

# Statistik verständlich mit J

Joachim Hoffmann

In diesem Artikel wird anschaulich dargestellt, wie einfach Statistik mittels J ausgedrückt werden kann. J erlaubt es Formeln so zu implementieren, daß die mathematischen Zusammenhänge klar zutage treten, die leider allzuoft durch die historisch gewachsene - und daher inkonsistente - Syntax und Semantik der konventionellen mathematischen Notation (KMN) verschleiert werden.

J ist eine *einfache*, in sich widerspruchsfreie und am Computer ausführbare mathematische Notation. Auf <http://www.jsoftware.com> ist J einschließlich Dokumentation kostenlos erhältlich.

### Verben, Adverbien, Konjunktionen, ...

Dem Erfinder von J, Kenneth E. Iverson, ist es gelungen, eine Verbindung zwischen menschlichen Sprachen und technischen Schreibweisen herzustellen, die sich unter anderem in der Bezeichnung der Wortarten ausdrückt:

KMN	J
Zahl; Variable	Nomen; Pronomen
Funktion (z.B. +)	Verb (Aktionswort)
Operator (mit einem Arg.)	Adverb
Operator (mit zwei Arg.)	Konjunktion (Bindewort)

Durch diese Wesensverwandtschaft zur Sprache verlieren schwierige mathematische Konzepte einiges an Schrecken. Z.B. ein Operator, eine Funktion die wiederum Funktionen als Argumente hat, heißt in J Konjunktion oder Bindewort, und funktioniert in der gleichen Weise wie das Bindewort „and“ in der Phrase „run and hide“, das die beiden Verben „run“ und „hide“ zu einer übergeordneten Handlung verbindet.

Weitere Beispiele sind weiter unten angeführt.

### Das arithmetische Mittel

**In Worten:** Das Arithmetische Mittel (am) ist die Summe dividiert durch die Anzahl der Elemente einer Zahlenliste.

KMN:

$$am = \frac{\sum_{i=1}^n x_i}{n} \quad n = \text{Anzahl der Elemente}$$

J:

- sum=: +/            Plus eingefügt
- div=: %            Dividieren
- anz=: #            Anzahl der Elemente

am=: sum div anz  
am 1 2 3 4 5

3

Bitte vergleichen Sie die drei Varianten. Die Definition in J ist rein funktional, d.h. ohne die explizite Erwähnung von Argumenten, wie z.B.  $x_i$  oder  $n$  in KMN, und kommt der Definition in Worten schon sehr nahe. Im Folgenden wird die J-Definition Schritt für Schritt erklärt.

### Das Einfüge-Adverb /

1+2+3+4+5  
15  
+/ 1 2 3 4 5  
15

Das Adverb Einfügen (/) fügt sein linkes Verb-Argument (+) zwischen die Elemente der Liste ein. +/ entspricht ungefähr  $\sum$ . Das Adverb / (in KMN Operator) verändert das Verb + (in KMN Funktion) auf ähnliche Weise, wie das Adverb „ökonomisch“ das Verb „denken“ zu der neuen Einheit „ökonomisch denken“ modifiziert. Vom programmiertechnischen Standpunkt aus, ist verb/ einfach einzutippen, schleifenfrei und schnell.

Generalisierend zur KMN läßt sich das Einfüge-Adverb auf jedes Zahlenfeld (Vektor, Matrix, Tensor, ...) anwenden und arbeitet mit jeder beliebigen Funktion (fn/) zusammen. Die KMN beschränkt hier auf einige wenige Symbole  $\sum \prod \Delta$  etc..

+/             $\Leftrightarrow \sum_{i=1}^n (1-)^{n-1} x_i$

-/             $\Leftrightarrow \sum_{i=1}^n x_i$

\*/             $\Leftrightarrow \prod_{i=1}^n x_i$

Aber wie wird z.B. das Maximum über eine Liste in KMN ausgedrückt? (>./  $\Leftrightarrow \max\{x|x \in I\}$ ). Oder der fortschreitende Exponent? ( $\wedge$  /  $\Leftrightarrow$  ?) Für den Exponenten gibt es in KMN kein Symbol, nur den hochgestellten Index  $x^y$ . Daher ist es etwas schwierig auf die Idee zu kommen, den „Exponent“ zwischen die Elemente eine Liste einzufügen.

### Die Gabel

Warum kann (sum div anz) das AM sein?

Eine isolierte Abfolge von drei Verben (KMN: Funktionen) hat Kenneth E. Iverson, als Gabel definiert:

1st=: 1 2 3 4 5            Liste definieren

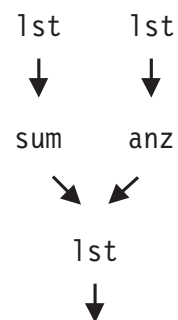
Die Gabel - eine isolierte Abfolge von drei Verben:

(sum div anz) 1st  
3

Zur Veranschaulichung die Gabel ausgeschrieben:

```
(sum 1st) div (anz 1st)
3
am            Verb am anzeigen
┌── sum
│   div      Tree display
└── anz
```

Das Argument 1st wird zuerst von den

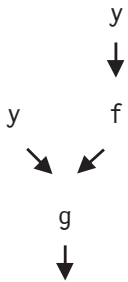


äußeren Verben sum, anz verarbeitet und die Zwischenergebnisse werden an das mittlere Verb div weitergegeben.

Die Gabel ist ein universelles Konzept, das auf allen Gebieten der Mathematik, Technik und EDV allgegenwärtig ist, aber von der Allgemeinheit leider nicht wahrgenommen wird.

**Der Haken**

Neben der Gabel hat Iverson nur noch den *Haken* definiert:



$$(f\ g)\ y \Leftrightarrow (y\ f\ g)\ y$$

Gabel und Haken legen nur den Datenfluß zwischen isolierten Abfolgen von Verben fest. Aus Gabeln und Haken läßt sich unter Zuhilfenahme von Klammern jeder beliebige *Funktionsbaum* zusammensetzen.

**Die Abweichung vom AM** ist ein typischer Haken:

```
lst=: 1 2 3 4 5  Liste definieren
lst - am lst  Abweichung vom AM
_2 _1 0 1 2

(- am) lst  Abweichung als Haken
_2 _1 0 1 2
```

*Das harmonische Mittel*

**In Worten:** Das harmonische Mittel (HM) ist der Kehrwert des arithmetischen Mittels der Kehrwerte.

$$KMN: HM = \frac{1}{\frac{\sum_{i=1}^n \frac{1}{x_i}}{n}} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

**J:**

```
ALLE =: "_  Verb-Rang Unendlich
NACH =: "@"  Hintereinanderausführen
UNTER=: "&." (f&g) = (g INVERS) f g
```

Das HM klassisch mittels Hintereinanderausführen:

```
hm=: div NACH am NACH (div ALLE)
```

Das HM etwas kürzer mit der Konjunktion UNTER:

```
hm=: am UNTER (div ALLE)
hm lst
2.18978
```

Bitte vergleichen Sie wieder die drei Varianten.

Das Bindewort UNTER führt zuerst sein rechtes Verb (div) aus, das könnte man als „Vorbereitung“ bezeichnen, dann das linke Verb (am), und zum Schluß wird die „Vorbereitung“ wieder rückgängig gemacht, indem die automatisch errechnete Umkehrfunktion des rechten Verbs (div) ausgeführt wird. Der Kehrwert ist angenehmerweise selbstinvers. Das Adverb ALLE sorgt dafür, daß die Skalarfunktion div nicht einzelne Zahlen, sondern die *ganze* Liste lst verarbeitet und an das nachfolgende Verb am weitergibt.

*Das geometrische Mittel*

**In Worten:** Das geometrische Mittel (GM) ist die n-te Wurzel aus dem Produkt über die Liste.

**K M N :**

$$GM = \sqrt[n]{x_1 x_2 \dots x_n} = \sqrt[n]{\prod_{i=1}^n x_i} \quad \text{mit } x_i > 0$$

**J:**

```
wzl=: %:  Wurzel
exp=: ^  Exponent
log=: ^.  Logarithmus
```

Klassisch - Das GM als Gabel:

```
gm =: anz wzl */
gm lst
2.60517
```

Besonders elegant - GM als AM unter dem Logarithmus:

```
gm =: am UNTER (log ALLE)
```

```
gm lst
2.60517
```

Das gleiche ausgeschrieben, ohne die Konjunktion UNTER:

```
exp am log 1 2 3 4 5
2.60517
```

Als kleine Übung versuchen Sie bitte die letzten beiden Definitionen des GM in KMN auszudrücken.

Das GM ergibt sich aus am UNTER log, weil das Addieren von Logarithmen der Multiplikation entspricht:

```
3 (+ UNTER log) 4
12
exp (log 3) + (log 4)
12
3*4
12
```

*Weiterführende Literatur*

„Teaching with J - An Example from Statistics“ von Prof. Keith Smillie, Univ. of Alberta, CA; (smillie@cs.ualberta.ca):

abw=: - am	Abweichung
qus=: sum @ sqr @ abw	Quadrat-Sum.
var=: qus div (1&+)@anz	Varianz
saw=: wzl @ var	Std.-Abweichung
kp =: sum . *	Matrizenprodukt
skp=: (kp~)  :@abw	Sum. Kreuzprod.
kov=: skp div (1&+)@anz	Kovarianz
kor=: kov div */~@saw	Korrelation
usw.	