

WIE VERMEIDET MAN RUNDUNGSFEHLER?

von David Green

Originaltitel:
Eliminating Bias Due to Rounding

Aus: Teaching Statistics, Vol 12 No 3, Autumn 1990

Übersetzt von Ingeborg Strauß, Kronberg im Taunus

In einem früheren Artikel [Vol 12 No 1] diskutierte Shrikant Bangdiwala die folgenden beiden Rundungs-Regeln:

Regel 1:

Ist die letzte Ziffer eine 0,1,2,3 oder 4 ist, lasse alle übrigen Ziffern unbeeinträchtigt (d.h. runde ab).

Ist die letzte Ziffer eine 5,6,7,8 oder 9 ist, erhöhe die vorletzte Ziffer um 1 (d.h. runde auf).

Regel 2:

Ist die letzte Ziffer eine 0,1,2,3 oder 4 ist, runde ab.

Ist die letzte Ziffer eine 6,7,8 oder 9 ist, runde auf.

Ist die letzte Ziffer eine 5 und die vorletzte Ziffer gerade, runde ab; andernfalls runde auf.

Der Autor gab Gründe dafür an, daß er Regel 1 favorisiert. Aber wachsame Leser schrieben uns und demonstrierten, daß Regel 2 die korrekte ist. Ein Leser beobachtete z.B. den mittleren Fehler, der bei 20 aufeinanderfolgenden ganzen Zahlen im einen oder anderen Fall entsteht:

Ohne Beschränkung der Allgemeinheit nehme man die zweistelligen Zahlen von 11 bis 30 und runde auf eine Stelle. Dabei ergeben sich die folgenden Fehler:

Regel 1:

-1,-2,-3,-4,5,4,3,2,1,0,-1,-2,-3,-4, 5,4,3,2,1,0;
mittlerer Fehler 0.5.

Regel 2:

-1,-2,-3,-4,5,4,3,2,1,0,-1,-2,-3,-4,-5,4,3,2,1,0;
mittlerer Fehler 0.

Derselbe Effekt tritt auf, wenn man Dezimal-Zahlen mit beliebiger Genauigkeit rundet.

In einem zweiten Leserbrief wurde angemerkt, daß es sich bei Regel 2 um die Britische und Internationale Standardregel handelt. Der springende Punkt ist, daß Nullen nicht zu runden sind und daher beim Runden keinen Einfluß haben sollten. (Das ist nicht selbstverständlich. Von daher wäre es besser, in Regel 2 explizit darauf hinzuweisen.) Hier ein Ausschnitt aus diesem instruktiven Brief:

"Nullen bedürfen keiner Rundung. Also sind nur 9 Ziffern zu betrachten. Wird die 5 stets aufgerundet, resultiert eine 5:4-Abweichung. Die 'Runde-gerade-Ziffer'-Regel korrigiert dies zu $4\frac{1}{2}:4\frac{1}{2}$ - und hat den zusätzlichen Vorteil, daß die aufeinanderfolgende Mittelung von ('benachbarten') Wertepaaren ebenfalls korrekt wird. Man nehme zum Beispiel 4.5 und 5.5:

Regel 1:

4.5 -> 5, 5.5 -> 6; Mittelwert 5.5 ist falsch.

Regel 2:

4.5 -> 4, 5.5 -> 6; Mittelwert 5 ist richtig.

(Natürlich ergeben sich bei Wertepaaren wie etwa 4.5, 6.5 immer noch Rundungsfehler, aber, betrachtet man alle möglichen Wertepaare, eben keine systematischen Fehler.)"

Shrikant Bangdiwala läßt sich durch diese Argumente nicht überzeugen. Ihm ist besonders lästig, daß durch Regel 2 die eine der Rundung unterliegende Ziffernmenge größer ist als die andere. Er betrachtet z.B. die Zahlenmenge ..., -0.2, -0.1, 0.0, 0.1, 0.2, ... Rundet man auf die Einerstelle, wird die Menge aller ganzen Zahlen produziert. Bei Regel 1 erscheint jede ganze Zahl 10 Mal; bei Regel 2 dagegen werden die ungeraden ganzen Zahlen jeweils 9 Mal und die geraden ganzen Zahlen jeweils 11 Mal generiert.

Zum Schluß noch der Beitrag eines weiteren Leserbriefschreibers, der ebenfalls die Überlegenheit (und Korrektheit) von Regel 2 herausstreicht und eine treffende Erklärung beisteuert. Als erstes betrachte man eine Menge von Zahlen zwischen 0 und 1 mit, sagen wir, 3 dezimalen Stellen. Es soll auf die nächstgelegene ganze Zahl gerundet werden. Um dabei Fehler zu vermeiden, benötigt man für jede aufgerundete Zahl eine andere, die abgerundet wird. So wird z.B. 0.123 abgerundet und das 'Komplement' 0.877 aufgerundet. Die Ausnahme ist 0.500, die kein Komplement besitzt. Egal, wie man runden würde, immer ergäbe sich ein Fehler. Genausowenig praktikabel wäre, diese Zahl in der Hälfte der Fälle, wenn sie auftritt, aufzurunden und ansonsten abzurunden. Daher muß ein anderer Weg gesucht werden, um die notwendige Balance zu erreichen, wofür Regel 2 sorgt. Das im Beispiel gewählte Intervall beschränkt nicht die Allgemeingültigkeit. Im Bereich zwischen 0 und 2 bewirkt Regel 2, daß 0.500 heruntergeht und 1.500 heraufgeht, der Mittel-

wert ist unverfälscht. Überträgt man diese Idee auf größere oder feiner unterteilte Intervalle, sieht man, daß die kleinen individuellen Fehler sich gegenseitig aufheben. (Wie schade, daß wir nicht gewöhnt sind, mit der Basis 11 zu arbeiten!)

Shrikant Bangdiwala ist speziell an statistischen Häufigkeiten interessiert. Es ist jedoch der Summationsprozeß (oder die Berechnung des Mittelwertes) bei einer Datenmenge, der von noch größerer Bedeutung ist. Fazit: Regel 2 regiert! Einverstanden?