AUSSIEBEN VON STATISTISCHEN AUSREISSERN

AUSGANGSLAGE

Ein beliebter Indikator bei der Auswertung von Zahlenreihen ist die Standardabweichung, bezogen auf den berechneten Mittelwert. Dabei geht die Modellannahme davon aus, dass die Häufigkeit zwischen Einzelwert und Mittelwert mit zunehmendem Unterschied abnimmt und in der Gänze einer gaussschen Normalverteilung nahekommt.

Allerdings ergibt sich auch für jeden Daten - Mix, der in keiner Weise der Normalverteilung entspricht, aus Berechnung immer ein Wert der Standardabweichung. Es ist daher erforderlich, nachträglich zu überprüfen, ob sich der Zahlen – Mix als Ganzes tatsächlich innerhalb der vorausgesetzten Normalverteilung hält – sofern dies von der Datenlieferung erwartet wird.

Eine solche Überprüfung kann beispielsweise über den Link: https://www.graphpad.com/quickcalcs/grubbs1/ für bis zu 2000 Einzelwerte erfolgen. Das Ergebnis liefert in Sekundenschnelle die Antwort, welcher maximale Z – Wert (als Multiplikator zur festgestellten Standardabweichung) sich für die vorhandene Anzahl erfasster Werte gerade noch einstellen darf. Dabei ist dieser Z – Wert NICHT von der Standardabweichung selbst, sondern nur vom gewählten Signifikanz – Niveau (i. d. R. $\alpha = 0.01$ oder $\alpha = 0.05$) sowie von der Anzahl erfasster Einzelwerte abhängig.

DAS PROBLEM UND SEINE LÖSUNG

Zwar liefert das digitalisierte «Verfahren Grubbs» (siehe Link) nicht nur den maximal zulässigen Z – Wert mit Direktverweis auf den (allfälligen) grössten Ausreisser, sondern es liefert auch für jeden der «n» Einzelwerte den individuell vorhandenen Z – Wert.

Die Schwierigkeit ist nun nicht so sehr, dass über eine allfällige Grosszahl von Einzelwerten all jene «von Hand» herausgesucht und eliminiert werden müssen, die einen zu grossen Z – Wert aufweisen. Das kann allenfalls mühsam sein.

Das **PROBLEM** liegt darin, dass eine erstmals ausgesiebte Zahlenreihe bei einer Neuberechnung nunmehr eine kleinere Standardabweichung (SD) aufweist, da die Werte mit den gröbsten Abweichungen zum Mittelwert als Ausreisser entfernt wurden. Weil sich der Z – Wert bei grösseren «Stückzahlen» dabei kaum ändert, folgt für die Zweitüberprüfung mit tieferer SD ein gegenüber der Erstüberprüfung u.U. deutlich

reduzierter Grenzwert der Abweichung zum neuen Mittelwert. Es kann daher sehr gut sein, dass ein erstes Aussieben der Ausreisser dazu führt, dass für den um die Ausreisser reduzierten zweiten Daten – Mix neue Ausreisser auftauchen. Dieses Prozedere kann sich wiederholt einstellen und muss so lange wiederholt werden, bis von Untersuch zu Untersuch keine Veränderung mehr erfolgt.

Statt sich des hilfreichen Online – Programms via Link zu bedienen, um damit alle Ausreisser einer Datenlieferung schrittweise zu eliminieren, kann als **LÖSUNG** das in die vorliegende ZIP – Datei eingefügte Excel – File verwendet werden. Dieses vollzieht nach Eingabe der Messwerte (Spalte A) alle weiteren Schritte automatisch und liefert – nach vorgängiger Wahl des Signifikanz – Niveaus – ohne weiteres Dazutun alle «sauberen» Einzelwerte, auch mit Angabe der Anzahl gefundener Ausreisser. Das im Excel – File eingetragene (und schreibgeschützte) Beispiel verdeutlicht die Problematik von stets neu auftauchenden Ausreissern.

Oktober 2026