

Über Ausreißertests

Bernd Paul Jäger
Ernst-Moritz-Arndt-
Universität Greifswald,
Institut für Biometrie und
Med. Informatik
Walther-Rathenau-Straße 48
17487 Greifswald
bjaeger@biometrie.uni-
greifswald.de

Paul Eberhard Rudolph
Ehemals: Leibnizinstitut für
Nutztierbiologie,
Forschungsbereich Genetik
und Biometrie
Wilhelm-Stahl-Allee 2
18196 Dummerstorf
pe.rudolph@kabelmail.de

Zusammenfassung

Mittels Simulation ‘verschmutzter’ Normalverteilungen mit einem SAS-Programm wird die Power einiger Ausreißertests (Boxplotmethode aus SAS, Maximummethode, MAD-Methode, variierte Boxplotmethode, Peirce-, David-Hartley-Pearson- und Dean/Dixon-Methode) ermittelt. Daraus werden Anwendungsempfehlungen für diese Tests abgeleitet.

Schlüsselwörter: Ausreißertest, Power, Simulation

1 Einleitung

Der Problemfall Ausreißer ist allen Anwendern intuitiv bekannt. Man versteht unter einem Ausreißer in der Statistik meist einen Messwert, der mit den übrigen erhobenen Werten nicht konsistent ist, sei es, dass er wesentlich größer oder kleiner als die übrigen ist.

Man möchte einerseits solche Werte als Messfehler betrachten. Andererseits weiß man, dass es Verteilungen gibt, zu denen besonders große oder kleine, allerdings selten auftretende Werte gehören (sogenannte heavy tailed distributions, Verteilungen mit relativ großen Wahrscheinlichkeitsanteilen am Rande ihres Definitionsbereichs). Natürlich kann sich in der Realität auch der „besondere Fall“ hinter einem solchen ungewöhnlichen Messwert verbergen, von dem man sich ohne Informationsverlust nicht trennen sollte.

Ausreißer sind nicht nur ein Makel für den Experimentator, sie beeinflussen in besonderer Weise auch aus der Stichprobe berechnete Parameter, zum Beispiel die Momente. Und je höher die Momente sind, umso größer ist der Einfluss eines Ausreißers. Beim ersten Moment, dem Mittelwert, gehen die Werte einschließlich des Ausreißers linear ein. Aber schon beim zweiten Moment, der Varianz, gehen diese quadratisch ein.

Die statistischen Entscheidungen werden ebenso von Ausreißern beeinflusst, lediglich die parameterfreien Verfahren sind ihnen gegenüber weitestgehend robust. Bei diesen auf Rangwerten beruhenden Entscheidungsverfahren geht ein Ausreißer genau wie ein

großer oder kleiner Messwert $x_{(n)}$ oder $x_{(1)}$ als Rangwert n oder 1 in die Analyse ein, je nachdem, ob er am oberen oder unteren Ende positioniert ist.

Ein weiteres Problem besteht darin, dass man ohne Kenntnis der unterliegenden Verteilung keine Ausreißertests durchführen kann. Der einzige parameterfreie Ausreißertest, der Test von WALSH (1958, 1973), soll nur erwähnt werden. Er funktioniert allerdings erst bei sehr großen Stichprobenumfängen, die man für die meisten Entscheidungen in der Praxis nicht zur Verfügung hat.

2 Die einzelnen Methoden

Abgesehen vom WALSH-Test beruhen alle besprochenen Ausreißertests allein auf der Normalverteilung. Zahlreiche Methoden sind ersonnen worden, um das Ausreißerproblem in den Griff zu bekommen, von denen die bekanntesten kurz erläutert werden sollen. In einem weiteren Abschnitt werden alle besprochenen Methoden einem Simulationsexperiment unterworfen und geprüft und bewertet, wie sie die Fehler 1. und 2. Art einhalten. Ein Ranking der Testverfahren ist möglich.

Der Einfachheit halber werden wir nur über Grundgesamtheiten mit Standardnormalverteilungen sprechen, bei denen die Elemente der Stichprobe bereits der Größe nach geordnet sind. Ansonsten muss eine Standardnormierung durchgeführt und die transformierte Stichprobe geordnet werden. Dann ist x_1 immer der kleinste und x_n immer der größte Messwert und man muss nicht zwischen Messwert x_i und i -tem Messwert $x_{(i)}$ der sortierten Reihe unterscheiden.

2.1 Boxplot-Methode

Bei der Boxplot-Methode geht die Breite der Box ein, die $Q_3 - Q_1$ Differenz des oberen und unteren Quartils (auch Interquartilabstand genannt). Vom unteren Quartil wird die 1.5-fache Boxbreite subtrahiert, zum oberen Quartil entsprechend addiert. Alle außerhalb dieses Bereichs liegenden Messwerte sind potentielle Ausreißer.

Die Boxbreite $Q_3 - Q_1$ ist ein parameterfreies Streuungsmaß. Die Boxplot-Methode ist damit einer „parameterfreien 2σ -Regel“ vergleichbar, denn wenn der Median mittig zwischen Q_1 und Q_3 liegt sind die Grenzen des Intervalls I gerade zweimal die „Streuung $Q_3 - Q_1$ “ vom Median entfernt. Man bezeichnet alle Punkte außerhalb des Intervalls

$$I = [Q_1 - 1.5 \cdot (Q_3 - Q_1); Q_3 + 1.5 \cdot (Q_3 - Q_1)]$$

als ausreißerverdächtig. In der SAS-Prozedur BOXPLOT werden genau nach dieser Methode die Ausreißerverdächtigen eingezeichnet. Mit der Programmzeile

```
plot x*gruppe/boxstyle = schematic IDSMBOL=circle;
```

beispielsweise werden die verdächtigen Punkte mit einem Kreissymbol gekennzeichnet. Die Methode stammt wahrscheinlich von BORIS IGLEWICZ und SHARMILA BANERJEE (2001).

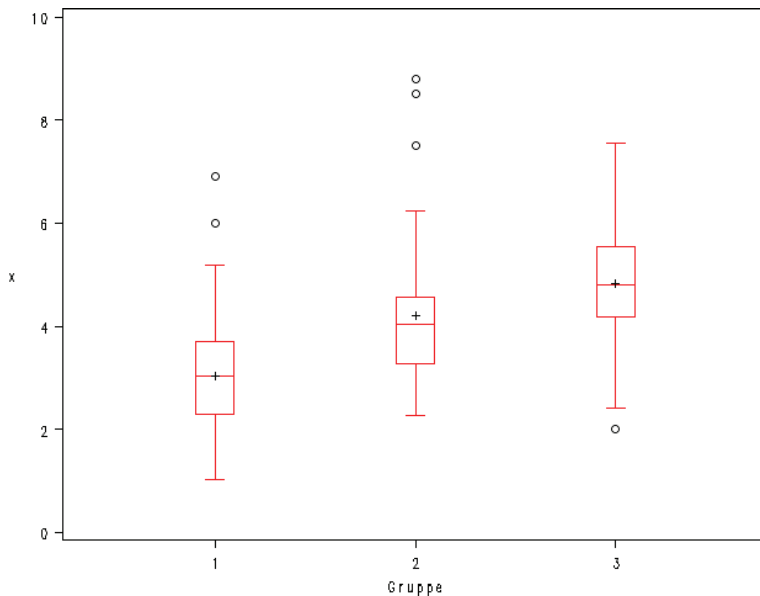


Abbildung 1: Grafikprozedur BOXPLOT markiert ausreißerverdächtige Punkte

FRIGGE, HOAGLIN und IGLEWICZ (1989) behaupten, dass man mit dieser Regelung mit Wahrscheinlichkeit $0.05 \leq p \leq 0.1$ einen oder mehrere Ausreißer in einer Stichprobe findet, eine recht vage Aussage. Vergleicht man diese mit den Simulationsergebnissen, muss man den Anteil von Stichproben mit ausreißerverdächtigen Werten bei Normalverteilungen bereits auf $p > 0.3$ schätzen, bei anderen (nicht so stark am Mittelwert konzentrierten) Verteilungen ist er sicher noch höher. (Vergleiche das Simulationsergebnis beim Umfang $n = 50$.)

Die schlechten Ergebnisse der Boxplotmethode im späteren Simulationsexperiment kann man verbessern, wenn man mit der „variieren Boxplot-Methode“ arbeitet. Dabei werden diejenigen Messwerte als ausreißerverdächtig eingestuft, die nicht im Intervall

$$I = [Q_1 - k(n) \cdot (Q_3 - Q_1); Q_3 + k(n) \cdot (Q_3 - Q_1)]$$

liegen. Dabei ist $k(n)$ eine Konstante, die vom Stichprobenumfang n abhängt, und die durch ein Simulationsexperiment erhalten werden kann. Für Stichprobenumfänge $n \geq 20$ kann diese Konstante mit 2.25 näherungsweise angenommen werden. Das folgende SAS-Programm realisiert die Bestimmung der Konstante und die Tabelle 1 gibt die $k(n)$ für $\alpha = 0.05$ an.

