

In diesem Merkblatt wird die mathematische Beschreibung eines 1967 ursprünglich für den Computer »LINC-8« der Firma Digital Equipment in FORTRAN-II (!) geschriebenen Programms wiedergegeben (P 125). Der »LINC-8« war ein Doppelcomputer (PDP-8 + LINC) in diskreter Transistortechnologie mit einer Wortlänge von *nur* 12 Bit und einem sehr einfachen Betriebssystem, dem »PROGOFOP« (Program of Operation).

Das Problem: Gegeben sei eine ungruppierte Stichprobe mit N Beobachtungswerten x_i ($i = 1, \dots, N$). Die Stichprobe soll aus einer normalverteilten Grundgesamtheit stammen. Die Parameter μ (Mittelwert) und σ (Standardabweichung) dieser Grundgesamtheit sind unbekannt. Sie werden daher durch die wirksamsten, biasfreien Schätzungen M bzw. s ersetzt [1].

1. Es wird berechnet:

- **Anzahl der Beobachtungen:** N (1)
(Anzahl der x -Werte, number of points, count)
- **Freiheitsgrad der Stichprobe:** $\nu = N - 1$ (2)
(degree of freedom)
- **Summe der Einzelwerte:** $\Sigma X = \sum_{i=1}^N x_i$ (3)
(Summe der x -Werte, sum)
- **Summe der Einzelquadrate:** $\Sigma X^2 = \sum_{i=1}^N x_i^2$ (4)
(Summe der Quadrate der x -Werte)
- **Arithmetischer Mittelwert:** $M = \frac{\Sigma X}{N}$ * (5) {491}
(Durchschnitt, Bestwert, mean)
- **Abweichungen vom Mittelwert:** $\Delta x_i = x_i - M$ (6)
(Scheinbare Fehler, zufällige Fehler)
- **Summe der Abweichungsquadrate:** $S = \sum_{i=1}^N (\Delta x_i)^2$ * (7) {493a}
(Quadratsumme)
- $S = \Sigma X^2 - N \cdot M^2$ (8) $S = \Sigma X^2 - M \cdot \Sigma X$ (9) {493b} + {493c} SUMME SQ
- $S = \Sigma X^2 - \frac{(\Sigma X)^2}{N}$ (10) {493d} SUMME SQ
- **Streuung:** $s^2 = \frac{S}{\nu}$ (11) {492b} STREUUNG
(Varianz, mittlere quadratische Abweichung, variance)
- **Standardabweichung:** $s = \sqrt{s^2}$ (12) {483} ST-ABWCHG.
(Mittlere Abweichung, mittlerer Fehler der Einzelmessung, Genauigkeit der Versuchsanordnung, standard deviation)
- **Standardabweichung des Mittelwerts:** $e = \sqrt{\frac{s^2}{N}} = \frac{s}{\sqrt{N}}$ (13) {495} ABS.FEHLER
(Mittlerer absoluter Fehler des Mittelwerts, Genauigkeit des Mittelwerts, standard error of the mean)
- **Mittlerer relativer Fehler des Mittelwerts:** $r = \frac{e}{M}$ (14) REL.FEHLER
- **Prozentualer Fehler des Mittelwerts:** $r\% = 100 \cdot r$ (15) —
- **Vertrauenswahrscheinlichkeit:** $\alpha = 100 \cdot [1 - 2P]$ (16) 100 [1 - 2P]
- **Signifikanzschranke:** t (17) T
(Aus [1]: Für $N \leq 201$: Tabellen auf Seite 32–35.
Für $N > 201$: $t = 1,96$ für $\alpha = 95\%$,
 $t = 2,58$ für $\alpha = 99\%$)
- **Vertrauensbereich des Einzelwerts x :** $(M - s \cdot t) < x < (M + s \cdot t)$ (18) {581c} ... BIS ...
(Zweiseitiger Vertrauensbereich)
- **Vertrauensbereich des Erwartungswerts μ :** $(M - e \cdot t) < \mu < (M + e \cdot t)$ (19) {580c} ... BIS ...
(Zweiseitiger Vertrauensbereich)

2. Ausgabebezeichner:

- ANZAHL N
- FR-GRAD NY
- SUMME X
- SUMME XX
- M-WERT M
-
- SUMME SQ
- SUMME SQ
- SUMME SQ
- STREUUNG
- ST-ABWCHG.
- ABS.FEHLER
- REL.FEHLER
-
- 100 [1 - 2P]
- T
- ... BIS ...
- ... BIS ...

* Besonderer Algorithmus siehe Punkt 3.

3. Besonderer Algorithmus zur Berechnung von M und S :

Aufgrund der sehr begrenzten Genauigkeit des LINC-8 Computers (Zahlendarstellung) mußte für die Berechnung des Mittelwertes M nach Gl.(5) und der Quadratsumme S nach Gl.(7) ein besonderer Algorithmus ‚erfunden‘ werden, der die Zahlen im Laufe der Berechnung stets klein hält.

