



TQU VERLAG

**Intelligente
Ausreißer-
behandlung**

Probieren und Studieren

Ausreißerbehandlung

Intelligent Ausreißer erkennen und bewerten

[Autoren: Dr. Walter Jahn](#)
[und Dr. Konrad Reuter](#)

In der Praxis der Datenauswertung ist immer wieder zu beobachten, dass Daten auftreten, die nicht so recht plausibel sind. Dafür wird gern der Begriff "Ausreißer" verwendet. Nicht selten wird diesen Daten keinerlei Beachtung geschenkt und die Auswertungen werden munter weiter betrieben. Die Fragwürdigkeit der Ergebnisse ist dann offensichtlich. Zunächst steht die Aufgabe, solche Ausreißer zu erkennen und nach festem Kriterium zu bewerten. Für diese Bewertung sind sowohl grafische als auch rechnerische statistische Methoden entwickelt worden. Es steht jedoch die Frage, ob diese Tests nach ihren Eigenschaften auch von Ausreißern selbst zu durchzuführen sind und der gesunde Menschenverstand nie auszuschalten ist.

Die Wissenschaftsgeschichte hat Beispiele parat, wo Messergebnisse, die zunächst vom Mainstream für Ausreißer gehalten worden waren, sich als bisher unbekannte Effekte herausgestellt hatten. Häufig hatten es diese Entdecker schwer, sich mit ihrer Erkenntnis durchzusetzen.

Um grobe Eingabe- oder Übertragungsfehler zu verhindern, kann bereits mit der Software gegengesteuert werden. EXCEL 2000, 2003 verfügen mit der Funktionalität "Gültigkeit" und EXCEL 2010 mit der analogen Funktion "Datenüberprüfung" über geeignete Mittel.

Die Komplexität der Aufgabenstellung führt zur Eliminierung von Ausreißern aus den Datensätzen sollte formale Kriterien sein.

Mit Berechnungen zu univariaten Datensätzen, hier ein unimodales Verteilungsgesetz, ist dies umso bemerkenswerter, als bei der Berechnung des "Aerk" Kriteriums die ausreißerverdächtigen Daten nicht für Parameterschätzungen verwendet werden.

Es wird verdeutlicht, dass univariate Methoden auf multivariate Datensätze angewendet nicht zum Erfolg führen, weil die Abhängigkeitsstruktur unberücksichtigt bleibt.

Ansprechpartner: Dr. Konrad Reuter
Telefon: 0171/6006604

TQU Verlag, Magirus-Deutz-Straße 18, 89077 Ulm Deutschland, Telefon 0731/14660200, verlag@tqu-group.com, www.tqu-verlag.com

QUALITY APPS Applikationen für das Qualitätsmanagement

Lizenzvereinbarung

Dieses Produkt Ausreißerbehandlung wurde von den Autoren Dr. Walter Jahn und Dr. Konrad Reuter mit großem Aufwand und großer Sorgfalt hergestellt.

Dieses Werk ist urheberrechtlich geschützt (©). Die dadurch begründeten Rechte, insbesondere die der Weitergabe, der Übersetzung, des Kopierens, der Entnahme von Teilen oder der Speicherung bleiben vorbehalten.

Bei Fehlern, die zu einer wesentlichen Beeinträchtigung der Nutzung dieses Softwareproduktes führen, leisten wir kostenlos Ersatz. Beschreibungen und Funktionen verstehen sich als Beschreibung von Nutzungsmöglichkeiten und nicht als verbindliche Zusicherung bestimmter Eigenschaften. Wir übernehmen keine Gewähr dafür, dass die angebotenen Lösungen für bestimmte vom Kunden beabsichtigte Zwecke geeignet sind.

Sie erklären sich damit einverstanden, dieses Produkt nur für Ihre eigene Arbeit und für die Information innerhalb Ihres Unternehmens zu verwenden. Sollten Sie es in anderer Form, insbesondere in Schulungs- und Informationsmaßnahmen bei anderen Unternehmen (Beratung, Schulungseinführung etc.) verwenden wollen, setzen Sie sich unter dieser Voraussetzung mit der entsprechenden Vereinbarung im Voraus mit uns in Verbindung. Unser Produkte werden kontinuierlich weiterentwickelt. Bitte melden Sie sich, wenn Sie neue Daten für uns haben.

Alle Ergebnisse basieren auf den vom Autor eingesetzten Formeln und müssen vom Anwender sorgfältig geprüft werden.

