

## Eliminierung von Ausreißern

Bei der Ableitung statistischer Aussagen aus vorliegenden Meßdaten können übermäßig fehlerbehaftete Daten stark verfälschend auf die zu treffende Aussage wirken. Meßverfahren, die in der Nähe der Auflösungsgrenze arbeiten, liefern zwangsläufig stärker fehlerbehaftete Meßdaten. Hier macht es sich für eine gesicherte Aussage erforderlich, die übermäßig fehlerbehafteten Meßdaten (*Ausreißer*) zu eliminieren. Eine der am häufigsten angewendeten Methode zur Ausreißerererkennung ist der NALIMOV-Test/1/.

Zur Erkennung der Ausreißer ist es nach dem NALIMOV-Test erforderlich, Mittelwert  $\bar{x}$  und Standardabweichung  $\sigma$  aus den Meßdaten zu berechnen. Mit einer statistischen Sicherheit  $p$  kann nun entschieden werden, ob der Meßwert  $x_i$  einen Ausreißer darstellt.

Zum Test wird die Beziehung  $|x_i - \bar{x}| \geq N(p)\sigma$  ( $1 \leq i \leq n$ ) herangezogen. Wird diese Beziehung durch den Meßwert  $x_i$  bei einer statistischen Sicherheit  $p = 0.99$  erfüllt, handelt es sich bei diesem Meßwert signifikant um einen Ausreißer. Die als NALIMOV-Faktoren  $N(p)$  bezeichneten Zahlenwerte sind von der Anzahl der Meßwerte  $n$  (Stichprobenumfang) abhängig und liegen tabelliert vor. In Ergänzung zu dem in/2/ vorgestellten Test auf Normalverteilung einer Stichprobe wurde ein Programm zur Durchführung des NALIMOV-Tests im Zusammenwirken mit dem Test auf Nor-

malverteilung entworfen. Die Vorgehensweise bei der Testdurchführung soll anhand der in den Abb. 1 bis 3 gezeigten Druckerprotokolle erläutert werden.

Eine Stichprobe von 62 Meßwerten sollte auf Normalverteilung getestet werden. Abb. 1 zeigt das in/2/ erwähnte Protokoll.

```

*** TEST AUF NORMALVERTEILUNG EINER STICHPROBE ***
=====
STICHPROBENUMFANG * 62
MAXIMALWERT * 50.0000
MINIMALWERT * 1.0000
KLASSENANZAHL * 9.0000
KLASSENBREITE * 5.4844
MITTELWERT * 11.6290
STANDARDABWEICHUNG * 6.6978

KLASSEN MITTEL HÄUFIGKEIT WAHRSCHEINLICHKEIT
ABSOLUT RELATIV NORMALVERTEILUNG
-----
3.72 9.000 0.145 0.160
9.17 27.000 0.435 0.321
14.61 21.000 0.330 0.306
20.06 4.000 0.065 0.138
25.50 0.000 0.000 0.030
29.94 0.000 0.000 0.003
36.39 0.000 0.000 0.000
41.83 0.000 0.000 0.000
47.28 1.000 0.016 0.000

SIGNIFIKANZWÄHRSCHEINLICHKEIT = 1%
FREIHEITSGRADE 6
TESTGRÖSSE U 2 = 436822
TAFELWERT CHI 2 = 17

HYPOTHESE AUF NORMALVERTEILUNG DER GRUNDGESAMTHEIT
DER STICHPROBE WIRD ABGELEHNT

```

Abb. 1 Test einer Stichprobe auf Normalverteilung

```

*** NALIMOV - TEST ***
=====
STICHPROBENUMFANG * 62
MITTELWERT * 11.6290
STANDARDABWEICHUNG * 6.6978
NEUER STICHPROBENUMFANG * 61
ELIMINIERTER SIGNIFIKANTE AUSREISSER * 1

```

Abb. 2 Eliminierung des Ausreißers der Stichprobe nach Abb. 1

```

*** TEST AUF NORMALVERTEILUNG EINER STICHPROBE ***
=====
STICHPROBENUMFANG * 61
MAXIMALWERT * 21.0000
MINIMALWERT * 1.0000
KLASSENANZAHL * 9.0000
KLASSENBREITE * 2.2222
MITTELWERT * 11.0000
STANDARDABWEICHUNG * 4.8461

KLASSEN MITTEL HÄUFIGKEIT WAHRSCHEINLICHKEIT
ABSOLUT RELATIV NORMALVERTEILUNG
-----
2.11 4.000 0.066 0.034
4.33 5.000 0.082 0.071
6.56 7.000 0.115 0.120
8.78 9.000 0.148 0.164
11.00 11.000 0.180 0.182
13.22 9.000 0.148 0.164
15.44 7.000 0.115 0.120
17.67 5.000 0.082 0.071
19.89 4.000 0.066 0.034

SIGNIFIKANZWÄHRSCHEINLICHKEIT = 1%
FREIHEITSGRADE 6
TESTGRÖSSE U 2 = 6
TAFELWERT CHI 2 = 17

HYPOTHESE AUF NORMALVERTEILUNG DER GRUNDGESAMTHEIT
DER STICHPROBE WIRD NICHT ABGELEHNT

