

Versuchsprotokoll

Versuch 30 (Statistik)

1. Aufgabenstellung

Es werden einige einfache BASIC-Programme erstellt um eine Veranschaulichung für statistische Fehler, den Einfluß vieler Messungen und die mathematische Behandlung von Meßwerten zu liefern.

2. BASIC-Programmierung

Die Programmiersprache BASIC hat eine gut verständliche Syntax; mit einem sehr geringen Pool von Rechenoperationen (+, -, *, /, ^, sqr, exp) und Befehlsstrukturen (FOR-NEXT-Schleife, IF-THEN-Entscheidung) können mathematische Operationen mit der großen Rechengeschwindigkeit eines Computers durchgeführt werden. Die Hauptaufgabe im naturwissenschaftlichen Bereich besteht darin, die Vorstellungen von einem Sachverhalt formelmäßig zu erfassen und die vom Computer ausgegebenen Werte zu interpretieren.

3. Lineare Regressionsrechnung

Um einen vermuteten Zusammenhang zwischen zwei Größen zu bestätigen werden die Ergebnisse einer Meßreihe oft so aufgetragen, daß sich eine Gerade ergibt (entsprechende Skalierung der Achsen, z.B. $1/x$, $\ln(x)$...). Die optimale Gerade zwischen den Meßpunkten kann dann durch eine lineare Regressionsrechnung ermittelt werden, bei welcher die Summe der Fehlerquadrate der einzelnen Meßpunkte minimiert wird. Es besteht auch die Möglichkeit, Funktionen höherer Ordnung an Meßgrößen anzupassen, dabei sind aber die Abweichungen weniger leicht zu erkennen.

Durch den Korrelationskoeffizienten (<1) wird angegeben, wie gut die Meßwerte mit der berechneten Geraden übereinstimmen.

Das Programm zur linearen Regressionsrechnung (s.u.) wurde mit einem vorgegebenen Satz von Meßwerten getestet:

X	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0
Y	1.6	1.9	2.6	3.1	3.4	4.0	4.4	5.1	5.6	6.1

Hierbei ergab sich eine Gerade

$$Y = 1 + 0,5054545 X$$

mit einem Korrelationskoeffizienten von 0,9981419, dieses Ergebnis stimmt mit der grafischen Darstellung der Meßpunkte überein.

4. Zufallszahlenziehung

Mit dem in BASIC enthaltenen Zufallszahlengenerator wurden Zahlen zwischen 0 und 1 gezogen und überprüft, ob sich eine gleichmäßige Verteilung der gezogenen Zahlen auf Wertebereiche in Zehntel-Schritten ergibt. Dies ist der Fall, wobei allerdings die

zufälligen Abweichungen bei einer geringen Anzahl von gezogenen Zahlen noch stark ins Gewicht fällt.

5. Bernoulli – und Gaußverteilung

Führt man N Versuche durch, bei welchen die Wahrscheinlichkeit für einen Erfolg p und die für einen Mißerfolg q ist, so ist die Wahrscheinlichkeit n Erfolge zu erhalten durch die Bernoulli-Verteilung gegeben. Für sehr große N geht diese in die Gaußverteilung über:

$$W_{Ber} = \binom{N}{n} p^n q^{N-n} \qquad W_{Ga} = \frac{1}{\sqrt{2\pi Npq}} e^{-\frac{(n-Np)^2}{2Npq}}$$

Die Bernoulli-Verteilung streut um den Wert $\sigma = (Npq)^{-1/2}$, die Gaußverteilung um $\sigma = (pqN^2/N-1)^{-1/2}$

Auch bei der Messung naturwissenschaftlicher Größen streuen die Ergebnisse gemäß einer Bernoulli-Verteilung, da man meist mit einer sehr großen Teilchenzahl (mole) arbeitet, kann diese gut mit einer Gaußverteilung angenähert werden.

Es wurde ein Programm erstellt, welches eine große Anzahl von Zufallsexperimenten simuliert (Würfel) und die Ergebnisse mit der Bernoulli- und Gaußverteilung vergleicht. Als Erfolg zählen hierbei nur 6er, die Anzahl der 6er für eine bestimmte Würfelanzahl wurde registriert.

Bei fünfzig Würfeln liegt der Erwartungswert bei 25/3 Sechsen, also zwischen 8 und 9, die Streuung für die Bernoulli- Verteilung bei $\pm 2,635$ und für die Gauß-Verteilung bei $\pm 2,662$.

Die Ergebnisse der einzelnen Experimente erfüllten die Bernoulli-Verteilung, auch die mathematisch sehr viel einfacher zu behandelnde Gauß-Verteilung lieferte schon recht gute Werte, bei einer Million Würfe lagen z.B. über 75% der Werte innerhalb des Bereichs der Standardabweichung.

Bei 10 Würfeln (Erwartungswert: 5/3) liegen die Streuungen bei $\pm 1,179$ bzw. $\pm 1,242$, Hier zeigt sich die Schwäche der Beschreibung durch eine Gauß-Verteilung, deren Maximalwert zu hoch liegt und die dann zu steil abfällt.

Programm zur linearen Regression:

```
CLS
DIM x(1000)
DIM y(1000)
40 INPUT „Anzahl der Messwerte (3-1000) =“; z
IF z=2 THEN PRINT „Das ist dann wohl eine Gerade...”
IF z=2 THEN GOTO 40
IF z<2 THEN PRINT „Spassvogel“
IF z<2 THEN GOTO 40
IF z>1000 THEN PRINT „Das ist etwas zu viel“
IF z>1000 THEN GOTO 40
C1=0
D1=0
E1=0
F1=0
G1=0
FOR i=1 TO z
INPUT „ersten x-Wert eingeben“; x(i)
C1=C1+x(i)
F1=F1+(x(i)*x(i))
INPUT „dazugehoerigen y-Wert eingeben“ ;y(i)
D1=D1+y(i)
E1=E1+(x(i)*y(i))
G1=G1+(y(i)*y(i))
NEXT i
FOR i=2 TO z
INPUT „noch einen x-Wert eingeben“; x(i)
C1=C1+x(i)
F1=F1+(x(i)*x(i))
INPUT „dazugehoerigen y-Wert eingeben“ y(i)
D1=D1+y(i)
E1=E1+(x(i)*y(i))
G1=G1+(y(i)*y(i))
NEXT i
CLS
PRINT „Es ist vollbracht“
QX=(F1-C1*C1/z)
QY=(G1-D1*D1/z)
QXY=(E1-C1*D1/z)
B=QXY/QX
A=(D1-B*C1)/z
PRINT „Die Geradengleichung lautet:“
PRINT „
PRINT „
PRINT „          „;a,“ + „;b,“x“
PRINT „
PRINT „
PRINT „
PRINT „Der Fehler des y-Wertes betraegt:“
SXY=SQR((QY-QXY*QXY/QX)/(z-2))
PRINT „          „;SXY
PRINT „Der Fehler der Steigung betraegt:“
SB=SXY/SQR(QX)
PRINT „          „; SB
PRINT „Der Fehler des Achsenabschnittes betraegt:“
SA=SB*SQR(F1/z)
PRINT „          „; SA
PRINT „Der Korrelationskoeffizient betraegt:“
R=QXY/SQR(QX*QY)
PRINT „          „; R
CALL lplot (x(),y(),z,a,b)
END
```

Programm zur Ziehung von Zufallszahlen:

```
CLS
RANDOMIZE TIMER
DIM H(1000)
FOR i=1 TO z
H(i)=0
NEXT i
INPUT „Anzahl der Zufallszahlen“; n
FOR i=1 TO n
X=RND
FOR j=1 TO 10
IF X>(j-1)/10 AND x<=j/10 THEN H(j)=H(j)+1
NEXT j
NEXT i
INPUT „Bitte Eingabetaste drücken“; i
FOR i=1 TO z
PRINT i, H(i)
NEXT i
INPUT i
CALL RNDPLOT (H(),z)
END
```