Diese Entwicklung eines iterativen Algorithmus war notwendig geworden, nachdem erkannt wurde, daß eine direkte algorithmische Auswertung der Gleichungen (5) bzw. (7) auf dem 12-Bit Computer in vielen Fällen zu falschen (zu ungenauen) Ergebnissen führte.

Der für den LINC-8 damals entwickelte iterative Algorithmus zur Berechnung des arithmetischen Mittelwertes bzw. der Quadratsumme bietet darüberhinaus den großen Vorteil, daß der Umfang der Stichprobe N vor dem Rechenbeginn *nicht bekannt* zu sein braucht. Das ist auch für ganz andere Anwendungen auf modernen Computer von großer Bedeutung, zumal dieser Algorithmus kaum bekannt ist. Mathematisch gesehen sind die im folgenden beschriebenen

Iterations-Algorithmen Gl.(20) bzw. Gl.(21) zu den genannten Gleichungen (5) bzw. (7) äquivalent.

Arithmetischer Mittelwert M :

$$1. \quad i = 0: \quad 0 \Rightarrow M \quad (20a)$$

$$2. \quad (i = 1, 2, 3, \dots, N): \quad \frac{i-1}{i} M_{i-1} + \frac{x_i}{i} \Rightarrow M_i \quad (20b)$$

$$3. \quad i = N: \quad M_N \Rightarrow M \quad (20c)$$

Quadratsumme S :

$$1. \quad i = 0: \quad 0 \Rightarrow S \quad (21a)$$

$$2. \quad (i = 1, 2, 3, \dots, N): \quad \frac{i-1}{i} (x_i - M_{i-1})^2 + S_{i-1} \Rightarrow S_i \quad (21b)$$

$$3. \quad i = N: \quad S_N \Rightarrow S \quad (21c)$$

4. Weitere Kennwerte der Stichprobe:

■ **Minimum:** x_{\min} (22)
(kleinster Wert, smallest value)

■ **Maximum:** x_{\max} (23)
(größter Wert, largest value)

■ **Extrembereich:** $\Delta x = x_{\max} - x_{\min}$ (24)
(Spannweite, range)

■ **Variationskoeffizient:** $v = \frac{s}{M}$ (25)
(coefficient of variation)

■ **Geometrischer Mittelwert:** $G = \sqrt[N]{\prod_{i=1}^N x_i}$ (26)
(geometric mean)

■ **Harmonischer Mittelwert:** $H = N / \sum_{i=1}^N \frac{1}{x_i}$ (27)
(harmonic mean)

■ **Die Momente 1., 2., 3. und 4. Ordnung:** $m_1 = \frac{1}{N} \cdot \sum (\Delta x_i)^1$ (28)

(Statistische Momente) $m_2 = \frac{1}{N} \cdot \sum (\Delta x_i)^2$ (29)

$$m_3 = \frac{1}{N} \cdot \sum (\Delta x_i)^3 \quad (30)$$

$$m_4 = \frac{1}{N} \cdot \sum (\Delta x_i)^4 \quad (31)$$

■ **Abweichung von der Symmetrie:** $w = \frac{m_3}{m_2 \sqrt{m_1^2}}$ (32)
(Schiefe der Verteilung, skewness, bei völliger Symmetrie ist $w = 0$)

■ **Abweichung von der Normalverteilung:** $k = \frac{m_4}{m_2^2} - 3$ (33)
(Kurtosis, bei Normalverteilung ist $k = 0$, $k > 0$ bedeutet Datenkonzentration in der Mitte, bei $k < 0$ liegen viele Daten in den Schwänzen (tails) der Verteilung)

5. Sonstige Hinweise:

■ Die unter Punkt 2 angegebenen Ausgabebezeichnungen beziehen sich ausschließlich auf den Protokollausdruck des ursprünglichen Programms (P 125).

■ Auf modernen Arbeitsplatzcomputern läßt sich heute das Berechnen der statistischen Kennwerte einer Stichprobe mit sehr vielen Anwendungsprogrammen quasi nebenbei erledigen. [Anm. im Mai 1992]

■ Ein Beispiel für eine solche Applikation ist das **KaleidaGraph** in der Macintosh-Welt, das mit dem Befehl »Statistics...« aus dem Menü »Functions« alle statistischen Kennwerte berechnet und anzeigt [2]. [Anm. im Juli 1994]

6. Literatur:

[1] Documenta Geigy: Wissenschaftliche Tabellen. Basel: Geigy 1969 (7. Auflage). – Die in diesem Merkblatt in { ... } angegebenen Gleichungsnummern beziehen sich auf dieses Tabellenwerk ab Seite 146. Die Tabellen für die Signifikanzschranken sind zu finden auf den Seiten 32–35 ($N \leq 201$) und Seite 31 unten ($N > 201$). – [BU 356].

[2] Dittberner, K.-H.: MAC-Software: KaleidaGraph V 3.0. FU Berlin (IfP): wdv-notes Nr. 269, 1993–1994.

Notizen: