Markus Paul

Der Zentrale Grenzwertsatz

Viele Vorgänge in der Natur und Technik lassen sich durch normalverteilte Zufallsvariablen beschreiben, z.B. Körpergröße von Personen, Länge von Stäben beim Zuschneiden, Geburtsgewicht von Neugeborenen, Füllgewicht von Packungen....
Die zentrale Bedeutung der Normalverteilung liegt im **Zentralen Grenzwertsatz**, der besagt, dass eine **SUMME von vielen unabhängigen, beliebig verteilten Zufallsvariablen gleicher Größenordnung angenähert normalverteilt ist**, und zwar umso besser angenähert, je größer ihre Anzahl ist.
Der Zentrale Grenzwertsatz bildet die Grundlage dafür, dass Stichprobenverteilungen ab einem bestimmten Stichprobenumfang durch die Normalverteilung approximiert werden können. Die Normalverteilung ist ein mathematisches Modell, das als ein *Grundpfeiler der mathematischen Statistik* angesehen werden kann. Da sich viele zufällige Variablen, die in der Natur beobachtet werden können, als Überlagerung vieler einzelner, weitgehend unabhängiger Einflüsse auffassen lassen, können diese Variablen durch die Normalverteilung beschrieben werden.

Mathematisch formuliert (Duden: Rechnen und Mathematik):
Es seien X_1, X_2, ..., X_n unabhängige Zufallsvariablen und $X := X_1 + X_2 + ... + X_n$.
Dann gilt die Näherungsformel

$$P(X \le x) \approx \Phi\left(\frac{x-\mu}{\sigma}\right)$$

mit $\mu = \sum_{i=1}^{n} E(X_i)$ und $\sigma^2 = \sum_{i=1}^{n} V(X_i)$.

Dabei ist Φ die Wahrscheinlichkeitsfunktion der standardisierten Normalverteilung. In der Regel ist die Näherung umso besser, je größer n ist.

Wie kann dieser sowohl in theoretischer als auch in praktischer Hinsicht bedeutendste Satz der Statistik Schülern sinnlich vermittelt werden?

Wir beschränken uns auf den Spezialfall, dass die Zufallsvariablen identisch verteilt sind, wählen als Beispiel das Würfeln und untersuchen die Zufallsvariable X = geworfene Augenzahl.
Diese Zufallsvariable ist gleichverteilt:

k	1	2	3	4	5	6
P(X=k)	$^1/_6$	$^1/_6$	$^1/_6$	$^1/_6$	$^1/_6$	$^1/_6$

Als Erwartungswert erhalten wir

$$\mu = E(X) = \sum_{i=1}^{6} k \cdot P(X=k) = 3{,}5;$$

als Varianz erhalten wir

$$\sigma^2 = V(X) = \sum_{i=1}^{6} (k-\mu)^2 \cdot P(X=k) = \frac{35}{12};$$

und damit die Standardabweichung

$$\sigma = \sqrt{\frac{35}{12}} = 1{,}7078.$$

Mit dem TI-83 Plus können wir dies elegant nachrechnen. Wir geben im Listen-Editor ($\boxed{\textbf{STAT}}$ > **1:Edit**) die Werte der Zufallsvariablen in Liste **L1** ein, die Wahrscheinlichkeiten in Liste **L2**.

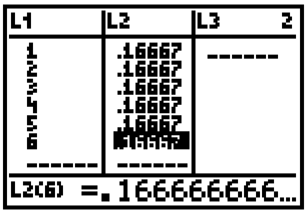

Abb. 1

Mit $\boxed{\textbf{STAT}}$ > **CALC** > 1:**1-Var Stats L1,L2** erhalten wir mit $\bar{x}$ den Erwartungswert und mit σx die Standardabweichung.

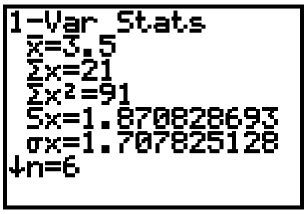

Abb. 2

Nun wollen wir mithilfe des Zufallsgenerators das Würfeln simulieren.
Mit **randInt(1,6)** (über $\boxed{\textbf{MATH}}$ **> PRB > 5:randInt**) wird eine ganzzahlige Zufallszahl zwischen 1 und 6 erzeugt. Durch Betätigen der $\boxed{\textbf{ENTER}}$ – Taste erhalten wir weitere Wurfergebnisse.

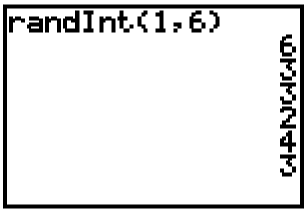

Abb. 3

Mit **randInt(1,6,5)** erhalten wir jeweils eine Liste von 5 Wurfergebnissen, aber bestimmt erhalten Sie andere Zufallszahlen als in der Abbildung.