Programm zur Durchführung eines Würfelexperiments und Vergleich mit der Bernoulli- und Gaußverteilung:

```
RANDOMIZE TIMER
CLS
DIM H(1000)
DIM BN(1000)
DIM GA(1000)
INPUT „Anzahl der Versuche“; k
INPUT „Anzahl der Würfel“; n
  FOR u=0 TO n
    GA(u)=(1/SQR(5/18*3.14159*n))*EXP(-(u-n*(1/6))^2/(n*5/18))
  NEXT u
  BN(0)=(5/6)^n
  FOR u=0 TO n
    BN(u+1)=((1/6)/(5/6))*(n-u)/(u+1)*BN(u)
  NEXT u
H(i)=0
  FOR b=1 TO k
    G=0
    FOR a=1 TO n
      X=RND
      IF x<=1/6 THEN g=g+1
    NEXT g
    H(g) =H(g) +1
  NEXT b
H(n) = 1E-20
FOR i=0 TO n
  IF H(i) > 0 THEN z=i
  IF GA(i)>0 THEN
    PRINT i, H(i), BN(i)*k, GA(i)*k
  END IF
NEXT i
CALL RNDPLOT(H())
CALL RNDPLOT(BN())
CALL RNDPLOT(GA())
END
```

WERTE der einzelnen Versuche:

Zunächst wurden Versuche mit 50 Würfeln durchgeführt, es wurden Reihen von 10, 1000, 10000 und 1000000 Würfeln berechnet, die Erfolge gezählt und mit der prozentualen Wahrscheinlichkeit für die Bernoulli- und Gaußverteilung verglichen.

	10er	1000er	10000er	10 ⁶ er	Bernoulli-%	Gauß-%
0	0	0	1	116	0,01099	0,10200
1	0	2	6	1077	0,10988	0,31514
2	0	5	45	5356	0,53844	0,84304
3	0	10	157	17359	1,72299	1,95279
4	1	37	401	40153	4,04904	3,91675
5	1	79	744	74855	7,45023	6,80230
6	2	109	1194	111756	11,17534	10,22935
7	1	134	1384	140344	14,04900	13,31993
8	1	165	1433	151259	15,10267	15,01817
9	2	144	1422	140855	14,09583	14,66203
10	2	112	1197	115389	11,55858	12,39460
11	0	84	822	83669	8,40624	9,07262
12	0	52	566	55191	5,46406	5,75035
13	0	35	310	32044	3,19437	3,15586
14	0	19	186	16793	1,68845	1,49969
15	0	8	74	8060	0,81046	0,61709
16	0	4	36	3565	0,35458	0,21987
17	0	1	15	1359	0,14183	0,06783
18	0	0	4	524	0,05200	0,01812
19	0	0	2	198	0,01752	0,00419
20	0	0	1	56	0,00543	0,00084
21	0	0	0	18	0,00155	0,00015
22	0	0	0	3	0,00041	0,00002
23	0	0	0	1	0,00010	0,00000
24	0	0	0	0	0,00002	0,00000
25	0	0	0	0	0,00000	0,00000
<i>gesamt</i>	<i>10</i>	<i>1000</i>	<i>10000</i>	<i>10⁶</i>	<i>100,00001</i>	<i>99,96273</i>

Anschließend wurde mit zehn Würfeln eine Reihe mit 1000000 Würfeln durchgeführt und wieder mit den rechnerischen Werten verglichen:

	10 ⁶ er	Bernoulli-%	Gauß-%
0	161393	16,15056	12,45323
1	322927	32,30112	28,84625
2	290770	29,07100	32,52406
3	155631	15,50453	17,84958
4	53988	5,42659	4,76825
5	12867	1,30238	0,62001
6	2144	0,21706	0,03924
7	258	0,02481	0,00121
8	21	0,00186	0,00002
9	1	0,00008	0,00000
10	0	0,00000	0,00000
<i>gesamt</i>	<i>10⁶</i>	<i>99,99999</i>	<i>97,10185</i>