SAS-Programm zur Bestimmung der $k(n)$ -Konstanten der variierten Boxplot-Methode

```

%let nj=90; /* Stichprobenumfang */
%let ni=10000; /* Simulationsumfang */

data johannes;
do i=1 to &ni;
  do j=1 to &nj;
    x=NORMAL(362971);
    output;
  end;
end;run;

proc means data=johannes noprint;
var x;by i;
output out=ergeb_joh min=mi p25=q1 p75=q3 max=ma;run;
data ergeb_joh;
set ergeb_joh;
drop _type_ _freq_;
n=_freq_;run;
proc iml;
use ergeb_joh;
read all var _num_ into x; /* Reihenfolge: i mi q1 q3 ma n */
Do aa=2.2 to 2.3 by 0.005; /* Bereich, in dem k(n) erwartet wird */
  box=J(&ni,2,0); /* 1.Spalte obere, 2. untere Grenze */
  Ausreisser=0;
  do i=1 to &ni;
    box[i,1]=x[i,4]+aa*(x[i,4]-x[i,3]);/* obere Grenze */
    box[i,2]=x[i,3]-aa*(x[i,4]-x[i,3]);/* untere Grenze*/
    if (x[i,5]>box[i,1]) | (x[i,2]<box[i,2]) then
Ausreisser=Ausreisser+1;
/* Wenn kleinster Wert unter unterer Grenze oder größter Wert
über oberer Grenze, dann um weiteren Ausreißer erhöhen */
  end;
  print aa, Ausreisser;
end;run; quit;

```

Tabelle 1: Konstanten $k(n)$ der variierten Boxplot-Methode für verschiedene n ($\alpha = 0.05$)

n	k	n	k	n	k	n	k
5	5.660	12	2.280	19	2.120	50	2.240
6	2.970	13	2.910	20	2.270	60	2.240
7	2.040	14	2.490	25	2.490	70	2.240
8	2.280	15	2.120	30	2.255	80	2.240
9	3.430	16	2.270	35	2.160	90	2.245
10	2.570	17	2.710	40	2.220	100	2.265
11	2.140	18	2.710	45	2.320	200	2.310

2.2 Methode von PEIRCE und CHAUVENET

Das historisch älteste Verfahren zur Ausreißererkenkung stammt von BENJAMIN PEIRCE (1852). Es ist von den Wissenschaftlern allerdings bis in unsere Zeit nicht zur Kenntnis genommen worden, wenn man einmal davon absieht, dass GOULD (1855) für das Verfahren ein Tafelwerk schuf zum leichteren Gebrauch der Ausreißer-erkennungsmethode. Ein Spezialfall von PEIRCE's allgemeiner Herangehensweise, die Methode von WILLIAM CHAUVENET (1863), ist - zumindest bei Ingenieurwissenschaftlern – in Anwendung geblieben.

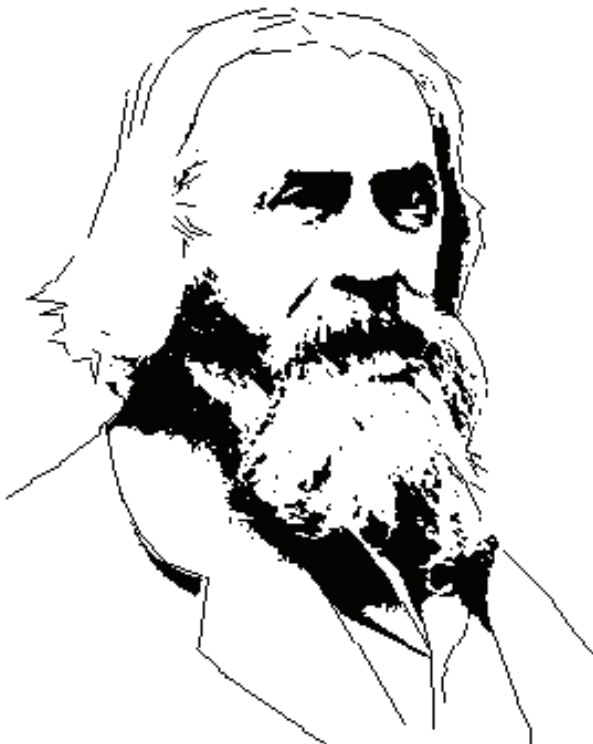


Abbildung 2: Benjamin Peirce
(*1809 in Salem, Massachusetts;
† 1880 in Cambridge, Massachusetts)

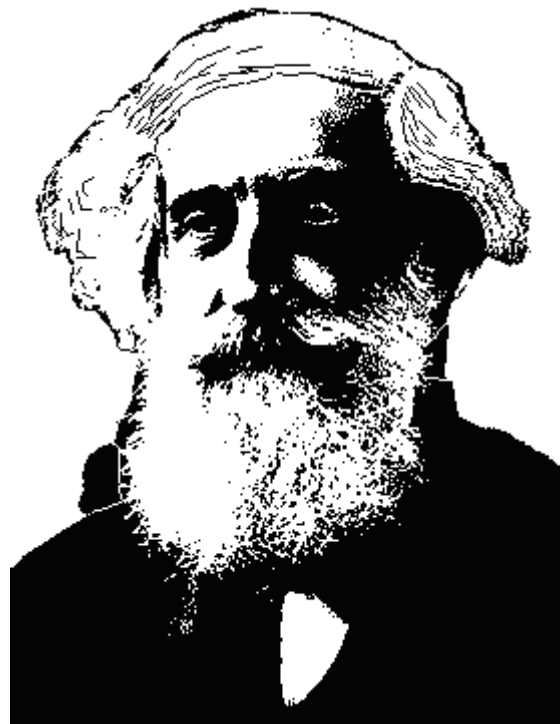


Abbildung 3: William Chauvenet
(* 24. Mai 1820 in Milford, Pennsylvania;
† 13. Dezember 1870 in St. Paul, Minnesota)

Der Unterschied zwischen beiden Methoden ist nur unwesentlich. Es ist das Verdienst von STEPHEN M. ROSS (2003), die Arbeit von PEIRCE wieder bekannt gemacht zu haben, eine späte Würdigung der Leistung von PEIRCE 150 Jahre nach seiner Publikation.

Wenn x_1, x_2, \dots, x_n eine bereits aufsteigend geordnete Stichprobe aus einer normalverteilten Grundgesamtheit ist, dann ist

$$R_1 = \text{MAX} \left(\text{ABS} \left(\frac{x_1 - m}{s} \right), \text{ABS} \left(\frac{x_n - m}{s} \right) \right)$$

die Prüfgröße des Peirce-Tests für ein als Ausreißer verdächtigtes Stichprobenelement. Der kleinste und der größte Wert der Stichprobe werden standardnormiert und man nimmt deren Maximum.

Die Methode von CHAUVENET besteht darin, dass nur auf einer Seite der Verteilung nach Ausreißern gesucht wird. Die Prüfgrößen von CHAUVENET sind folglich:

$$CH_1 = \frac{x_n - m}{s} \text{ bzw. } CH_2 = \frac{m - x_1}{s}$$

für den Test, dass der größte Wert x_n ein Ausreißer ist, bzw., dass der kleinste Wert x_1 ein Ausreißer ist.

Die Verallgemeinerung des PEIRCE- und des CHAUVENET-Tests für mehrere als Ausreißer verdächtige Stichprobenelemente auf einer Seite der Verteilung liegt förmlich auf der Hand, für zwei beispielsweise

$$R_2 = \text{MAX} \left(\text{ABS} \left(\frac{x_2 - m}{s} \right), \text{ABS} \left(\frac{x_{n-1} - m}{s} \right) \right)$$

und

$$CH_{1,2} = \frac{x_{n-1} - m}{s} \text{ bzw. } CH_{2,2} = \frac{m - x_2}{s}.$$

Theoretisch könnte mit dieser Verallgemeinerung bis auf $n/2$ mögliche Ausreißer auf einer Seite der Verteilung getestet werden. Die Tabellen im Artikel von GOULD (1855) hören bei 10 Ausreißern auf. Allerdings tabelliert Gould nicht die kritischen Werte der Prüfgröße, sondern den Erwartungswert! Mit dem SAS-Simulationsprogramm kann man für vorgegebene Umfänge n Verteilung, Quantile und kritische Werte der Prüfgröße bestimmen (siehe Tab. 2 und Abb. 4).

Tabelle 2: Quantile der PEIRCE-Prüfgröße bei 10 000 Simulationen und Vergleich mit den Werten von GOULD

n	Erwartungswerte		Quantile					
	GOULD	Mittelw.	Q ₅₀	Q ₉₀	Q ₉₅	Q _{97.5}	Q ₉₉	Q _{99.5}
5	1.509	1.439	1.441	1.671	1.715	1.743	1.764	1.773
6	1.610	1.547	1.540	1.824	1.888	1.933	1.974	1.994
7	1.693	1.632	1.618	1.938	2.021	2.079	2.138	2.171
8	1.763	1.702	1.683	2.029	2.124	2.200	2.274	2.317
9	1.824	1.762	1.738	2.109	2.213	2.298	2.388	2.439
10	1.878	1.816	1.789	2.176	2.289	2.383	2.484	2.543
11	1.925	1.863	1.835	2.234	2.356	2.455	2.562	2.627
12	1.969	1.905	1.874	2.287	2.412	2.518	2.632	2.708
13	2.007	1.943	1.910	2.332	2.460	2.573	2.700	2.779
14	2.043	1.976	1.941	2.370	2.503	2.623	2.754	2.841
15	2.076	2.008	1.973	2.411	2.548	2.668	2.804	2.899

16	2.106	2.037	2.001	2.444	2.588	2.713	2.853	2.950
17	2.134	2.066	2.029	2.479	2.622	2.749	2.898	2.992
18	2.161	2.089	2.051	2.505	2.649	2.781	2.934	3.039
19	2.185	2.113	2.075	2.533	2.685	2.819	2.971	3.076
20	2.209	2.134	2.094	2.558	2.707	2.846	3.003	3.110
25	2.307	2.228	2.186	2.661	2.822	2.968	3.139	3.254
30	2.385	2.301	2.257	2.743	2.913	3.057	3.242	3.366
35	2.450	2.362	2.316	2.811	2.977	3.133	3.315	3.448
40	2.504	2.412	2.367	2.865	3.036	3.187	3.373	3.510
45	2.551	2.457	2.412	2.914	3.086	3.246	3.434	3.563
50	2.592	2.496	2.450	2.957	3.131	3.293	3.481	3.615
60	2.663	2.561	2.515	3.022	3.194	3.356	3.552	3.693