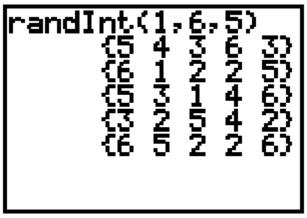

Abb. 4

Wir können den Zufallsgenerator durch einen Anfangs-Code (Seed) intialisieren. Dazu speichern wir etwa die Zahl 1143 in der Variablen **rand** (**MATH > PRB > 1:rand**): **1143 STO> rand**

Mit **randInt(1,6,120) STO> L1** erhalten wir auf jedem TI-83 Plus dieselbe Liste von 120 Wurfzahlen, die in **L1** gespeichert werden.

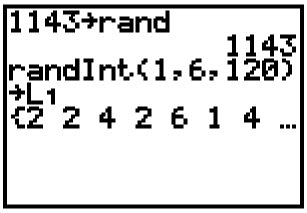

Abb. 5

Die Verteilung der Wurfzahlen können wir durch einen Statistik-Plot grafisch darstellen:

2nd [STAT PLOT] > 1:Plot1, wählen Sie bei Type das Histogramm, Xlist: L1, Freq: 1.

Fenstereinstellung:

als Grenzen Xmin = -0.5 und Xmax = 7.5,

als Klassenbreite Xscl = 1;

als Grenzen für die Häufigkeit Ymin = -10, Ymax = 30.

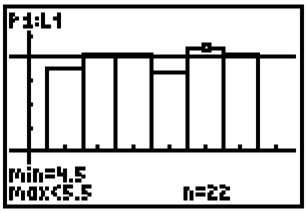

Abb. 6

Wir geben im Y= - Editor zusätzlich **Y1** = 20 ein. Mit **GRAPH** erhalten wir ein Histogramm der Häufigkeits-verteilung (mit **TRACE** können wir z.B. abfragen, dass 22mal die Augenzahl 4 geworfen wurde); mit STAT > CALC > 1:1-Var Stats erhalten wir die statistische Auswertung.

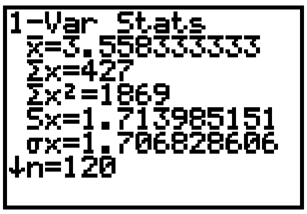

Abb. 7

Augensumme von zwei Würfeln

Nun untersuchen wir die Verteilung der Summe der Augenzahlen zweier Würfel, d.h. wir bilden die Summenvariable $X = X_1 + X_2$, wobei X_i die Augenzahl des i-ten Würfels ist.

Dies lässt sich mit dem TI-83 Plus durch den Summenbefehl realisieren: **sum(randInt(1,6,2))** liefert die Augensumme zweier Würfel.

Wir erzeugen eine Folge von 120 Augensummen-Zahlen, stellen diese grafisch dar und berechnen statistische Kennzahlen:

seq(sum(randInt(1,6,2)), X, 1, 120) STO> L1

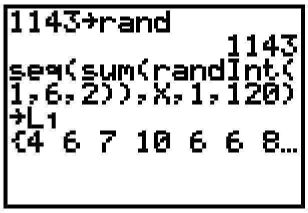

Abb. 8

Da die Augensumme die Werte 2, 3, ..., 12 annehmen kann, wählen wir als Grenzen Xmin = 0.5 und Xmax = 13.5, als Klassenbreite Xscl = 1. Grenzen der Häufigkeiten: Ymin = -10, Ymax = 30.

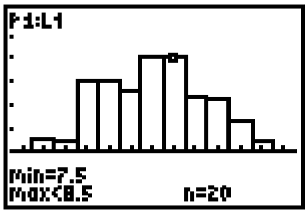

Abb. 9

Deutlich ist zu erkennen, dass die Augensumme nicht mehr gleichverteilt ist, die Augenzahlen in der Mitte treten häufiger auf als die Augenzahlen an den Rändern.

Als Mittelwert dieser Stichprobe erhalten wir die Augensumme 7,033 und als Standardabweichung 2,316.

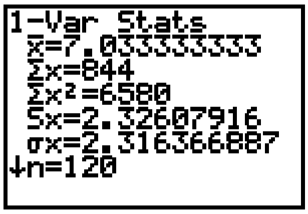

Abb. 10

Was wird wohl der Erwartungswert und was wird die Standardabweichung der theoretischen Verteilung sein?

Die Schüler können Vermutungen aufstellen. Dazu können weitere Listen mit 120 Zufallszahlen erzeugt werden.

Klar erkennbar ist, dass der Mittelwert um 7 schwankt, das ist der doppelte Erwartungswert der Augenzahl eines Würfels.

