anschauen, kopieren (und auch ändern, was Sie aber nicht tun sollten, um Ihre Rechenergebnisse später vergleichen zu können), indem Sie im Windows-Explorer auf den Dateinamen doppelklicken und dann zum Anzeigen entweder den EDT oder den Word-Pad wählen. Der erste Zahlenwert in jedem Satz der Datei ist die Nasenlänge, der zweite die Schuhgröße.

Der "Dreh" des zu erstellenden Programms besteht darin, in einer einzigen Programmschleife alle Wertepaare n_i , s_i einzulesen und beim Einlesen sofort die vorgenannten 5 Summen und die Anzahl der Wertepaare N zu berechnen. N wird durch Hochzählen einer Zählvariablen in der Programmschleife gewonnen. Die Programmschleife bricht ab, wenn der Lesezeiger am Ende der ASCII-Datei angekommen ist (s. VB-Beispiel weiter unten). Mit den Werten für N und den 5 Summen (6) werden dann die gewünschten Größen (1) bis (5) berechnet. Die Einzelwerte n_i , s_i brauchen bei diesem Vorgehen also nicht in Feldern zwischengespeichert zu werden, und dementsprechend dürfen in Ihrem Programm auch keine Felder verwendet werden.

Das folgende Beispiel zeigt das Öffnen und Schließen einer ASCII-Datei und das Einlesen einer ASCII-Zeile

```
Dim Dateinr as integer
Dateinr = FreeFile
Open "a:\nase.dat" For Input As Dateinr           ' ASCII-Datei zum Lesen
                                                    ' öffnen (input=Lesen)
Input #Dateinr, x, y                             ' die beiden Single-Variablen x und y werden
                                                    ' gelesen
Close Dateinr                                    ' ASCII-Datei schließen (nicht vergessen!!!)
```

Um die Vergabe einer Dateinummer braucht man sich nicht zu kümmern, in `FreeFile` steht stets eine freie Nummer. Der Dateiname ist mit vollständiger Pfadbezeichnung anzugeben. Um Zugriffsprobleme zu vermeiden, empfehle ich, die Datei `nase.dat` vom handout-Verzeichnis auf Ihre Diskette zu kopieren, so daß die obige Pfadangabe zutrifft. Ihr Programm liest die Dateidaten dann von der Diskette.

Nach jedem Lesevorgang wandert der Dateizeiger automatisch um einen Satz weiter. Um daher beliebig viele Sätze der ASCII-Datei lesen zu können, ist die o.g. einzige Leseanweisung durch eine Schleife mit Leseanweisung zu ersetzen, die so oft durchlaufen wird, wie sich Sätze auf der ASCII-Datei befinden.

Wenn Sie in der Datei `nase.dat` nachsehen, werden Sie allerdings nirgendwo einen Hinweis darauf finden, wieviele Sätze sie hat. Woher weiß dann die Programmschleife, wann mit dem Lesen Schluß ist? Würde die Programmschleife immer weiterlaufen, wäre eine Fehlermeldung und der Programmabbruch die Folge, weil unzulässigerweise versucht wird, über den letzten Satz der Datei hinaus zu lesen. Das Prinzip, wie ohne Kenntnis der Satzanzahl ein definiertes Schleifenende ohne Fehlerabbruch zu erreichen ist, stützt sich auf die `EOF()`-Funktion (`EOF=End Of File`). `EOF()` gibt einen logischen Wert zurück, der "true" ist, wenn das Dateiende erreicht wurde. Im folgenden Beispiel ist gezeigt, wie es geht:

```
Do While Not EOF(Dateinr)           ' while_1
Input #Dateinr, xnl, xsh
...
... ' Verarbeitung
...
Loop                                 ' while_1
```

Die Schleife bleibt also nur solange aktiv, bis das Dateiende erreicht ist.

Die Oberfläche Ihres VisualBasic-Programms sollte etwa wie folgt aussehen:

mittlere Nasenlänge	4,26
mittlere Schuhgröße	26,29
Varianz der Nasenlänge	0,28
Varianz der Schuhgröße	2,11
Korrelationskoeffizient	0,12

Berechnung starten

Bei der Abgabe der Übung müssen Sie über die hier verwendeten statistischen Größen Bescheid wissen, etwa, welche Bedeutung ein Korrelationskoeffizient hat und welche Werte er annehmen kann. Zur abzugebenden Übungsleistung gehört ein Struktogramm sowie Ihre Handrechnung zur Ermittlung der Summen (6).

Bei ca. 110 abgegebenen Übungen ist für mich das Einhalten einer gewissen Ordnung von Ihrer Seite hilfreich. Daher wiederholt:

- Ihre Übungsleistung sollte auf möglichst wenigen Blättern enthalten sein, in der Regel sollten 3-6 DIN A4-Blätter für Text und Computerausdrucke wirklich ausreichen.
- Die Blätter sind mit einer Klammer zusammenzutackern. Keine Sichthüllen, Mappen o. dergl. verwenden!!!
- Links oben auf jedem Blatt bitte Ihren Namen u. Ihre Matr.-Nr. eintragen (in Word-Text, nicht handschriftlich, denn manche Handschrift ist nicht zu entziffern)
- Das erste Blatt unterschreiben.

So schwer ist das doch nicht, oder?