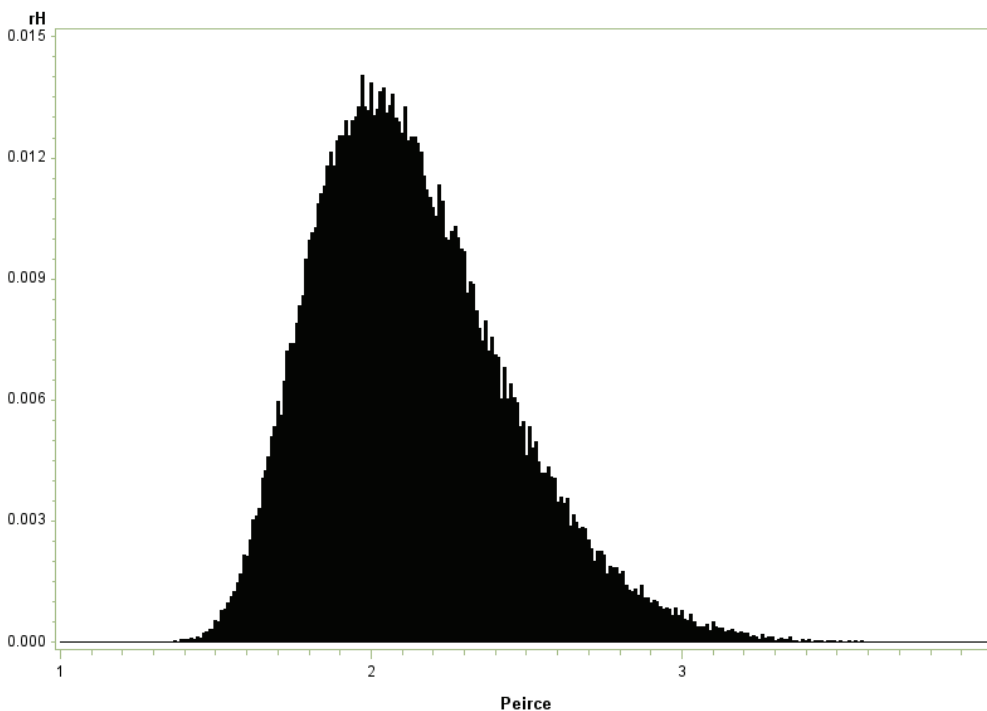


Abbildung 4: Häufigkeitsfunktion der Prüfgröße von PEIRCE für $n = 20$ bei 100 000 Simulationen für den Fall eines ausreißerverdächtigen Punktes

2.3 Die Maximum-Methode

Bei der Maximum-Methode geht man von der Verteilung der Extremwerte einer Stichprobe aus, die bei bekannter Verteilung F bestimmbar sind. Bekanntlich sind das für das Maximum $\text{MAX}(X_1, X_2, \dots, X_n)$ einer Stichprobe die Verteilung

$$F(\text{MAX}(X_1, X_2, \dots, X_n) = x) = F^n(x) = \text{CDF}(\text{'NORMAL'}, x)^n$$

und für das Minimum $\text{MIN}(X_1, X_2, \dots, X_n)$ die Verteilung

$$F(\text{MIN}(X_1, X_2, \dots, X_n) = x) = 1 - (1 - F(x))^n = 1 - (1 - \text{CDF}('NORMAL', x))^n,$$

wobei diese Verteilungen durch die CDF-Funktion aus SAS leicht bestimmbar sind. Man kann sich auf die Maximumverteilung beschränken, weil durch Multiplikation der Stichprobenelemente mit (-1) das Minimum- zu einem Maximumproblem gewandelt werden kann. Alle Werte oberhalb des $(1-\alpha/2)$ -Quantils von $F^n(x)$ und unterhalb des $(\alpha/2)$ -Quantils von $1 - (1 - F(x))^n$ werden als mögliche Ausreißer deklariert. Die Quantile des Minimums unterscheiden sich von den Werten in Tab. 3 nur durch das Vorzeichen.

Tabelle 3: Quantile der Maximummethode für Stichprobenumfänge n von 1 bis 50 für $\alpha=0.05$

n	MAX	n	MAX	n	MAX	n	MAX	n	MAX
1	1.95996	11	2.83393	21	3.03445	31	3.14999	41	3.23076
2	2.23896	12	2.86159	22	3.04844	32	3.15925	42	3.23764
3	2.39089	13	2.88685	23	3.06177	33	3.16821	43	3.24434
4	2.49435	14	2.91006	24	3.07448	34	3.17687	44	3.25088
5	2.57233	15	2.93154	25	3.08663	35	3.18527	45	3.25726
6	2.63469	16	2.95151	26	3.09826	36	3.19341	46	3.26349
7	2.68650	17	2.97017	27	3.10943	37	3.20131	47	3.26958
8	2.73073	18	2.98767	28	3.12015	38	3.20898	48	3.27553
9	2.76927	19	3.00415	29	3.13047	39	3.21644	49	3.28135
10	2.80337	20	3.01971	30	3.14041	40	3.22369	50	3.28704

SAS-Programm zur Berechnung der Quantile des Maximumtests für n = 1 bis 50

```
data;
do n=1 to 50;
  x=Probit(0.975** (1/n));
  output;
end; run;
proc print;run;
```

2.4 Die modifizierten Z-Scores

Im Allgemeinen bezeichnet man die standardnormierten Werte z_i , die durch Transformation aus beliebigen normalverteilten Werten x_i hervorgehen, als Z-Scores: $z_i = (x_i - \bar{x})/s$, wobei s die Standardabweichung und \bar{x} der Mittelwert sind. Die modifizierten Z-Scores sind das parameterfreie Pendant, bei dem man den Mittelwert durch den empirischen Median $p_{50} = p_{50}(x_i)$ und die Streuung durch den Median der absoluten Abweichung (**M**edian of **A**bsolute **D**eviation **MAD**) ersetzt. Der MAD-Wert ist ein robustes Maß der Variabilität der Verteilung, der gegenüber der

Standardabweichung weniger von Ausreißern betroffen wird. Ausgehend von den Abweichungen der Messwerte vom Median der Stichprobe ($x_i - p_{50}(x_i)$) ist

$$\text{MAD} = p_{50} \left(\text{ABS}(x_i - p_{50}(x_i)) \right)$$

und das Äquivalent zum Z-Score

$$\text{Mz}_i = 0.6745 \cdot \left(\frac{x_i - p_{50}}{\text{MAD}} \right).$$

Alle Messwerte x_i mit $\text{Mz}_i > 3.5$ gelten unter Normalverteilungsvoraussetzung nach IGLEWICZ und HOAGLIN als potentielle Ausreißer.

Beispiel:

Betrachtet man die Stichprobe (0.5, 1, 2, **2**, 4, 5, 9) vom Umfang $n = 7$, bei der Median mit dem viertgrößten Wert von 2 zusammen fällt (Wert fett markiert). Die absoluten Abweichungen von 2 sind (1.5, 1, 0, 0, 2, 4, 7). Sie haben den Median von 1.5 = MAD. Die Mz_i sind (0.6745, 0.44967, 0, 0, 0.8993, 1.7987). Kein Mz_i -Wert überschreitet die 3.5, kein Wert ist ausreißerverdächtig.

2.5 Der Test von DEAN und DIXON

Voraussetzungen für den Ausreißertest von DEAN und DIXON sind Daten aus normalverteilten Grundgesamtheiten, die bereits aufsteigend geordnet sind. Der Test wurde von R.B. DEAN und J.W. DIXON 1951 entwickelt und 1953 von DIXON vereinfacht. Durch das Ordnen testet man, ob der kleinste Messwert x_1 ein Ausreißer ist. Man testet:

H_0 : Die Stichprobe enthält keinen Ausreißer versus H_1 : x_1 ist ein Ausreißer.

Als Testgröße Q haben DEAN und DIXON verschiedene Varianten von Q für die verschiedenen Stichprobenumfänge n ersonnen:

$$Q = r_{10} = \frac{x_2 - x_1}{x_n - x_1} \quad \text{für } 3 \leq n \leq 7,$$

$$r_{11} = \frac{x_2 - x_1}{x_{n-1} - x_1} \quad \text{für } 8 \leq n \leq 10,$$

$$r_{21} = \frac{x_3 - x_1}{x_{n-1} - x_1} \quad \text{für } 11 \leq n \leq 13 \text{ und}$$

$$r_{22} = \frac{x_3 - x_1}{x_{n-2} - x_1} \quad \text{für } n \geq 14.$$

Durch Vergleich des Wertes der Prüfgröße mit dem kritischen Wert der folgenden Tabelle 4 wird zwischen beiden Hypothesen entschieden. Falls der Wert von $Q = r_{10}$ (entsprechend für r_{11} , r_{21} und r_{22}) größer als der kritische Wert ist, lehnt man H_0 ab, und x_1 wird als Ausreißer deklariert. Diese statistische Entscheidung ist einseitig.

Ein SAS-Programm zur Simulation der Prüfgrößen kann aus dem SAS-Programm des Anhangs abgeleitet werden. Für jeden Stichprobenumfang n wird die Prüfgröße des DEAN-DIXON-Tests und ihre Quantile näherungsweise bestimmt (siehe Abb. 5 u. 6 sowie Tab. 4).

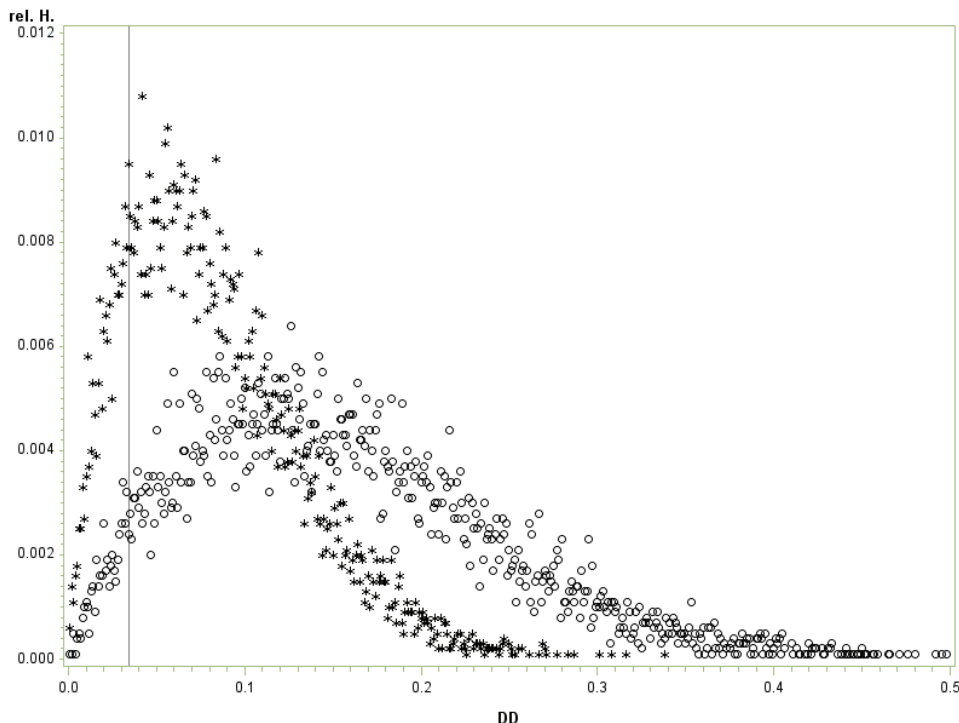


Abbildung 5: Relative Häufigkeitsverteilungen der Prüfgröße von DEAN und DIXON für Stichprobenumfang $n = 50$ bei 100 000 Simulationen unter $H_0: X \sim N(0,1)$ bzw. $H_1: Y \sim N(0,1)$ mit 2% Verschmutzung mit $N(5,1)$, wobei Kreis Symbol für X bzw. Stern für Y . Als Referenzlinie eingezeichnet ist der kritische Wert 0.03454 des Tests für $\alpha = 0.05$.

Tabelle 4: Kritische Werte des Ausreißertests von DEAN und DIXON (simul. Werte fett)

	N	Q99	Sim Q99	Q95	simQ95	Sim Q1	SimQ5
$Q = r_{10} = \frac{x_2 - x_1}{x_n - x_1}$	3	0.988	0.988	0.941	0.941	0.01202	0.05935
	4	0.889	0.889	0.766	0.766	0.00673	0.03259
	5	0.782	0.778	0.643	0.644	0.00488	0.02361
	6	0.698	0.702	0.563	0.564	0.00387	0.01978
	7	0.636	0.643	0.507	0.509	0.00342	0.01690
$r_{11} = \frac{x_2 - x_1}{x_{n-1} - x_1}$	8	0.682	0.681	0.554	0.555	0.00395	0.01967
	9	0.634	0.631	0.512	0.509	0.00338	0.01733
	10	0.597	0.599	0.477	0.479	0.00312	0.01555
$r_{21} = \frac{x_3 - x_1}{x_{n-1} - x_1}$	11	0.674	0.674	0.575	0.571	0.03677	0.08359
	12	0.643	0.645	0.546	0.546	0.03354	0.07763
	13	0.617	0.617	0.522	0.521	0.03129	0.07251

$r_{22} = \frac{x_3 - x_1}{x_{n-2} - x_1}$	14	0.640	0.640	0.546	0.547	0.03473	0.07860
	15	0.617	0.616	0.524	0.523	0.03265	0.07352
	16	0.598	0.598	0.505	0.504	0.03016	0.06920
	17	0.580	0.579	0.489	0.488	0.02905	0.06609
	18	0.564	0.565	0.475	0.474	0.02755	0.06359
	19	0.551	0.550	0.462	0.461	0.02668	0.06095
	20	0.538	0.538	0.450	0.450	0.025	0.059
	25		0.487		0.405	0.02184	0.05026
	30		0.456		0.375	0.01920	0.04519
	35		0.432		0.353	0.01778	0.04199
	40		0.413		0.338	0.01682	0.03873
	45		0.398		0.323	0.01576	0.03635
	50		0.383		0.311	0.01473	0.03454
	100		0.253		0.318	0.01155	0.02647

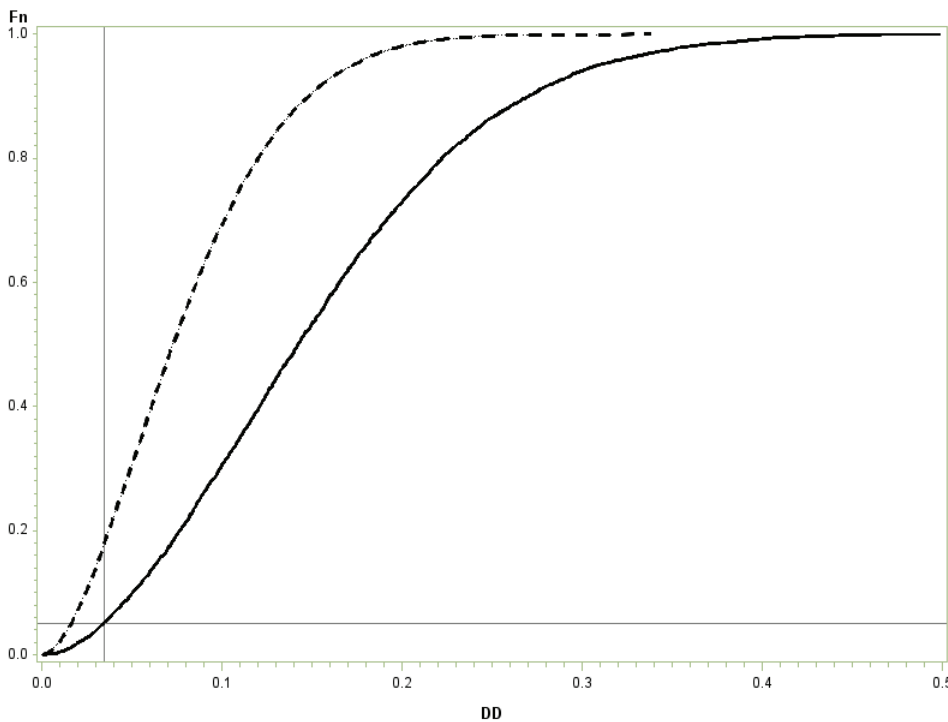


Abbildung 6: Empirische Verteilungsfunktionen der Prüfgröße von DEAN und DIXON für Stichprobenumfang $n = 50$ bei 100 000 Simulationen unter $H_0: X \sim N(0,1)$ bzw. $H_1: Y \sim N(0,1)$ mit 2% Verschmutzung mit $N(5,1)$, wobei Kreis Symbol für X bzw. Stern für Y. Als Referenzlinien sind der kritische Wert des Tests und $\alpha = 0.05$ eingezeichnet.

2.6 Der DAVID-HARTLEY-PEARSON-Test

Die Nullhypothese H_0 , der kleinste oder der größte Wert einer Datenreihe gehört zur Stichprobe, wird zum Niveau α verworfen, wenn gilt:

$$Q = \frac{R}{s} = \frac{x_{(n)} - x_{(1)}}{s} > Q_{n,1-\alpha}$$

R ist die Spannweite und s die Standardabweichung, $Q_{n,1-\alpha}$ steht für die Quantile des DAVID-HARTLEY-PEARSON-Tests. Wird die Nullhypothese verworfen, wird der kleinste bzw. größte Wert als Ausreißer betrachtet, je nachdem, welcher am weitesten vom Mittelwert entfernt liegt.

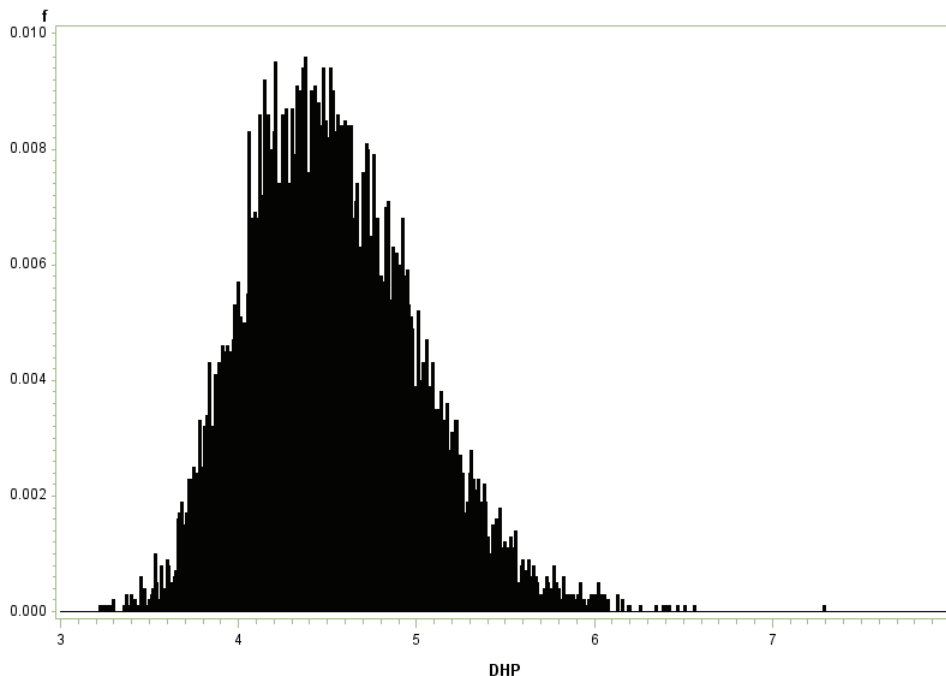


Abbildung 7: Häufigkeitsverteilung der Prüfgröße des Tests von DAVID-HARTLEY-PEARSON bei 10 000 Simulationsläufen für Stichprobenumfang $n = 50$

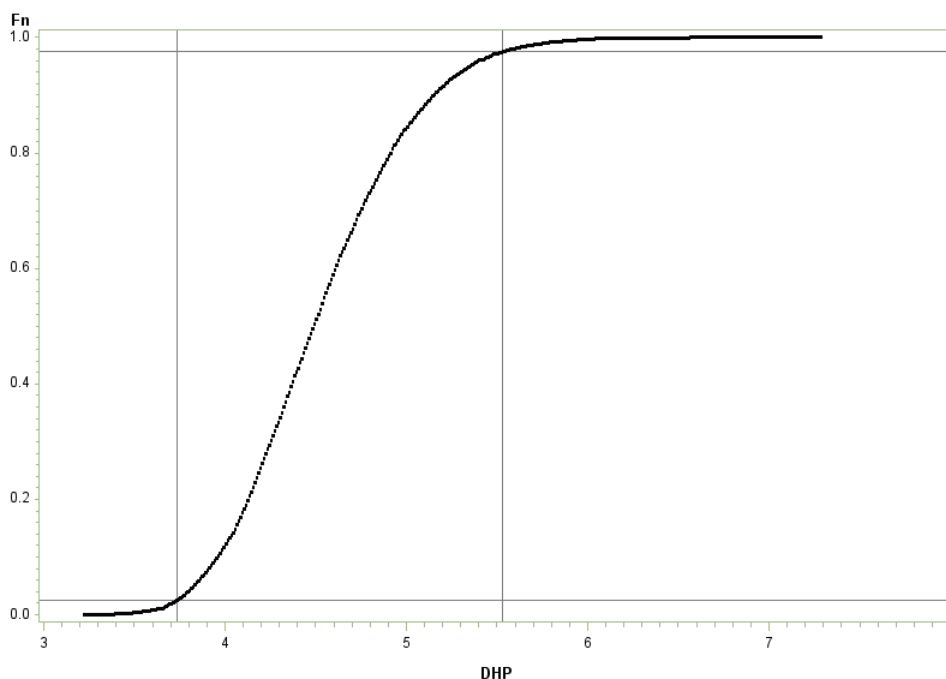


Abbildung 8: Empirische Verteilungsfunktion der Prüfgröße des Tests von DAVID-HARTLEY-PEARSON bei 10000 Simulationsläufen für Stichprobenumfang $n = 50$ (eingezeichnet sind die empirischen Quantile $Q_{2.5} = 3.73401$ und $Q_{97.5} = 5.52671$, vergleiche Tab. 5)

Tabelle 5: Empirische Quantile der der Prüfgröße des DAVID-HARTLEY-PEARSON-Tests mittels Simulationsmethode bei 10 000 Simulationsläufen

n	Q _{0.5}	Q ₁	Q _{2.5}	Q ₅	Q ₉₅	Q _{97.5}	Q ₉₉	Q _{99.5}
3	1.73516	1.73794	1.74652	1.75804	1.99932	1.99980	1.99996	1.99999
4	1.81494	1.85205	1.91693	1.99176	2.42941	2.43967	2.44536	2.44721
5	1.98352	2.02117	2.07867	2.14281	2.75173	2.77984	2.80150	2.81152
6	2.11571	2.15034	2.21996	2.28629	3.01238	3.05468	3.09387	3.11124
7	2.21171	2.26099	2.34335	2.41121	3.22649	3.28729	3.34981	3.38030
8	2.29436	2.36863	2.43347	2.50580	3.40382	3.47244	3.54852	3.58714
9	2.39648	2.45105	2.52555	2.60086	3.55800	3.64519	3.72665	3.77792
10	2.45913	2.51163	2.59730	2.67221	3.69737	3.79788	3.89212	3.94853
11	2.52272	2.58380	2.66591	2.74557	3.80982	3.91235	4.02531	4.09021
12	2.58532	2.64603	2.72937	2.80314	3.91734	4.02751	4.14188	4.22384
13	2.63405	2.70609	2.79960	2.87777	4.02760	4.13885	4.25044	4.34209
14	2.70081	2.76918	2.84601	2.93126	4.09557	4.22065	4.36647	4.46014
15	2.72822	2.79755	2.88309	2.97217	4.17184	4.28584	4.44965	4.54950
16	2.76558	2.84036	2.93274	3.02836	4.23525	4.36880	4.50386	4.59816
17	2.81530	2.88110	2.97545	3.06543	4.31813	4.45254	4.59347	4.70327
18	2.85880	2.91921	3.01298	3.10687	4.38438	4.52798	4.66333	4.76736
19	2.89970	2.95944	3.05979	3.14968	4.44371	4.58540	4.73988	4.88609
20	2.95577	3.01144	3.09956	3.18524	4.51041	4.64390	4.79413	4.90550
30	3.18687	3.25758	3.36942	3.46396	4.88532	5.05892	5.26579	5.43092
40	3.41273	3.48716	3.58208	3.68458	5.16269	5.31500	5.54512	5.69490
50	3.53288	3.62674	3.73401	3.82893	5.34887	5.52671	5.76595	5.91372
100	4.03259	4.09491	4.21027	4.31353	5.91918	6.10302	6.34227	6.48697

2.7 GRUBBS-Test und TIETJEN-MOORE-Test

Beim GRUBBS-Test wird jeweils für das Maximum und das Minimum der Stichprobe bezüglich des empirischen Mittelwertes und der Standardabweichung standardisiert und mit den von n abhängigen kritischen Werten des GRUBBS-Tests verglichen. Damit ist der GRUBBS-Test als einseitiger Test formuliert. Die Nullhypothese, dass das Minimum (min) kein Ausreißer ist, wird zum Niveau α verworfen, wenn gilt:

$$G_{\min} = \frac{\bar{x} - \min}{s} > T_{n,1-\alpha} .$$

Entsprechendes gilt für das Maximum (max):

$$G_{\max} = \frac{\max - \bar{x}}{s} > T_{n,1-\alpha} .$$

\bar{x} entspricht dabei dem Mittelwert der Datenreihe, s der Standardabweichung, $T_{n,1-\alpha}$ steht für den von n abhängigen kritischen Wert des GRUBBS-Tests. GRUBBS-Test und CHAUVENET-Test sind identisch. Der kritische Wert ist für beide Prüfgrößen G_{\min} und G_{\max} gleich. Im Simulationsexperiment wird man sehen, dass sogar beide Prüfverteilungen gleich sind. Da die Prüfgrößen als Standardnormierungen des

extremwertverdächtigen Maximums oder des Minimums aufgefasst werden können bezüglich der geschätzten s und \bar{x} liegt die Vermutung nahe, dass die Prüfverteilung aus einer t-Verteilung abgeleitet werden kann. Unter H_0 , dass keine Ausreißer vorhanden sind, besitzen sowohl G_{\min} als auch G_{\max} eine Prüfverteilung, die als Funktion einer t-Verteilung dargestellt werden kann:

$$z_{\alpha} = \frac{n-1}{\sqrt{n}} \sqrt{\frac{t_{\alpha, n-2}^2}{n-2+t_{\alpha, n-2}^2}} .$$

Dabei bezeichnet $t_{\alpha, m}$ das α -Quantil einer t-Verteilung mit m Freiheitsgraden. Wenn der Test einen Ausreißer entdeckt, wird dieser aus der Stichprobe entfernt und ein neuer GRUBBS-Test mit den Daten der restlichen Stichprobe des Umfangs $n - 1$ durchgeführt. Das kann so lange geschehen, bis kein Ausreißer mehr entdeckt wird. Der GRUBBS-Test ist damit ein statistischer Test, der wiederholt nacheinander angewandt werden kann. Die Berechnung der kritischen Werte mit Hilfe des SAS-Systems ist einfach, weil die Quantilfunktion der t-Verteilung zu den Standardfunktionen gehört.

Große Probleme bereiten Ausreißer, wenn sie gehäuft auftreten. Ein statistischer Test, der genau einen Ausreißer erkennt, darf nach Elimination dieses Ausreißers mit der entsprechenden Reduktion des Stichprobenumfangs nicht erneut angewandt werden (von einigen Ausnahmen abgesehen, die zum iterativen Gebrauch ersonnen wurden), weil man die Stichprobenelemente nicht zufällig, sondern der Größe nach reduziert.

Die TIETJEN-MOORE-Tests, von den Autoren auch Ausreißertests vom GRUBBS-Type genannt, arbeiten analog zum GRUBBS -Test, wenn mehrere Ausreißer k auf einer Seite der Stichprobe vermutet werden. Für den Fall $k = 1$ fällt der TIETJEN-MOORE -Test mit dem GRUBBS -Test zusammen.

Von GRUBBS wurde ein Test entwickelt, der zwei Ausreißer am gleichen Ende der Stichprobe, entweder beide am oberen oder beide am unteren Ende oder zwei Ausreißer, je einen am oberen und einen am unteren Ende der Stichprobe erkennt.

TIETJEN und MOORE (1972) entwickelten die Vorgehensweise von GRUBBS weiter. Ihre Tests nennen sie dementsprechend auch GRUBBS-Typ-Statistiken. Mit diesen Tests könnte man bei einem Stichprobenumfang n bis maximal $\text{INT}(n/2)$ Ausreißer erkennen. Wenn die größten k Stichprobenelemente ausreißerverdächtig sind, nimmt man als Prüfgröße

$$L_k = \sum_{i=1}^{n-k} (x_i - \bar{x}_k)^2 / \sum_{i=1}^n (x_i - \bar{x})^2 \text{ mit } \bar{x}_k = (\sum_{i=1}^{n-k} x_i) / (n - k).$$

L_1 ist gleich der Prüfgröße, die bei GRUBBS S_n^2/S^2 genannt wird, wobei S_n^2 die Summe der Abweichungsquadrate der $n - 1$ kleinen Werte der Stichprobe von ihrem Mittel ist. L_2 ist gleich der GRUBBS-Prüfgröße $S_{n,n-1}^2/S^2$, wobei $S_{n,n-1}^2$ Summe der Abweichungsquadrate der $n - 2$ kleinen Werte der Stichprobe von ihrem Mittel ist. Ebenso hat GRUBBS Prüfgrößen für Ausreißer am unteren Ende der Stichprobe definiert,

die man allerdings durch Multiplikation der Stichprobenelemente mit -1 auf obiges Problem zurückführen kann. Liegen gleichzeitig am unteren und oberen Stichprobenende ausreißerverdächtige Elemente, dann soll man übergehen zur Prüfgröße für die der Größe nach geordneten Absolutbeträge $y_i = \text{ABS}(x_i)$

$$E_k = \sum_{i=1}^{n-k} (y_i - \bar{y}_k)^2 / \sum_{i=1}^n (y_i - \bar{y})^2 \quad \text{mit } \bar{y}_k = (\sum_{i=1}^{n-k} y_i) / (n - k) .$$

Die kritischen Werte der Prüfgrößen erhielten TIETJEN und MOORE durch ein Simulationsprogramm in FORTRAN IV, ausgeführt auf einer CDC6600. Der verwendete Zufallszahlengenerator war ein multiplikativer Kongruenzgenerator des Typs

$$X_{n+1} = \text{MOD}(X_n \cdot 8.5 \cdot 10^{16} + 5, 2^{48}).$$

Die $N(0,1)$ -verteilten Zufallszahlen wurden mittels BOX-MULLER-Methode aus gleichverteilten Zufallszahlen erzeugt. Es wurden jeweils 10 000 Stichproben des Umfangs n gezogen. Da mittlerweile „verbesserte“ Zufallszahlengeneratoren zur Verfügung stehen, ist es berechtigt, die Tabelle der kritischen Werte zu verbessern.