Die berechneten Ergebnisse sind als Hinweise und Anregungen zu verstehen. Das APP enthält Makros, die vor der Benutzung aktiviert werden müssen.

Wir wünschen viel Spaß und Erfolg mit dieser Applikation.

TQU Verlag, Magirus-Deutz-Straße 18, 89077 Ulm Deutschland, Telefon 0731/14660200, verlag@tqu-group.com, www.tqu-verlag.com

Anwendungshinweise

1. Statistik

1.1 zum Begriff "Ausreißer" (*outlier*)

Ausreißer sind nach allgemeiner Auffassung der statistischen Literatur, überraschend extrem liegende Werte, wobei der Beobachter ein Modell voraussetzt und den "Ausreißer" als Modellabweichung auffasst [4].

Anders ausgedrückt sind Ausreißer Werte, die nicht zur beobachteten Grundgesamtheit gehören.

Für beliebige Verteilungen gilt die Ungleichung nach Tschebyscheff.

Ausreißer können entstehen als Fehler in der Stichprobe, Messfehler, Eingabefehler, Übertragungsfehler und Rechenfehler. Saubere statistische Entscheidungen sind nur selten möglich [4].

Es ist also erforderlich, solche Daten sachlogisch zu klären. Nur dann sollte eine Elimination erfolgen, die unbedingt zu protokollieren ist. Siehe auch [2].

Excel (und andere Software) kann die Dateneingabe überwachen und damit weitgehend formale Fehler unterbinden (Daten - Gültigkeit).

Siehe als einen Datenparameter der Verteilung geschätzt werden sollen, können sich "robuste" Schätzverfahren, die gegenüber Ausreißern weniger empfindlich sind.

Als Beispiel für die Schätzung des Mittelwertes wird der gestutzte Mittelwert (*trimmed mean*) vorgestellt. Der Grad der Stützung ist wählbar.

Zur Identifikation von Ausreißern sind verschiedene grafische und rechnerische Methoden brauchbar, von denen einige wichtiger sind als andere.

Zusätzlich sind die folgenden Grundsätze zu beachten:

- Eingabefehler - Prüfung der originalen Quelldaten
- ist der Ausreißer technisch nach dem Stand der Wissenschaft möglich?
Achtung! Es gibt auch neue Erkenntnisse. Die Wissenschaftsgeschichte hält hierzu Beispiele parat!
- ist die Annahme der Normalverteilung gerechtfertigt?
Achtung! Schiefe Verteilungen (Lognormal oder Weibull) produzieren extrem weit liegende Werte!
Die Anwendung von Tests auf Basis Normalverteilung führt dann in die Irre!
- Haben Sie für die Ausreißerelimination eine Berechtigung?

Ausreißerbehandlung ist ein diffiziles Problem und sollte auch mittels statistischer Tests unter Berücksichtigung deren Eigenschaften durchgeführt werden.

Beachten Sie, dass die mehrfache Anwendung statistischer Tests die Irrtumswahrscheinlichkeit erhöht!

1.2 univariate Methoden

1.2.1 Standardisierte Sigmagrenzen

Ein praktikabler Einstieg in die Ausreißeranalyse ist die Darstellung der Daten in standardisierter Form bezogen auf die Standardabweichung.

$$z_i = \frac{x_i - \bar{x}}{s}$$

Das Diagramm zeigt die absteigend sortierten Daten mit den Sigmagrenzen $\sigma = 3$ und $\sigma = 4$.

Die Lage von möglichen Ausreißern an den Enden wird deutlich.

Diese möglichen Ausreißer können jetzt für die weitere Berechnung entfernt werden und mit den neu berechneten Schätzwerten der Abstand zu den Sigmalinien verglichen werden.

Durch Umschalten kann die Wirkung der Ausreißer beobachtet werden. Die Daten bleiben im Datensatz erhalten.

Achtung: Die Bewertung zeitabhängiger Daten ist damit eingeschränkt möglich!

1.2.2 Boxplot

Box-Whisker Plots beruhen auf Quantilen und sind nicht von einer Verteilung abhängig.

Die Box erstreckt sich vom unteren Quartil bis zum oberen Quartil (die Werte seien aufsteigend sortiert) und umfasst somit 50% der Daten. Der Abstand vom oberen bis zum unteren Quartil wird als Interquartilsabstand IQR bezeichnet.

Die Box wird vom Median geteilt, ein Kreuzsymbol kann für den arithmetischen Mittelwert verwendet werden.

Die Whiskers werden von den Rändern der Box als Linien dargestellt und repräsentieren Minimum und Maximum der Datenreihe (in EXCEL Fehlerindikatoren formatieren).