Die Standardabweichung schwankt um den Wert 2,4. Das ist nicht die doppelte Standardabweichung der Augenzahl eines Würfels (σ = 1,708)! Der Quotient 2,4/1,7 ergibt 1,41. Ist hier vielleicht die Zahl $\sqrt{2}$ im Spiel?

Untersuchen wir statt der Standardabweichung die Varianz, so erhärtet sich dieser Verdacht: Für die Varianz der Summenvariablen erhalten wir ca $2{,}4^2 = 5{,}76$; die doppelte Varianz für einen Würfel ergibt 5,8333.

Folgende Vermutungen können aufgestellt werden:
Der Erwartungswert der Summe zweier Zufallsvariablen ist die Summe der Erwartungswerte:

$E(X_1+X_2) = E(X_1) + E(X_2)$.

Die Varianz der Summe zweier unabhängiger Zufallsvariablen ist die Summe der Varianzen:

$V(X_1+X_2) = V(X_1) + V(X_2)$.

Sind die Zufallsvariablen X_i identisch verteilt mit $E(X_i) = \mu$ und $V(X_i) = \sigma^2$, dann gilt:

$E(X_1+X_2) = 2\mu$ und

$V(X_1+X_2) = 2\sigma^2$ bzw $\sigma_{X1+X2} = \sqrt{2}\,\sigma$.

Für die Augensumme zweier Würfel lässt sich noch relativ leicht die Wahrscheinlichkeitsfunktion aufstellen:

Wir erhalten eine Dreiecksverteilung:

k	2	3	4	5	6	
P(X=k)	$^1/_{36}$	$^2/_{36}$	$^3/_{36}$	$^4/_{36}$	$^5/_{36}$	
k	7	8	9	10	11	12
P(X=k)	$^6/_{36}$	$^5/_{36}$	$^4/_{36}$	$^3/_{36}$	$^2/_{36}$	$^1/_{36}$

Erwartungswert: $\mu = E(X) = \sum_{i=2}^{12} k\cdot P(X=k) = 7{,}0$;

Varianz: $\sigma^2 = V(X) = \sum_{i=2}^{12} (k-\mu)^2 \cdot P(X=k) = \frac{35}{6}$;

Standardabweichung: $\sigma = \sqrt{\frac{35}{6}} = 2{,}415$.

Mit dem TI-83 Plus können wir dies wieder elegant nachrechnen. Wir geben im Listen-Editor ($\boxed{\text{STAT}}$ > **1:Edit**) die Werte der Zufallsvariablen in Liste **L1** ein, die Wahrscheinlichkeiten in Liste **L2**.

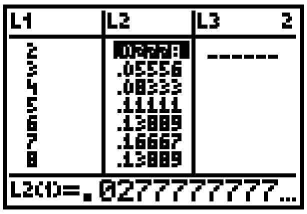

Abb. 11

Die Verteilung können wir mit einem Statistik-Plot grafisch darstellen. (Geben Sie dazu Xlist:L1 und Freq: L2 ein).
Für Xmin=0.5, Xmax=13.5, Xscl=1, Ymin=-0.05, Ymax=0.2,Yscl=0.1 erhalten Sie die Dreiecksverteilung wie in der Abbildung.

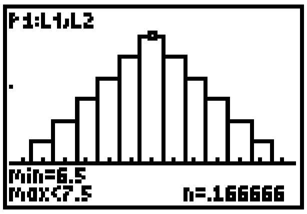

Abb. 12

Mit $\boxed{\text{STAT}}$ > **CALC** > **1:1-Var Stats L1,L2** erhalten wir den Erwartungswert 7 und die Standardabweichung 2,415.
Für σx^2 erhalten wir die Varianz 5,833333, das ist 35/6.

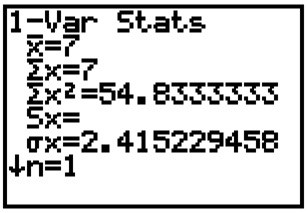

Abb. 13

Augensumme von fünf Würfeln
Nun machen wir einen Sprung und untersuchen die Verteilung der Summe der Augenzahlen von fünf Würfeln, wir bilden die Variable $X = X_1+X_2+X_3+X_4+X_5$, wobei X_i die Augenzahl des i-ten Würfels ist.
Was können wir vermuten?
Keine Hexerei: $E(X) = 5\cdot 3{,}5 = 17{,}5$;

$V(X) = 5\cdot\frac{35}{12} = 14{,}58333$; $\sigma = 3{,}8188$.