Ein SAS-Programm, das die relativen Häufigkeiten (Abb. 9), die empirischen Verteilungsfunktionen der Prüfgröße L_k (Abb. 10) und die simulierten kritischen Werte für Tab. 6 erzeugt, wird hier nicht angegeben. Tab. 6 unterscheidet sich nur wenig von den Originaltabellen von TIETJEN und MOORE (1971).

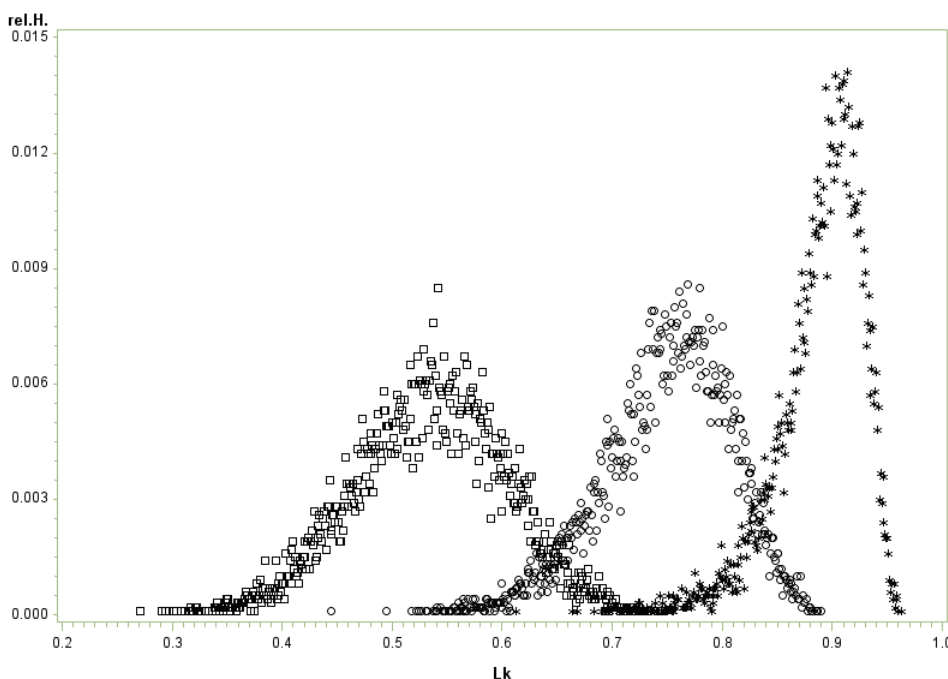


Abbildung 9: Häufigkeitsfunktion der Prüfgröße L_k bei 10 000 Simulationsläufen für Umfang $n = 50$ und $k = 1, 3$ und 8 ausreißerverdächtige Stichprobenelemente (von rechts nach links)

Tabelle 6: Quantile der Testgröße von Tietjen/Moore im Simulationsexperiment vom Umfang 10 000 für einen ausreißerverdächtigen Punkt zu verschiedenen Stichprobenumfängen n (fett gedruckt sind die Quantile aus der Originalarbeit von Tietjen/Moore)

n	k = 1										
	q _{0.5}	q ₁	Grubbs q ₁	q _{2.5}	Grubbs q _{2.5}	q ₅	Grubbs q ₅	q ₉₅	q _{97.5}	q ₉₉	q _{99.5}
5	.024	.039	.045	.080	.084	.124	.125	.793	.825	.859	.875
6	.059	.088	.091	.143	.146	.198	.203	.812	.837	.866	.881
7	.113	.147	.148	.203	.209	.259	.273	.827	.853	.875	.891
8	.144	.184	.202	.259	.262	.329	.326	.834	.856	.879	.890
9	.198	.239	.235	.314	.308	.373	.372	.847	.867	.886	.899
10	.228	.275	.280	.347	.350	.413	.418	.854	.873	.891	.903
11	.278	.328	.327	.385	.366	.446	.454	.861	.878	.897	.907
12	.308	.354	.371	.421	.440	.481	.489	.865	.882	.899	.909
13	.346	.397	.400	.456	.462	.508	.517	.872	.887	.904	.914
14	.366	.413	.424	.472	.493	.532	.540	.876	.891	.905	.913
15	.397	.445	.450	.508	.498	.555	.556	.879	.894	.908	.915
16	.423	.473	.473	.530	.537	.577	.575	.885	.898	.910	.918
17	.443	.491	.480	.545	.552	.592	.594	.886	.899	.912	.920
18	.461	.506	.502	.569	.570	.612	.608	.890	.904	.915	.923
19	.484	.526	.508	.578	.573	.622	.624	.892	.905	.917	.925
20	.492	.539	.533	.596	.595	.637	.639	.896	.907	.919	.926
25	.575	.609	.603	.655	.656	.693	.696	.908	.918	.927	.933
30	.630	.662	.650	.699	.699	.730	.730	.917	.924	.933	.938
35	.672	.696	.690	.732	.732	.761	.762	.922	.930	.937	.941
40	.702	.726	.722	.758	.755	.783	.784	.929	.936	.941	.944
45	.730	.746	.745	.776	.773	.802	.802	.934	.939	.945	.949
50	.748	.767	.768	.796	.796	.818	.820	.938	.943	.948	.951

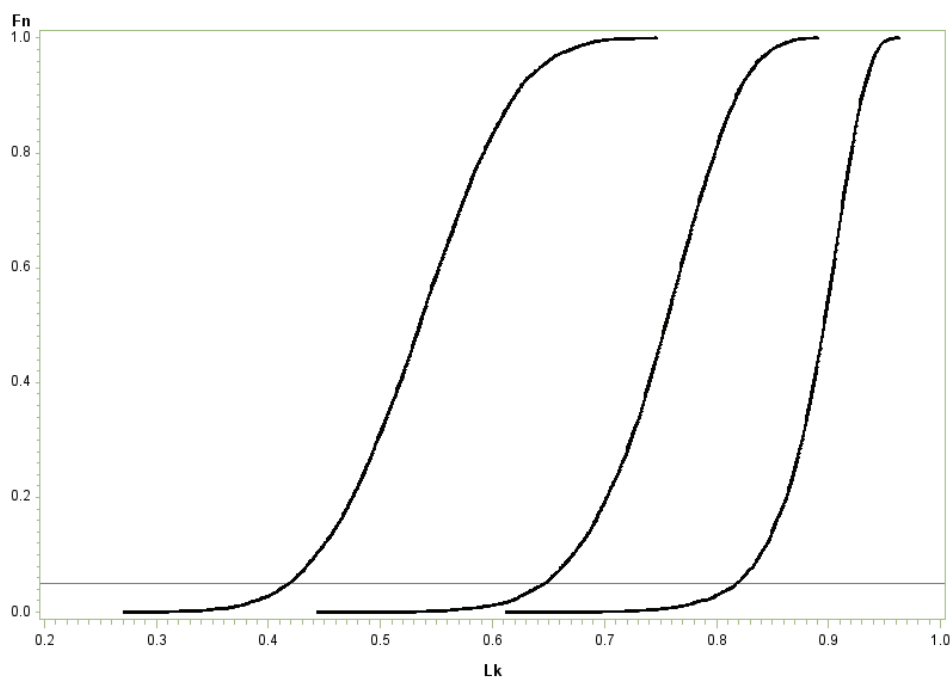


Abbildung 10: Emp. Verteilungsfunktion der Prüfgröße L_k bei 10.000 Simulationsläufen für den Stichprobenumfang $n=50$ und $k=1,3$ und 8 ausreißerverdächtige Stichprobenelemente (von rechts nach links), Referenzlinie $\alpha = 0.05$

Bemerkungen:

- Während der bisher beschriebene Test nur Ausreißer auf einer Seite der Verteilung finden kann, ist man mit E_k in der Lage, auf beiden Seiten der Stichprobe gleichzeitig nach Ausreißern zu fahnden. Die zweite Methode wird dem Leser als Übungsaufgabe überlassen.
- Die Werte der Zufallsgröße L_k variieren zwischen 0 und 1. Wenn die Summen der Abweichungsquadrate in Zähler und Nenner von L_k in etwa übereinstimmen – wie das bei der Gültigkeit der Nullhypothese der Fall ist – erhält man große Werte von L_k . Gegen H_0 und für H_1 entscheidet man sich, wenn die Prüfgröße L_k kleine Werte annimmt.
- Sinnvoll ist damit ein einseitiger Test, der H_0 ablehnt, wenn die Werte von L_k unter die Quantile $q_{0,01}$ bzw. $q_{0,05}$ fallen, je nachdem man $\alpha = 0.01$ bzw. $\alpha = 0.05$ wählt.

3 Simulationsexperiment zur Wertung der Testmethoden zur Ausreißerererkennung mittels Powerbestimmung

Die Powerbestimmungen werden bezüglich zweier Simulationsexperimente vom Stichprobenumfang $n = 50$ durchgeführt. Unter H_0 befindet sich kein Ausreißer in der Stichprobe. H_1 wird auf folgende Weise parametrisiert. In der Stichprobe werden den 90% einer Standardnormalverteilung 10% einer weiteren $N(\mu, 1)$ -Verteilung beigemischt.