Bezüglich der Darstellung von Ausreißern werden folgende Regeln angewendet:

- Daten außerhalb des 1,5 fachen IQR werden als Ausreißer (*outside points*) angezeigt.
 - Daten außerhalb des 3 fachen IQR werden als ferne Ausreißer (*far outside points*) angezeigt.
 - Treten solche Werte auf, dann stellen die Whiskers diejenigen Werte dar, die kleiner sind als die genannten Ausreißer.
- Professionelle Boxplots können noch die Vertrauensbereiche der Mediane als sogenannte Kerben (*median notches*) zeigen.

Mit der Breite der Box ist ein Bezug zum jeweiligen Stichprobenumfang n hergestellt.

Bemerkung zum vorliegenden Diagramm:

Die EXCEL Lösung zeigt tabellarisch die Anzahl der Ausreißer jenseits der genannten Grenzen auf und stellt grafisch nur den jeweils größten Wert von ggf. mehreren Werten dar.

1.2.3 Grubbs Test

Der Grubbs Test kann für normalverteilte Daten verwendet werden (ab $n > 8$). Ein einzelner Ausreißer kann entfernt und der Test weitergeführt werden.

Achtung: Die Mehrfachanwendung von Tests führt generell zur Erhöhung der Irrtumswahrscheinlichkeit.

$$\hat{G} = \frac{\max(|x_i - \bar{x}|)}{s}$$

Ein kritischer Wert kann wie folgt berechnet werden.

$$\hat{G} > G_{n,\alpha} = \frac{(n-1)}{\sqrt{n}} \sqrt{\frac{t_{n-2,\alpha/2n}^2}{n-2+t_{n-2,\alpha/2n}^2}}$$

Damit entfällt das Nachschlagen in den früher üblichen Tabellen.

Hinweis EXCEL: Die Anwendung von Matrixfunktionen lässt eine elegante Lösung ohne Zwischenwerte zu.

1.2.4 Dixon-Test

Bei normal verteilten Werten kann für Stichprobenumfänge $n \leq 30$ der Dixon-Test angewendet werden.

In der Fassung nach DIN 30104 ist die Bezeichnung der Prüfgröße unterschiedlichen n zugeordnet.

Die entsprechenden Formeln sind im Arbeitsblatt dargestellt.

1.2.5 MAD-Test

Für größere ($n > 30$) Stichprobenumfänge kann verwendet werden.

Nach [4] kann ein robustes Ausreißerkriterium aus dem Median und der Medianabweichung gewonnen werden.

Der MAD ist der Median der Abweichungen der Einzelwerte vom Median.

$$|\tilde{x} - x_a| > 5,2MAD$$

Nach [6] wird M_i wie folgt ermittelt.

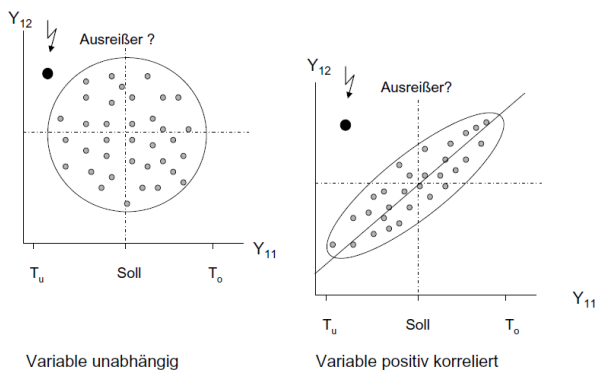
$$M_i = \frac{0,6745(x_i - \tilde{x})}{MAD}$$

Überschreitet $|M_i|$ den Wert von 3,5 kann dieser Wert als Ausreißer betrachtet werden.

Hinweis EXCEL: Die Anwendung von Matrixfunktionen lässt eine elegante Lösung ohne Zwischenwerte zu.

1.3 Multivariate Ausreißertests

Die univariaten Ausreißertests sind im multivariaten Fall nicht anwendbar, weil die Abhängigkeitsstruktur damit nicht erfasst wird.



Ein Hinweis auf Ausreißer kann aus den Werten der T^2 Statistik nach Hotelling abgeleitet werden. Allerdings geht in diese Berechnung auch der ausreißerverdächtige Wert mit ein. Dieses Dilemma weisen auch die parametrischen univariaten Ausreißertests wie z.B. Grubbs auf. Nach Jahn/Braun [5] kann die Teststatistik "Aerk" ohne dieses Dilemma nach folgender Beziehung berechnet werden.

$$Aerk = Y_i X_i^T (X_{N-i}^T X_{N-i})^{-1} X_i Y_i, \forall i = 1, \dots, N$$

Dieses Matrizenprodukt wird für alle Datensätze berechnet.
 Der Vektor $Y_i X_i^T$ ist die i -te Datensatz bzw. Messwert für die Variable Y_i .
 Die Matrix $X_{N-i}^T X_{N-i}$ ist die Matrix des "restlichen" Datensatzes X_{N-i}^T der $N-i$ Transponierten.
 Der Ausdruck $(X_{N-i}^T X_{N-i})^{-1}$ ist die inverse Matrix des Matrizenproduktes der "restlichen" $N-i$ Datenmatrix mit ihrer Transponierten.
 Die Umsetzung in Excel gelingt mit der Partitionierung der Datenmatrix und den in EXCEL verfügbaren Matrixfunktionen.
www.tqu-verlag.de
 Die Häufigkeitsverteilung der Aerk-Werte kann sehr gut durch eine logarithmische Normalverteilung approximiert werden.

Diese Eigenschaft kann für ein Entscheidungskriterium ausgenutzt werden.
 Die Aerk-Werte werden logarithmiert und davon Mittelwert und Standardabweichung berechnet.
 Mit diesem Mittelwert und Standardabweichung werden die transformierten Aerk-Werte normiert (t -Verteilung).
 Die Wahrscheinlichkeit für den maximalen t -Wert kann über die t -Verteilung mit $N - 1$ Freiheitsgraden berechnet werden.

Die drei Beispieldatensätze sind so gewählt, dass der Einfluss der Abhängigkeitsstruktur deutlich wird.
 Mit höherer Abhängigkeit (höherer Grad der Multikollinearität bzw. kleinerer Determinante von R) werden die Ausreißer immer klarer erkennbar.

2. Anwendung

Im Blatt Merkmal sind die zugehörigen Angaben einzutragen.

Die Auswertung ist für 6 Variable ausgelegt.

Für die Auswertung können drei vorbereitete Beispieldatensätze per Makro übernommen werden.

Für die Variablen sind Excelnamen vergeben. Der Namensbereich für die Variablen ist auf $n = 80$ ausgelegt.

Sollen aus Anwendungen Daten übernommen werden mit $n < 80$, so sind die Namen der Variablen dem Umfang der Bereiche anzupassen.

Achtung: Multivariate Datensätze müssen vollständig sein, d.h. fehlende Daten oder ungleiche n , sind nicht erlaubt!

Die folgenden grafischen Analysen und Tests stützen sich auf den gewählten Datensatz. Bei den univariaten Methoden können die interessierenden Variablen ausgewählt werden.

Für den DIXON Test ist die Anzahl der Daten auf < 30 begrenzt!

3. EXCEL

Die EXCEL-Lösung stützt sich auf folgende Prinzipien:

Funktionelle Aufteilung auf verschiedene Blätter,
Optische Hervorhebung von Zellen in Abhängigkeit von ihrer Funktion,
Kommentierung wesentlicher Zellen,

Vergaben von Namen für Variable,

Die Liste muss nach Änderungen mit F3 manuell aktualisiert werden.

Einsatz sog. dynamischer Namen, deren Umfang sich den Inhalten automatisch anpasst,

Textliche Kommentierungen in rot kursiv,

Auslesen von Tabelleninhalten,

Bezug auf benannte Zellbereiche (ist ein sog. volatile Funktion in EXCEL),

Feststellen von Zellinhalten,

Reagieren auf Bedingungen /Verzweigungen,

Erzwingung einer geeigneten Zahlenformatierung im Protokoll mit vorgegebener Stellenzahl,

Verknüpfung von Zellinhalten über "&".

Zellen werden in Berechnungsblättern und Protokollen nicht über "Zellen verbinden" formatiert!

Als Lösung dient die Formatierung schmaler Spalten und die Formatierung benachbarter Zellen mit "Über Auswahl zentrieren".

Die Nachteile verbundener Zellen sind damit vermieden.

Ausblenden von Zellinhalten, die nicht zutreffend sind.

Ausblenden von Blättern, die nicht ständig gebraucht werden.

Blattschutz gegen versehentliches Überschreiben, bitte nutzerintern anpassen.

Matrixfunktionen werden als "nicht-matrixfunktionen" angeboten, als beim Eingeben die Funktion mit "strg", "shift" und "eintr" gleichzeitig (gleichzeitig arbeiten) abgeschlossen wird.
Die "Matrixfunktionen" können in diesem Prozess für von EXCEL automatisch eingefügt, sind also nicht mit der Tastatur erzeugbar!

Matrixfunktionen arbeiten gewissermaßen als "Schleife" und ersparen so häufig eine sonst notwendige VBA Prozedur!

(Anmerkung: Die Formatierung der Protokolle ist weitestgehend standardisiert.)

In den Grafiken werden z.B. Grenzlinien nur mit einem Datenpunkt erzeugt.

Die Linie wird dann mit der Formatierung als "Fehlerindikator" erzeugt.

Dieser Fehlerindikator ist sehr flexibel einsetzbar und erweitert die Möglichkeiten grafischer Gestaltung bedeutend.

Bereitstellung von Testdaten zur Überprüfung der Funktion der Datei,

Protokoll bitte in geeigneten Abständen aktualisieren.

4. Quellen

[1] DIN 53 408-1:2002-4 Statistische Auswertungen messbare (kontinuierliche) Merkmale

[2] DIN ISO 5725-1:1994 Präzision von Messverfahren

[3] Müller et al.: Lexikon der Stochastik, Berlin, 1985

[4] Sachs, L.: Angewandte Statistik, 12. Auflage, Springer, Berlin, Heidelberg, New York 2006

[5] Jahn, W. Braun, L.: Praxisleitfaden Qualität - Prozessoptimierung mit multivariater Statistik in 150 Beispielen. Carl Hanser Verlag, München 2006

[6] STATGRAPHICS XVI statpoint inc.

[7] Jahn, W. Reuter, K.: Multivariate Statistik im Qualitätsmanagement, Workbook TQU Verlag 2012

TQU Verlag, Magirus-Deutz-Straße 18, 89077 Ulm Deutschland, Telefon 0731/14660200, verlag@tqu-group.com, www.tqu-verlag.com

Quality APP Ausreißerbehandlung

Kunde	Automobil OEM
Artikel	Modell_A
Nummer	20
Bezeichnung	Säule_B

	1	2	3	4	5	6	7	8	9
	Variable	Einheit	Nummer	Messmittel	Auflösung	Einheit	Mittelwert	gestutzter Mittelwert	Standardabweichung
1	Y_	Maß1	20-01	Messmaschine MM	0,001	mm	12,3773371	12,33096	2,98279
2	x_1	Maß2	20-02				7,859246657	7,88681	1,45873
3	x_2	Maß3	20-03				8,859970356	8,85422	2,16057
4	x_3	Maß4	20-04				0,592588307	0,59230	0,15087
5	x_4								
6	x_5								
m	4						Stutzungsgrad	5%	

Mittelwert	μ	0,3292
Standardabweichung	σ	0,0078
Wert aus X	x_i	0,42
	k	11,6710
Wahrscheinlichkeit innerhalb	P	91,26%
Wahrscheinlichkeit außerhalb	$1-P$	8,734%

Werte einsetzen und die Wahrscheinlichkeit berechnen

oder

k wählen und den Bereich berechnen

k	2,00	Wahrscheinlichkeit
		P
		75,00%
		$1-P$
		25,00%

oder

P wählen und k berechnen

P	k
95%	4,47

Ohne Voraussetzung einer Verteilung gilt der Satz von Tschebyscheff:

Falls eine Zufallsvariable X den Mittelwert μ und die Standardabweichung σ hat, dann ist die Wahrscheinlichkeit P , dass die Zufallsvariable X Werte x_i in einem symmetrisch zum Mittelwert μ reichhaltigen Bereich

$$P(\mu - k\sigma \leq X \leq \mu + k\sigma) = P(|X - \mu| < k \cdot \sigma) > 1 - \frac{1}{k^2} \quad k > 0$$

k	Variable	Y _i	Mittelwert alle	12,37734	○
1	22,1	3,53	Mittelwert ohne 1	12,25427	●
2	18,407792	2,21	Standardabweichung alle	2,98279	
3	17,6266915	1,93	Standardabweichung ohne 1	2,78995	
4	17,4230999	1,85	Werte entfernen oben	1	
5	17,1061323	1,74	Werte entfernen unten		
6	17,0035677	1,70	Ergebnisse		
7	16,8888964	1,66	Maximum	22,1000	EXCEL Zeile 57
8	16,5904555	1,55	Minimum	5,32884	EXCEL Zeile 5
9	16,3378647	1,46			
10	16,1884378	1,41			
11	16,1484776	1,40			
12	15,8327176	1,28			
13	15,8192895	1,28			
14	15,1958063	1,05			
15	14,8550779	0,93			
16	14,767219	0,90			
17	14,6146558	0,85			
18	14,5684248	0,83			
19	14,5381603	0,82			
20	14,3844723	0,76			
21	14,1638894	0,68			
22	14,0283437	0,64			
23	13,7564768	0,54			
24	13,3907303	0,41			
25	13,3529728	0,39			
26	13,3160783	0,38			
27	13,2902584	0,37			
28	13,2804461	0,37			
29	13,1605287	0,32			
30	13,1056959	0,31			
31	13,0653158	0,29			
32	13,0256234	0,28			
33	12,995637	0,27			
34	12,9904034	0,26			
35	12,8615765	0,22			
36	12,7574879	0,18			
37	12,6182219	0,13			
38	12,4952477	0,09			
39	12,402178	0,05			
40	12,2998519	0,02			
41	12,1443289	-0,04			
42	12,1344815	-0,04			
43	12,0580615	-0,07			
44	11,8754771	-0,14			
45	11,7700521	-0,17			

2

n_i	80
-------	----

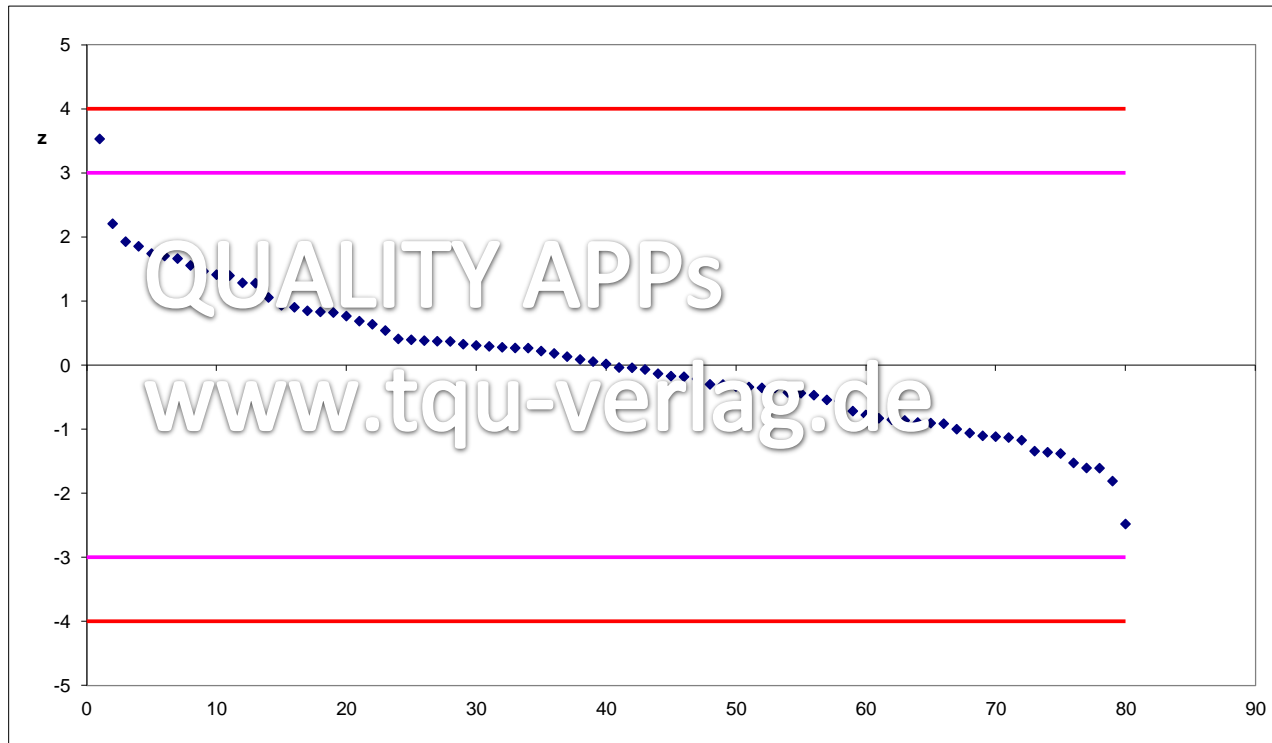
Schätzung für Mittelwert und Standardabweichung ohne vermutete Ausreißer

Standardisierte Daten z_i mit Sigmagrenzen

Aus der Berechnung entfernte Werte sind im Diagramm weiterhin sichtbar!

Ergebnisse		z alle	z ohne
Maximum	22,1000	EXCEL Zeile 57	3,26
Minimum	5,32884	EXCEL Zeile 5	2,36

Die Entscheidung zur Entfernung aus dem Datensatz (mit Protokoll) wird nur von berechtigten Personen getroffen!



Quality APP Ausreißerbehandlung

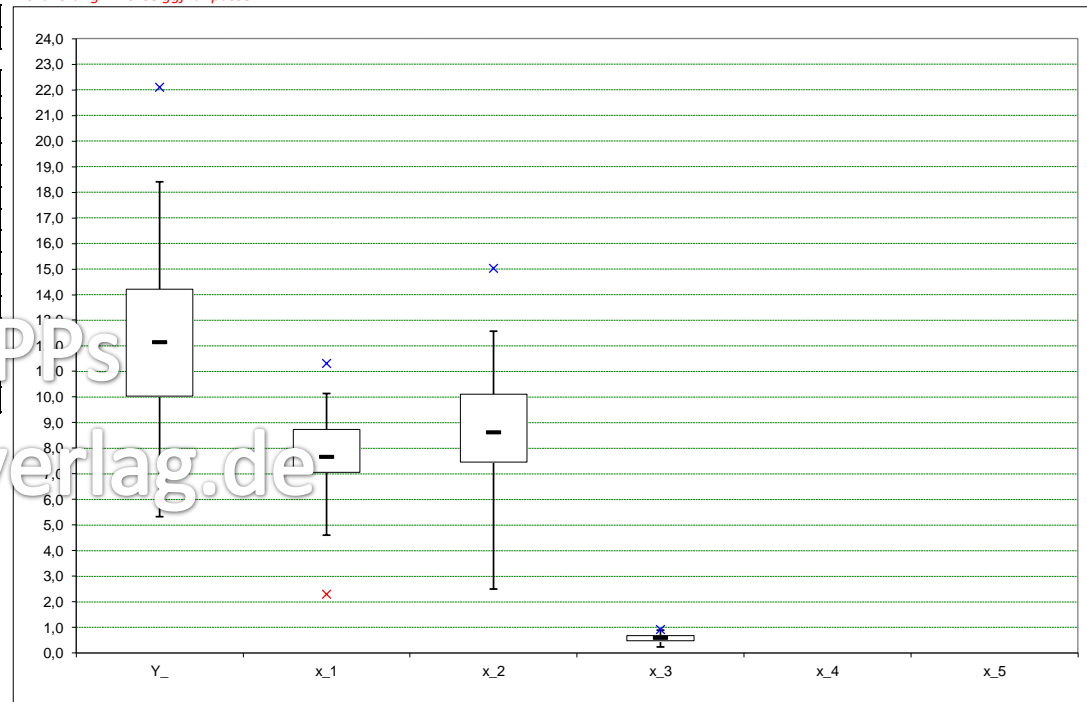
TQU Verlag, Magirus-Deutz-Straße 18, 89077 Ulm Deutschland, Telefon 0731/14660200, verlag@tqu-group.com, www.tqu-verlag.com

Variable	Y_	x_1	x_2	x_3	x_4	x_5
IQ-Abstand	4,17589447	1,65915782	2,64312123	0,20883665		

max ₃	#NV	#NV	#NV	#NV	#NV	#NV
> 3 IQR						
max _{1,5}	22,1	11,31360256	15,02643345	0,921458794	#NV	#NV
> 1,5 IQR	1	5	5	3	#NV	#NV
max	18,40779205	10,13939208	12,5730834	0,898733132	#NV	#NV
+ Whisker	4,189	1,408	2,469	0,203	#NV	#NV
3.Quartil	14,219035	8,731447	10,104086	0,696135	#NV	#NV
Median	12,144329	7,664796	8,623545	0,597665	#NV	#NV
1.Quartil	10,043141	7,072289	7,460965	0,487298	#NV	#NV
- Whisker	4,714	2,465	4,958	0,246	#NV	#NV
min	5,328837177	4,607201352	2,503462266	0,241159063	#NV	#NV
< 1,5 IQR						
min _{1,5}	#NV	#NV	#NV	#NV	#NV	#NV
< 3 IQR		1				
min ₃	#NV	2,3	#NV	#NV	#NV	#NV

In den Bereichen > 1,5 und > 3 IQR wird nur der jeweilige Wert grafisch angezeigt!

Skalierung Y-Achse ggf. anpassen



Quality APP Ausreißerbehandlung

TQU Verlag, Magirus-Deutz-Straße 18, 89077 Ulm Deutschland, Telefon 0731/14660200, verlag@tqu-group.com, www.tqu-verlag.com

Grubbs Test (alle Variablen gleiches n)

Variable	mean	stdev	max	min	max	min	\hat{G}	\hat{G}	Signifikanz	Entscheidung	Zeilennummer
Y_	12,37734	2,982789	22,1000	5,3288	9,7227	7,0485	3,2596	3,2596	FALSCH	nicht signifikant	
x_1	7,859247	1,458726	11,3136	2,3000	3,4544	5,5592	3,8110	3,8110	WAHR	signifikant	73
x_2	8,85997	2,160569	15,0264	2,5035	6,1665	6,3565	2,9421	2,9421	FALSCH	nicht signifikant	
x_3	0,592588	0,150869	0,9215	0,2412	0,3289	0,3514	2,3294	2,3294	FALSCH	nicht signifikant	
x_4			0,0000	0,0000							
x_5			0,0000	0,0000							

n =	alpha	t ²	G _{nα}
80	5,0%	14,236325	3,470

$$\hat{G} = \frac{\max(|x_i - \bar{x}|)}{s}$$

$$\hat{G} > G_{n,\alpha} = \frac{(n-1)}{\sqrt{n}} \sqrt{\frac{t_{n-2,\alpha/2n}^2}{n-2 + t_{n-2,\alpha/2n}^2}}$$

Quality APP Ausreißerbehandlung

TQU Verlag, Magirus-Deutz-Straße 18, 89077 Ulm Deutschland, Telefon 0731/14660200, verlag@tqu-group.com, www.tqu-verlag.com

DIN 53 804 T1

Variable	Y_	Werte aus aufsteigend geordneter Folge				
n =	80	Prüfen Sie, ob MIN oder MAX als Ausreißer interpretiert werden können.				
<i>sorry, no test</i>			Testreißer	Schweller	Signifikanz	
$\alpha =$	5,0%	2	nach unten	nach oben	nach unten	nach oben
$X_{(n)}$	21,0	1 - 7				
$X_{(n-1)}$	18,41	8 - 10				
$X_{(n-2)}$	17,63	11 - 13				
$X_{(3)}$	7,6	1 - 25	0,16	0,3197	#/V	#/V
$X_{(2)}$	7,20					
$X_{(1)}$	5,33					

Quality APP Ausreißerbehandlung

TQU Verlag, Magirus-Deutz-Straße 18, 89077 Ulm Deutschland, Telefon 0731/14660200, verlag@tqu-group.com, www.tqu-verlag.com

MAD Test

MAD = Median der Medianabweichungen

nach [4]	Kennwerte		Extreme		Differenzen zum Median		Kriterium	5,2	Ausreißer	Datensatz
Variable	Median	MAD	max	min	max	min	max	min	Wert	Zeile EXCEL
Y_	12,2220904	2,152792	22,1	5,328837	9,87791	6,893253	FALSCH	FALSCH		
x_1	7,68059237	0,863067	11,3136	2,3	3,63301	5,380592	FALSCH	FALSCH		
x_2	8,71106241	1,3141532	15,02643	2,503462	6,315371	6,2076	FALSCH	FALSCH		
x_3	0,59172748	0,1056572	0,921459	0,241159	0,329731	0,350568	FALSCH	FALSCH		
x_4			0	0	#ZÄHL!	#ZÄHL!	#ZÄHL!	#ZÄHL!		
x_5			0	0	#ZÄHL!	#ZÄHL!	#ZÄHL!	#ZÄHL!		

Irrtumswahrscheinlichkeiten werden von den Autoren nicht angegeben!

$$M_a = \frac{0,6745 (x_a - \tilde{x})}{MAD}$$

nach [6]	Kennwerte		Extreme		Differenzen zum Median		Kriterium	3,5	Ausreißer	Datensatz
Variable	max	min	max	min	max	min	max	min	Wert	Zeile EXCEL
Y_	6,66265	4,649499	FALSCH	FALSCH						
x_1	2,450465	3,62921	FALSCH	WAHR					2,3	73
x_2	4,259718	4,187026	FALSCH	FALSCH						
x_3	0,222404	0,236458	FALSCH	FALSCH						
x_4			#ZÄHL!	#ZÄHL!						
x_5			#ZÄHL!	#ZÄHL!						

Quality APP Ausreißerbehandlung

TQU Verlag, Magirus-Deutz-Straße 18, 89077 Ulm Deutschland, Telefon 0731/14660200, verlag@tqu-group.com, www.tqu-verlag.com