Wir erzeugen mit dem TI-83 Plus eine Folge von 120 Augensummen-Zahlen von fünf Würfeln:
(Initialisierung wieder mit 1143)
seq(sum(randInt(1,6,5)), X, 1, 120) $\boxed{\text{STO>}}$ **L1**

Um die Annäherung an die Normalverteilung zu zeigen, geben wir im $\boxed{\text{Y=}}$ - Editor die Dichte der Normalverteilung ein:
Y1 = normalpdf(X, 5*3.5, √(5*35/12))*120
(normalpdf über $\boxed{\text{2nd}}$ **[DISTR] > 1:normalpdf**)
Grenzen für das Häufigkeitshistogramm:
Xmin = 3.5; Xmax = 31.5, Klassenbreite Xscl = 1;
Grenzen der Häufigkeiten: Ymin = -5, Ymax = 20.

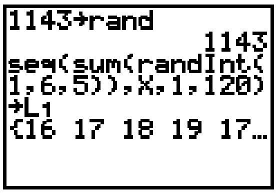

Abb. 14

Abb. 15

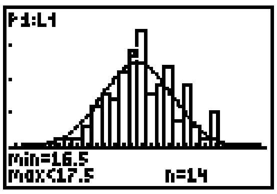

Abb. 16

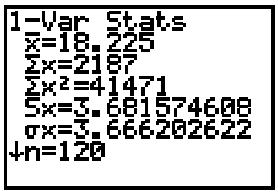

Abb. 17

Erzeugen Sie mehrere Listen von Augensummen-Zufallszahlen, stellen Sie diese grafisch dar und berechnen Sie die statistischen Kennzahlen. Vergleichen Sie diese mit den Werten der theoretischen Zufallsvariable.

Wie kann die Approximation an die Normalverteilung verbessert werden?

Wir erhöhen die Versuchsanzahl auf 200 (Vorsicht: wir stoßen an die Grenzen der Rechnerkapazität, unter Umständen müssen Sie RAM-Speicher freimachen, indem Sie Programme, Listen und Variablen in den Archivspeicher verschieben): (Initialisieren wieder mit 1143)
seq(sum(randInt(1,6,5)), X, 1, 200) STO> L1

Nun erzeugen wir ein Histogramm mit Klassenbreite Xscl = 2. (Ymin = -5, Ymax = 50).
Normalverteilungsdichte:
Y1 = normalpdf(X, 5*3.5, √(5*35/12))*200*2
Wir erhalten eine passable Annäherung an die Normalverteilung

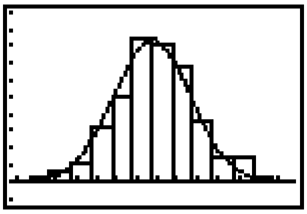

Abb. 18

Für eine größere Anzahl von Zufallsversuchen müssen wir ein kleines Programm WURFK (Menü **PRGM > New > 1:Create New)** schreiben. Der Verarbeitungsteil des Programms ist im Wesentlichen der Bucketsort-Algorithmus:

```
Input "ANZ WUERFEL: ",K
Input "ANZ VERSUCHE:",N
seq(0,X,K,6K)üW
seq(X,X,K,6K)üL
For(J,1,N)
  sum(randInt(1,6,K))üI
  áW(I-K+1)+1üáW(I-K+1)
End
```

Will man auch den Ausgabeteil ins Programm integrieren (Benutzerfreundlichkeit), kann man das Programm erweitern:

```
Input "ANZ WUERFEL: ",K
Input "ANZ VERSUCHE:",N
seq(0,X,K,6K)üW
seq(X,X,K,6K)üL
For(J,1,N)
  sum(randInt(1,6,K))üI
  áW(I-K+1)+1üáW(I-K+1)
End
ClrHome
PlotsOff :FnOff
K-1.5üXmin:6K+1.5üXmax:1üXscl
úmax(áW)/3üYmin:max(áW)*6/5üYmax
0üYscl
"normalpdf(X,K*3.5,ð(K*35/12))*N"üY
Plot1(Histogram,L ,áW)
DispGraph
```

Die Initialisierung mit 1143 ergibt für K = 5 Würfel und N = 1000 Versuche folgendes Bild (Achtung: lange Rechenzeit, die kleinen grauen Männchen im Taschenrechner müssen 5000mal würfeln):

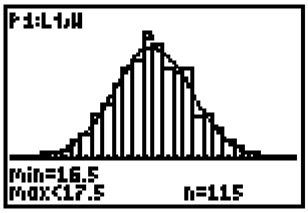

Abb. 19

Für K= 5 Würfel und N = 5000 Versuche (25000 Würfe, Rechenzeit ca. 20 Minuten beim TI-83 Plus!!!) erhalten wir eine ausgezeichnete Anpassung an die Normalverteilung.

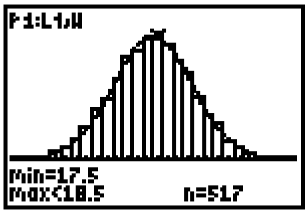

Abb. 20

Der Autor:
Dr. Markus Paul
Peter-Mayr-Str. 19
A-6020 Innsbruck
mailto: markus.paul@utanet.at