Wenn $\mu = 0$, dann liegt H_0 vor. Bei jedem $\mu > 0$ kommen Werte hinzu, die nicht zur Standardnormalverteilung gehören und ausreißerverdächtig sind. Je größer μ wird, umso mehr Ausreißer erwartet man in der Stichprobe und die Entscheidung für H_0 fällt immer seltener. (Man beachte aber, dass bei manchen Ausreißertests die Ausreißer so in die Prüfgröße eingehen, dass sie ihr Wesen verbergen können.)

Vergleiche dazu die Abb.11. Das folgende SAS-Programm realisiert die Erzeugung von 10000 „verschmutzten“ standardnormalverteilten Größen. Diese werden in die jeweiligen SAS-Programme eingebunden und für $\alpha = 0, 0.5, 1, 2, 3, 4, 5$ und 6 durchgerechnet. In den jeweiligen Programmen wird mit der PROC FREQ ermittelt, wie oft H_0 angenommen wurde. Das ist gerade der Fehler 2. Art β des Tests, denn für $\mu > 0$ gilt bekanntlich H_1 .

SAS-Programm zur Erzeugung von verschmutzten Normalverteilungen

```
%let sim=10000; /* Simulationsumfang */
%let n=20;      /* Stichprobenumfang */
%let proz=0.1; /* Verschmutzungsgrad mit N(my,1) */
%let my=5;

data test; /* Erzeugung von sim verschmutzten Stichproben vom Umfang
n */
drop j;
```

```
do i=1 to &sim;
  do j=1 to &n;
    if UNIFORM(-i)<&proz then do x=NORMAL(i)+&my;output;end;
    else do;x=NORMAL(i);output;end;
  end;
end;run;
```

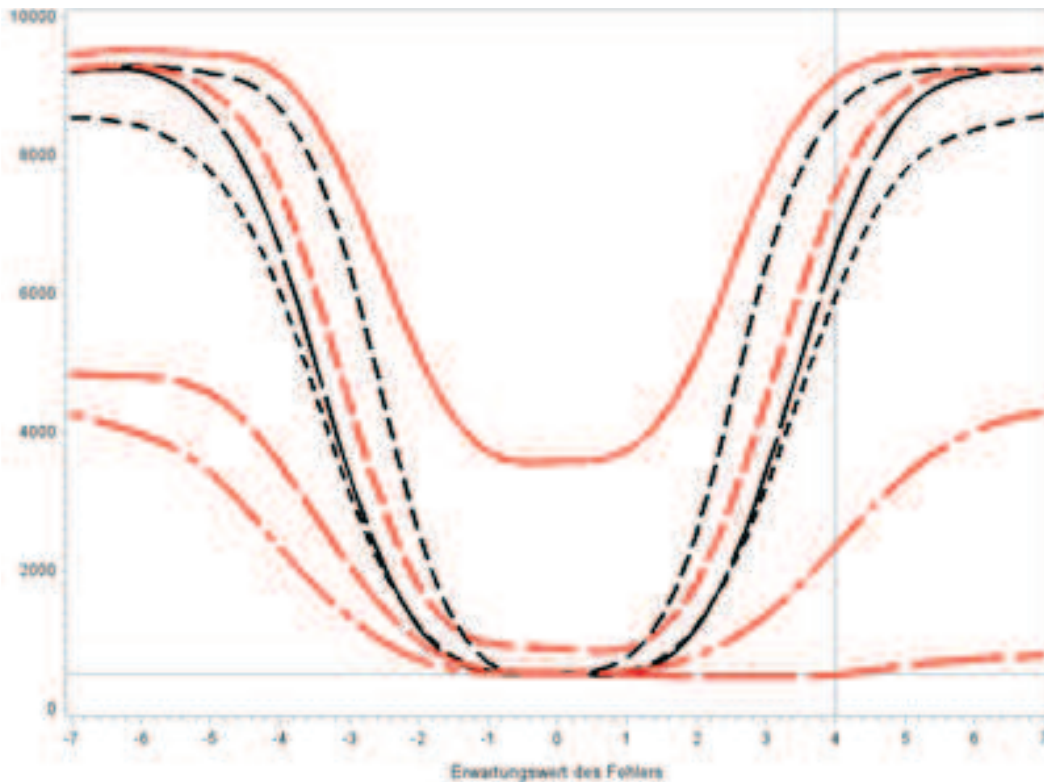


Abbildung 11: Simulierte Powerfunktionen für die Ausreißertests – Anzahl Entscheidungen für H_1 bei 5% Verschmutzung, variierenden Erwartungswerten μ von -7 bis +7 und $\sigma = 1$ des Fehlers bei 10000 Simulationen (Powerfunktionen von oben nach unten an der Referenzlinie $\mu = 4$: Boxplotmethode aus SAS, Maximummethode, MAD-Methode, variierte Boxplotmethode, Peirce-Methode, David-Hartley-Pearson und Dean/Dixon)

Es ergeben sich folgende Schlussfolgerungen:

1. Die Boxplot-Methode ist ein denkbar schlechter Test, um Ausreißer zu erkennen. Bei $\mu = 0$, also an der H_0 -Stelle, tritt ein Entscheidungsfehler ein, der bei etwa 38% liegt. Da hilft auch nicht weiter, dass sie bei Vorliegen von H_1 gegenüber allen anderen Tests besser ist. Sie ist in Abb. 10 als rot - nicht geeignet - markiert.
2. Die variierte Boxplot-Methode (in Abb. 10 lang gestrichelt) schneidet nach der Maximum-Methode als zweitbeste ab. Der α -Fehler wird für $\mu = 0$ eingehalten.
3. Die MAD-Methode liegt in großen Bereichen von H_1 in der Power zwischen der Maximum- und der variierten Boxplot-Methode. Der Fehler 1. Art wird mit ca. 8% verfehlt. Schuld daran ist sicher der für alle Stichprobenumfänge einheitliche Parameter von 3.5. Eine variierte MAD-Methode, bei der man 3.5 durch eine

Konstante $k(n)$ ersetzt, könnte ähnlich wie bei der Boxplot-Methode Abhilfe schaffen.

4. Vom David-Hartley-Pearson-Test ist die geringe Power bekannt.
5. Der Dean/Dixon-Test ist in der Power kaum besser als der David-Hartley-Pearson-Test. Als Test, der ein Minimum erkennen soll, ist folglich seine Power für $\mu > 0$ sehr schlecht.

Es wurde ein zweites Simulationsexperiment durchgeführt, bei dem der Verschmutzungsgrad von 0 bis 10% ansteigt. Die verschmutzende Normalverteilung ist mit $N(3,1)$ fixiert. Die Ergebnisse sind in Abb.12 illustriert. Für die Powerfunktionen in Abb. 12 gelten die gleichen Aussagen wie in Abb.11. Die Schlussfolgerungen sind analog zu denen aus dem ersten Simulationsexperiment.

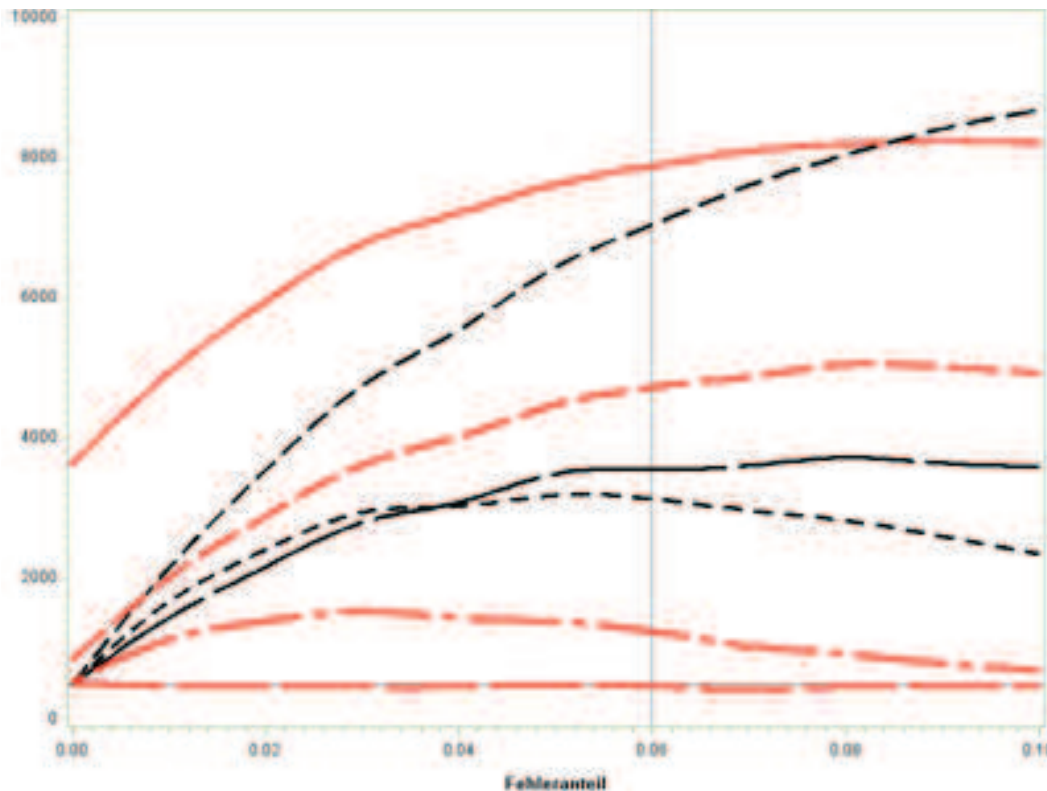


Abbildung 12: Simulierte Powerfunktionen für die Ausreißertests – Anzahl Entscheidungen für H_1 bei variierendem Verschmutzungsgrad von 0 bis 10% und fixierter Verschmutzungsverteilung $N(3,1)$ bei 10000 Simulationen (Powerfunktionen von oben nach unten an der Referenzlinie Fehleranteil 0.06: Boxplotmethode aus SAS, Maximummethode, MAD-Methode, variierte Boxplotmethode, Pierce-Methode, David-Hartley-Pearson und Dean/Dixon)