```

Abb. 3 Test der bereinigten Stichprobe auf Normalverteilung

Die Häufigkeitsverteilung der Stichprobe weicht deutlich von der Normalverteilung ab, weshalb die Hypothese auf Normalverteilung zwangsläufig abgelehnt werden muß. Aus der Häufigkeitstabelle ist ersichtlich, daß es nur einen Wert im Bereich von 44.56 bis 50.00 gibt, während sich die anderen 61 Werte über einen Bereich von 1.00 bis 22.78 erstrecken. Mit dem NALIMOV-Test werden die 62 Meßwerte nun auf einen Ausreißer hin

untersucht. Hierzu werden durch das Programm auf Magnetband abgespeicherte Datenfiles geladen, welche die Stichprobe einschließlich Stichprobenumfang, Mittelwert und Standardabweichung und die NALIMOV-Faktoren  $N(p)$  für eine statistische Sicherheit von  $p = 0.99$  und für einen Stichprobenumfang bis zu 100 Werten enthalten. Der sich nun anschließende Test verläuft nach der oben angegebenen Beziehung. Erkannte Ausreißer werden aus der Datenmenge gestrichen. Der sich nach Abb. 1 bereits andeutende Ausreißer wurde als solcher erkannt, weshalb der neue Stichprobenumfang auf 61 reduziert wurde (Abb. 2). Das bereinigte Datenmaterial wird wieder auf ein Datenfile abgelegt und kann erneut auf Normalverteilung getestet werden. Abb. 3 zeigt das nunmehr erhaltene Ergebnis. Die Programmorganisation wurde so vorgenommen, daß über den Dialog das Anketten des jeweils erforderlichen Programms erfolgen kann. Anderenfalls wird das jeweils laufende Programm definiert beendet.

Mit dem erläuterten Rechenprogramm kann der in/2/ beschriebene Test auf Normalverteilung vorteilhaft ergänzt werden. Übermäßig fehlerbehaftete Meßdaten lassen sich auf die beschriebene Weise effektiv aus der zu bearbeitenden Datenmenge eliminieren. Die hier gezeigte Verbindung mit dem Test auf Normalverteilung stellt dabei nur ein mögliches Beispiel dar.

Claus Kühnel

## Literatur

- /1/ Metzner, K.: Gaschromatografische Spurenanalyse. Akademische Verlagsgesellschaft Geest & Portig, Leipzig 1977
- /2/ Kühnel, C.: Statistischer Test auf Normalverteilung der Grundgesamtheit einer Stichprobe mit dem  $\chi^2$  - (Chi-Quadrat-)Anpassungstest. rechentechnik/datenverarbeitung (zur Veröffentlichung vorgesehen)

## Sichere Dateiverarbeitung auch bei Folgebändern (Multi-Volume-Dataset)

Beim Ausgeben von Datenbeständen auf Magnetbändern läßt sich häufig der Datenumfang nicht im voraus exakt bestimmen. Beim Überschreiten der Speicherkapazität eines Magnetbandes werden in einem solchen Fall vom Betriebssystem OS/ES während der Ausgabe Arbeitsbänder als Folgebänder angefordert. Die Systemnachricht zum Montieren dieser Datenträger führt in der Regel dazu, daß die Daten nicht als Folgebänder erkannt und entsprechend weitergereicht werden, sondern nur als Arbeitsbänder behandelt werden. Wegen der dabei auftretenden Datenverluste kommt es zu Wiederholarbeiten erheblichen Umfangs. Im Rahmen einer NVE wurde auf der Basis einer Veränderung im Betriebssystem OS/ES die *Technologie Folgebänder* erarbeitet. Die Lösung sichert:

- die eindeutige Erkennung nichtgeplanter Folgebänderanforderungen
- einen einheitlichen technologischen Ablauf während der Jobbearbeitung sowie
- die Weitergabe der Datenbestände zur Verarbeitung. Interessenten wenden sich bitte an den VEB ZOD Bauwesen Berlin, Bereich Produktionsdurchführung, 1086 Berlin, Charlottenstraße 36, PSF 1254, Tel.: 2 05 22 56.

Wäsch, Berlin

## Neuerscheinungen

- Anwendung der Mikroelektronik in der Metallurgie. FFHB 243. Von einem Autorenkollektiv. (Freiberger Forschungshefte, Gruppe Metallurgie und Werkstofftechnik.) 1. Aufl. Etwa 112 S. mit 34 Abb. und 4 Tab. Br. Etwa 34,- M. VEB Deutscher Verlag für Grundstoffindustrie
- Löffler, H., Rechnerverbundsysteme. (Informatik – Kybernetik – Rechentechnik.) 1. Aufl. Etwa 250 S. mit 176 Abb. und 12 Tab. L 6. KE. Etwa 54,- M. Akademie Verlag