Literatur

- [1] Chauvenet, W. (1863): A manual of spherical and practical astronomy V.II.,Lippincott, Philadelphia
- [2] Dixon, W. J. (1953): Processing data for outliers. J. Biometrics, 9, pp.74-89

- [3] Dean, R. B.; Dixon, W.J. (1951): Simplified statistics for small numbers of observations. *Anal. Chem.*, 23, pp. 636-638
- [4] Frigge, M.; Hoaglin, D. C.; Iglewicz, B. (1989): Some implementations of the boxplot, *The American Statistician*, Vol. 43, No. 1 pp. 50-54
- [5] Gould, B. A. (1855): On Peirce's criterion for the rejection of doubtful observations, with tables for facilitating its application, *Astronomical Journal*, iss. 83, vol. 4, no. 11, pp. 81--87.
- [6] Grubbs, F. E. (1950): Sample Criteria for Testing Outlying Observations. *Annals of Mathematical Statistics*, 21, Nr.1, pp. 27–58,
- [7] Grubbs, F. E.; Beck, G. (1972): Extension of sample sizes and percentage points for significance tests of outlying observations, *Technometrics*. Vol. 14, pp. 847-854.
- [8] Iglewicz, B.; Banerjee, S. (2001): A simple univariate outlier identification procedure, *Proceedings of the Annual Meeting of the American Statistical Association*, August, 5-9,
- [9] Peirce, B. (1852): Criterion for the rejection of doubtful observations , *Astronomical Journal II*, 45, pp. 161-163
- [10] Ross, S. M. (2003): Peirce's criterion for the elimination of suspect experimental data," *Journal of Engineering Technology*, vol. 2, no. 2, pp. 1-12
- [11] Tietjen, G. L.; Moore, R. (1972): Some Grubbs-type statistics for the detection of several outliers. *Technometrics*. Vol.14. No. 3. pp.
- [12] Walsh, J. E. (1950): Some nonparametric tests of whether the largest observations of a set are too large or too small. *Annals of Mathematical Statistics*, 21 , pp. 583-592

Anhang Simulationsprogramm

Für das Simulationsprogramm zur Bestimmung der Power der Ausreißertests (Boxplotmethode aus SAS, Maximummethode, MAD-Methode, variierte Boxplotmethode, Pierce-, David-Hartley-Pearson- und Dean/Dixon-Methode wird auf Folgendes hingewiesen:

- Die Stellen, an denen man die Dateien für die Quantilbestimmung der Tests erhält, sind durch entsprechende Kommentare gekennzeichnet. Die Tabellen im Beitrag sind in einem vorangehenden Lauf erzeugt worden.
- Man beachte, dass im Programm für andere Umfänge andere kritische Werte der Tests einzufügen sind!
- Im ersten data-Step werden ‚n mal sim‘ Datensätze in eine Datei gespielt. Bei 10 000 Simulationen und einem Stichprobenumfang von $n = 200$ sind das 2 000 000 Datensätze.

Simulationsprogramm zur Bestimmung der Power der Ausreißertests

```

%let sim=10000; /* Simulationsumfang */
%let n=50;      /* Stichprobenumfang */
%let proz=0.09; /* Verschmutzungsanteil */
%let my=3;      /* Parameter der Verschmutzungsverteilung N(my,1)
*/

data test;      /* Erzeugung von sim verschmutzten Stichproben vom
Umfang n */
do i=1 to &sim;
  do j=1 to &n;
    if UNIFORM(-39763)<&proz then do;x=NORMAL(i)+&my;goto
marke;end;
    else x=NORMAL(i);
    marke:output;
  end;
end;run;

/* Boxplotmethode *****/
proc means data=test noprint;
by i;var x;
output out=help_box p25=QU p50=Med p75=QO max=maxi min=mini std=s
mean=m;run;

data help_box;
set help_box;
drop _FREQ_ _TYPE_;
flag_box1=0;flag_box2=0;
if qu-1.5*(qo-qu)>mini or qo+1.5*(qo-qu)<maxi then flag_box1=1; /*
Boxplot-Methode */
if qu-2.25*(qo-qu)>mini or qo+2.25*(qo-qu)<maxi then flag_box2=1;
/* variierte Methode, Fausformel k=2.25, bei kleinem n Tabellenwert
*/
run;
/* Datei help_box kann im Fall von proz=0 oder my=0 (H0)
zur Bestimmung der Quantile von Boxplot- und variiertes
Boxplotmethode verwandt werden */

proc freq data=help_box noprint ;
tables flag_box1/out=ausr_box1;run;
data ausr_box1;
set ausr_box1;
where flag_box1=1;
keep box1 Proz my;
box1=count;Proz=&proz;my=&my;run;

proc freq data=help_box noprint;
tables flag_box2/out=ausr_box2;run;
data ausr_box2;

```

```
set ausr_box2;
where flag_box2=1;
keep boxJ Proz my;
boxJ=count;Proz=&proz;my=&my;run;

/***** PEIRCE-Methode *****/
data help_peirce;
set help_box;
r1=Max (ABS ((mini-m) /s) ,ABS ((maxi-m) /s));
flag_Peirce=0;
if r1>=3.131 then flag_Peirce=1;
/* bei anderem Stichprobenumfang anderen Tabellenwert einsetzen*/
run;

proc freq data=help_peirce noprint;
tables flag_Peirce/out=ausr_Peirce;run;
/* Datei ausr_Peirce kann im Fall von proz=0 oder my=0 (H0)
zur Bestimmung der Quantile des Peirce-Tests verwandt werden */
data ausr_Peirce;
set ausr_Peirce;
where flag_Peirce=1;
keep Peirce Proz my;
Peirce=count;Proz=&proz;run;

/***** Maximummethode *****/
data help_maximum;
set help_box;
flag_max=0;
if maxi >=3.28704 or mini<=-3.28704 then flag_max=1;run;
/* Datei help_maximum kann im Fall von proz=0 oder my=0 (H0)
zur Bestimmung der Quantile des Maximum-Tests verwandt werden */

proc freq data=help_maximum noprint;
tables flag_max/out=ausr_Max;
run;
data ausr_max;
set ausr_max;
where flag_max=1;
keep MAX Proz my;
MAX=count;Proz=&proz;run;

/***** Median of Absolute Deviation, MAD-Methode*****/
proc means data=test noprint;
by i;
var x;
output out=abc Median=med; /* p50(xi) */run;
data gesamt;
merge test abc;
by i;
y=ABS (x-med) ;run;

proc means data=gesamt noprint;
```

```

var y;
by i;
output out=MAD_DAT p50=MAD; /* MAD*/run;
data gesamt;
merge gesamt mad_dat;
by i;
flag_MAD=1;
Mz=0.6745*((x - med)/MAD);
if -3.5<=mz<=3.5 then flag_MAD=0;run;

proc means data=gesamt noprint;
var flag_MAD;
by i;
output out=aaa max=z;run;

proc freq data=aaa noprint;
tables z/out=ausr_MAD;run;
data ausr_MAD;
set ausr_MAD;
where z=1;
keep MAD Proz my;
MAD=count;Proz=&proz;run;

/** Dean/Dixon-Test *****/

data index; /* Erzeugung von j-Index*/
do i=1 to &sim;
  do j=1 to &n;
    output;
  end;
end;run;
proc sort data=test;
by i x;
run;
data lkj;
merge test index;
by i;
run;
data zwei;
set lkj;
keep i x2;where j=2;
x2=x;run;
data drei;set lkj;keep i x3;where j=3;x3=x;run;
data n_minus_2;set lkj;keep i xn_2;where j=&n-2;xn_2=x;run;
data n_minus_1;set lkj;keep i xn_1;where j=&n-1;xn_1=x;run;

data dean_dixon;
merge help_box zwei drei n_minus_2 n_minus_1;
by i;
keep i mini maxi x2 x3 xn_2 xn_1;run;
data dean_dixon;
set dean_dixon;

```

```
if &n<=7 then DD=(x2-mini)/(maxi-mini);
if 8<=&n<=10 then DD=(x2-mini)/(xn_1-mini);
if 11<=&n<=13 then DD=(x3-mini)/(xn_1-mini);
if 14<=&n then DD=(x3-mini)/(xn_2-mini);
flag_dd=1;
if 0.02446<dd<0.34681 then flag_dd=0;
/* bei anderem Umfang andere Tabellenwerte einsetzen */run;
/* Datei dean_dixon kann im Fall von proz=0 oder my=0 (H0)
   zur Bestimmung der Quantile des Dean/Dixon-Tests verwandt werden
*/
```

```
proc freq data=dean_dixon noprint;
tables flag_dd/out=ausr_dd;run;
/* Die Datei ausr_dd kann bei proz=0 oder my=0 (H0)
   zur Bestimmung der Quantile von Tab.4 verwandt werden */
data ausr_dd;
set ausr_dd;
where flag_dd=1;
keep dd Proz my;
DD=count;
Proz=&proz;run;
```

```
/** David-Hartlay-Pearson *****/
data david;
set help_box;
DHP=ABS(maxi-mini)/s;
flag_DHP=1;
if 3.73401<DHP<5.52671then flag_dhp=0;
/* Anderer Stichprobenumfang, andere Tabellenwerte */
run;
/* Datei david kann im Fall von proz=0 oder my=0 (H0)
   zur Bestimmung der Quantile des David-Hartlay-Pearson-Tests
   verwandt werden */
```

```
proc freq data=david noprint;
tables flag_dhp/out=ausr_dhp;run;
data ausr_dhp;
set ausr_dhp;
where flag_dhp=1;
keep DHP Proz my;
DHP=count;Proz=&proz;run;
```

```
***** Ausgabe
*****/
data Ausgabe;
merge ausr_box1 ausr_box2 ausr_Peirce ausr_MAX ausr_MAD ausr_DD
ausr_DHP;
by proz;run;
proc print data=ausgabe noobs;run;
```